

АКАДЕМИЯ НАУК СССР

А.Б.ВИСТЕЛИУС

ОСНОВЫ МАТЕМАТИЧЕСКОЙ ГЕОЛОГИИ

$\delta' (z'z) \overset{-1}{z}' \Sigma z (z'z) \overset{-1}{z}' \delta$

АКАДЕМИЯ НАУК СССР
ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ ОРДЕНА ЛЕНИНА
МАТЕМАТИЧЕСКОГО ИНСТИТУТА им. В. А. СТЕКЛОВА
ЛАБОРАТОРИЯ МАТЕМАТИЧЕСКОЙ ГЕОЛОГИИ

А.Б.ВИСТЕЛИУС

ОСНОВЫ МАТЕМАТИЧЕСКОЙ ГЕОЛОГИИ

(определение предмета,
изложение аппарата)

3196



ЛЕНИНГРАД
«НАУКА»
ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ
1980



Основы математической геологии (определение предмета, изложение аппарата). Вистелиус А. Б. Л., «Наука», 1980. 389 с.

Рассматриваются идеи, определявшие развитие геологии на протяжении последних 300 лет. Показывается, что геология подготовлена для создания собственной математической дисциплины. Такой дисциплиной является математическая геология, опирающаяся на вероятностную природу геологических явлений. Для изучения стохастической организации геологических явлений требуется специфический математический аппарат. Дается изложение основ этого аппарата, некоторых основных положений теории вероятностей и математической статистики. Детально рассматриваются методы исследования случайных последовательностей, играющие большую роль при построении математических моделей геологических явлений. Лит. — 141 назв., ил. — 29, табл. — 12.

Ответственные редакторы:

академик АН УССР Б. В. ГНЕДЕНКО
доктор геол.-минер. наук М. А. РОМАНОВА

ОГЛАВЛЕНИЕ

Предисловие	6
Введение	9
Глава I. Математическая геология и развитие геологических наук	15
I.1. Введение	15
I.2. Развитие геологии и смена парадигмов	16
I.3. Организация среды и типичные структуры	23
I.4. Постановка задач, роль модели в поисках решения	29
I.5. Математическая геология и ее развитие	33
Литература	37
Глава II. Вероятностное пространство и случайные величины	41
II.1. Введение	41
II.2. Дискретное пространство элементарных событий	42
II.2.1. Вероятностное пространство	42
II.2.2. Случайные величины	45
II.3. Аксиоматика Колмогорова. Интеграл Лебега	48
II.3.1. Вероятностное пространство и случайные величины	48
II.3.2. Интеграл Лебега	52
II.3.3. Числовые характеристики случайных величин	56
II.4. Примеры распределений случайных величин	59
II.4.1. Дискретные распределения	59
II.4.2. Абсолютно непрерывные распределения	63
II.5. Векторные случайные величины	69
II.5.1. Произведение вероятностных пространств	70
II.5.2. Распределение векторных случайных величин	71
II.5.3. Характеристики векторных случайных величин	76
II.5.4. Примеры распределений векторных случайных величин	80
II.5.5. Условные распределения относительно случайных величин	89
II.6. Преобразования случайных величин	98
II.6.1. Линейное преобразование	99
II.6.2. Некоторые нелинейные преобразования	102
Пример II.4. Генетические соотношения между концентрациями элементов и процентный пересчет	110
II.6.3. Рандомизация	114
II.7. Последовательности независимых случайных величин и предельные теоремы	116
II.7.1. Неравенство Чебышева и закон больших чисел	116
II.7.2. Центральная предельная теорема	119

II.8.	Случайные процессы и геометрические вероятности	120
II.8.1.	Случайные последовательности	121
II.8.2.	Стационарные случайные последовательности и эргодичность	125
II.8.3.	Случайные процессы	129
II.8.4.	Точечные процессы	135
	Пример II.6. О разное обломочного материала в седиментационном бассейне	139
II.8.5.	Геометрические вероятности	142
	Литература	150
Глава III. Основные статистические понятия — вопросы оценивания и проверка гипотез		153
III.1.	Вводные замечания	153
III.2.	Точечное оценивание	156
III.2.1.	Основные идеи и определения	157
	Пример III.1. Оценки переходных вероятностей в предположении их пропорциональности	159
	Пример III.2. О смещении информационных статистик	164
	Пример III.3. О достаточности максимального наблюдения при оценивании правой точки усечения	175
III.3.	Проверка статистических гипотез	178
III.3.1.	Основные идеи и важнейшие определения	179
III.3.2.	Построение наилучших критических областей при проверке гипотез о параметре θ на примере равномерного распределения	183
III.3.3.	О некоторых важных характеристиках статистического критерия (продолжение исследования гипотезы равномерного распределения)	191
III.3.4.	Дальнейшее развитие теории проверки гипотез — сложная гипотеза против сложной альтернативы	197
III.3.5.	Критерий отношения правдоподобия	201
	Пример III.4. О распределении вероятностей концентрации Na_2O в базальтах мира	205
III.3.6.	О доверительных интервалах	212
	Литература	216
Глава IV. Случайные последовательности и их марковские модели		218
IV.1.	Введение	218
IV.2.	Вероятностные структуры марковских последовательностей	224
IV.2.1.	Случайные последовательности и связанные с ними вероятности	224
IV.2.2.	Стационарность, однородность и обратимость	227
IV.2.3.	Марковские последовательности	229
IV.2.4.	Последовательности независимых испытаний и простые цепи Маркова	234
IV.2.5.	Сложные цепи Маркова и существенно немарковские последовательности	242
IV.2.6.	Распределения серий в марковских цепях	246
IV.3.	Матричные методы изучения цепей Маркова	250
IV.3.1.	Марковские матрицы и стохастические векторы	250
IV.3.2.	Предельные распределения в марковских цепях	253
IV.3.3.	Некоторые расчеты, связанные с цепями Маркова	264
	Пример IV.10. О структуре ковариационных последовательностей для чередования мощностей слоев некоторых осадочных толщ	268
IV.3.4.	Обратимость, периодичность и число состояний	272
IV.3.5.	Степени свободы	274
IV.3.6.	Вычисления со сложными цепями Маркова	276

IV.4. Некоторые обобщения марковского свойства	283
IV.4.1. Восстанавливающие события	283
IV.4.2. Типы марковских переходов	286
IV.5. Трехмерные упаковки и марковские последовательности	292
Литература	294
Глава V. Преобразования марковских цепей	296
V.1. Введение	296
V.2. Укрупнение по множеству состояний	302
V.2.1. Сильное и слабое укрупнение	303
V.2.2. Некоторые достаточные условия слабого укрупнения цепи и частных переходов	307
V.2.3. Критерий слабого укрупнения при дополнительных ограничениях	314
V.3. Сгущение и разрежение	319
V.3.1. Внедрение между исходами смежных испытаний	321
Пример V.10. Алгоритм для получения реализаций существенно немарковских последовательностей	325
V.3.2. Схема увеличения серий	332
V.3.3. Простейшие схемы разрежения	335
V.4. О некоторых схемах замещения	340
V.5. Последовательности пакетов	343
Литература	347
Глава VI. Статистические заключения о свойствах случайных последовательностей и марковских гипотезах	348
VI.1. Введение	348
VI.2. Проверка однородности	350
VI.3. Проверка обратимости	353
VI.4. Критерий отношения правдоподобия Λ для проверки пропорциональности переходных вероятностей	359
VI.5. Марковские гипотезы	362
VI.6. Частная марковость	369
VI.7. Опытная проверка состоятельности статистических тестов о порядке марковости	374
Литература	379
Указатель имен	380
Указатель географических названий	382
Предметный указатель	384

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

- (Ω, \mathcal{F}, P) — вероятностное пространство.
 R^n — n -мерное векторное пространство.
 $\Omega_1 \times \Omega_2$ — декартово произведение двух множеств.
 \mathcal{F} — поле событий, σ -алгебра измеримых подмножеств множества элементарных событий.
 \mathcal{B}^n — борелевская σ -алгебра подмножеств пространства R^n .
 $\mathcal{F}_1 \otimes \mathcal{F}_2$ — произведение σ -алгебр \mathcal{F}_1 и \mathcal{F}_2 .
 $P_1 \times P_2$ — произведение вероятностных мер.
 $P(\cdot | B)$ — условная вероятность относительно события B .
 $F_\xi(x)$ — функция распределения.
 $p_\xi(x)$ — плотность распределения.
 $P(\cdot | \xi)$ — условная вероятность относительно случайной величины ξ .
 $p_{\xi_1} * p_{\xi_2}$ — свертка плотностей распределения.
- $$E f(\xi) = \int_{\Omega} f(\xi(\omega)) P(d\omega) = \int_{R^1} f(x) \mu_\xi(dx) = \int_{-\infty}^{\infty} f(x) dF_\xi(x) =$$
- $$= \int_{-\infty}^{\infty} f(x) p_\xi(x) dx$$
- математическое ожидание случайной величины $f(\xi)$, выраженное в виде интеграла Лебега от функции $f(\xi)$ по мере P или от функции f по мере μ_ξ , или в виде интеграла Стильтьеса от функции f по функции распределения F_ξ , или в виде интеграла Римана от произведения функции f на плотность распределения p_ξ .
- I_A — индикаторная функция множества A .
 $\pi(i_1, \dots, i_n)$ — проектирование вектора на подпространство, соответствующее координатам (i_1, \dots, i_n) .
 $J_g = \left(\frac{\partial g_i}{\partial x_j} \right)_{n \times n}$ — матрица Якоби.
 $H_k(x)$ — энтропия Шеннона.
 W — выборочное пространство.
 w_α — критическая область уровня α .
 Ω — параметрическое пространство.
 ω_0, ω_1 — параметрические множества для гипотезы и альтернативы.
 $\alpha(\theta)$ — функция уровня значимости.
 $\beta(\theta)$ — функция мощности.
 $R(x) = \frac{p_0(x; \hat{\theta}_0(x))}{p_1(x; \hat{\theta}_1(x))}$ — отношение правдоподобия.
 \mathcal{H}_{h-r-1} — σ -алгебра событий, зависящих лишь от исходов испытаний с номерами меньшими, чем $h-r$.
 \mathcal{H}_h — σ -алгебра событий, зависящих лишь от исходов испытаний, с номерами большими, чем h .
 $P^{(-)}$ — марковская матрица для обращенной цепи.
 $\mathcal{H}'_{(-)}$ — σ -алгебра событий из «прошлого».
 $\mathcal{H}'_{(0)}$ — σ -алгебра событий из «настоящего».
 $\mathcal{H}'_{(+)}$ — σ -алгебра событий из «будущего».
 Σ — множество состояний до укрупнения.
 \mathcal{S} — множество состояний; в главе 5 — множество состояний после укрупнения.
 G — система укрупнения.
 $\Sigma_I, \Sigma_J, \Sigma_K$ — классы эквивалентности при укрупнении.
 $a_i^{(k)}$ — «пакетное состояние» на i -м месте в пакете типа I , занимающее k -ю позицию в последовательности.
 Δ — статистика критерия отношения правдоподобия.

ПРЕДИСЛОВИЕ

В первой половине 40-х годов стало ясно, что в обозримом будущем развитие геологических наук не сможет проходить нормально, если не будет разработана специфическая для геологии дисциплина, которая может быть названа математической, или аналитической, геологией. В принципе ее положение в геологии должно быть аналогично положению математической физики в физических науках. Было также достаточно ясно, что до того, как будет получено в этом направлении достаточное количество нетривиальных результатов, пройдет по крайней мере несколько десятилетий. Действительно, прежде всего предстояло найти и если нужно развить раздел математики, аксиоматика которого была бы наиболее адекватна аксиоматике, описывающей геологические явления, вопрос о которой в то время не ставился. Затем следовало выделить наиболее перспективные для разработки области геологии и хотя бы на них показать возможности, создаваемые новым подходом. Для того же, чтобы это сделать, требовался в первую очередь большой объем полевых экспедиционных работ, так как новые идеи в геологии порождаются именно такими работами и сопровождающей их обработкой целеустремленно собранных материалов. Коллекции, накапливающиеся в результате отмеченных работ, необходимо было обрабатывать также самому исследователю, так как отрыв его от такой обработки повел бы к потере чувства реальности в окончательных выводах. Отмеченная деятельность нуждалась в специальной организации, способной на каждом этапе исследований давать результаты ясной научной и практической значимости, что далеко не просто, когда намеченная цель далека, а путь к ней сложен и требует обширных поисков и промежуточных разработок.

В итоге длительных усилий, в которых большую роль сыграло руководство Математического института АН СССР, все необходимое для работ такого типа было создано. Это позволило накопить материал для настоящей книги. В нее вошли личные исследования автора и в переработанном виде итоги многолетней деятельности

сотрудников руководимой им группы. Многочисленные примеры в тексте в значительной мере основаны на результатах полевых экспедиционных работ М. А. Романовой и Д. Н. Иванова. Некоторые рисунки иллюстрируют обработку исключительно трудоемкими методами коллекций, собранных автором, М. Е. Деминой и химиками нашей лаборатории; новые теоремы, фигурирующие в этой книге, собраны из различных исследований, проводившихся автором совместно с А. В. Фаасом, которому принадлежат и доказательства этих теорем. Систематическая переписка со многими коллегами и активное участие в работах производственных организаций позволили использовать в книге труднодоступные цифровые данные, в частности данные, накопленные за десятки лет многими архивами как у нас, так и за рубежом. Без указанных материалов книга никогда не могла бы быть написана.

Как отмечалось, подготовка материала для книги заняла длительное время. В течение этого периода автор пользовался помощью большого числа лиц, всем этим лицам автор выражает свою признательность. Невозможно не отметить влияние, которое имели на автора длительные контакты с Ю. А. Жемчужниковым, А. Н. Колмогоровым и Н. Н. Михайловым. Их помощь сыграла решающую роль в разработке основ математической геологии.

При подготовке рукописи весь ее текст и доказательства были любезно прочитаны Б. П. Харламовым, в соавторстве с которым написана вторая глава. Кроме того, первые три главы были внимательно прочитаны Н. А. Сапоговым, а третья глава — К. П. Латышевым. Их замечания способствовали улучшению текста.

При работе над текстом автор пользовался помощью своего постоянного сотрудника А. В. Фааса, инициатива и отзывчивость которого содействовали успешному завершению работы. Подготовка варианта рукописи для издательства осуществлена И. Э. Сирот.

Автор признателен редакторам книги — Б. В. Гнеденко и М. А. Романовой — за внимание к пожеланиям автора и неизменную благожелательность. Необходимо также отметить большую помощь, оказанную автору руководителями тех учреждений, в которых протекала его деятельность, — М. Ф. Двали, Н. Г. Келлем и Г. И. Петрашенем.

Разработка основ математической геологии не всегда проходила гладко. В трудные периоды жизни автора всегда поддерживало самоотверженное и чуткое участие в нем его матери — Натальи Леонидовны Вистелиус.

ВВЕДЕНИЕ

В книге об основах той или иной науки всегда предполагается, что соответствующая наука уже сформировалась и что из ее материала могут быть выделены основные факты. Для математической геологии такая схема не подходит, так как в монографическом плане интересующая нас наука еще не определена. Говоря об основах математической геологии, следует в первую очередь иметь в виду ее достаточно полное определение и те конкретные результаты, на которых она может быть развита.

Термин «математическая геология» и определение ее задач были даны автором этих строк в начале 60-х годов. Этот термин был принят международной общественностью, за чем последовало множество публикаций, авторы которых стремились внести свой вклад в новую область знания. Однако таких исследований, которые развивали бы математическую геологию и давали бы принципиально новые результаты для геологии вообще, оказалось немного. Поэтому для изложения основ математической геологии необходимо было выделить те результаты (по возможности в разных разделах геологии), которые бы наиболее точно удовлетворяли специфике интересующей нас дисциплины, развивали бы эту дисциплину и приносили бы факты, неизвестные ранее с геологической точки зрения.

Тщательный анализ материала показал, что результаты, удовлетворяющие поставленным требованиям, имелись в петрологии, седиментологии и геохимии. Действительно, в петрологии, как неоднократно отмечал Е. Х. Т. Уиттен, были открыты совершенно неожиданные свойства структур, имеющие глубокий генетический смысл, развиты новые феноменологические схемы метасоматоза, решившие ряд конкретных задач, созданы специфические методы картирования интрузивных массивов, вскрывающие их черты, не выявляемые обычными методами геологии, важные для решения не только научных, но и производственных задач. В результате детального анализа слоистых структур в седиментологии были внесены принципиальные изменения в концепцию циклического осадконакопления таких толщ, как молассы и флиш. Разработаны

методы, способные значительно уточнить представление об осадочных формациях. Наконец, в геохимии удалось уточнить понятия, связанные с таким ее основным разделом, как учение о средних концентрациях, сильно углубить понимание термина «распределение частот» и развить методы анализа парагенезисов элементов.

Все эти результаты создали реальную основу, на которой возможно построение основ математической геологии. Кроме отмеченных петрологии, седиментологии и геохимии в математической геологии должны найти отражение проблемы тектоники, палеонтологии и биостратиграфии, но пока удовлетворительных разработок этих проблем нет.

Успехи в отмеченных задачах были достигнуты потому, что для их решения был найден адекватный им математический аппарат. Без знания этого аппарата невозможно оценить полученные результаты и развивать их далее. Аппарат этот носит специальный характер и требует знакомства с рядом разделов математики. Для того чтобы работать с этим аппаратом и получать конкретные результаты, нужна не популяризация аппарата, а его систематическое изложение.

Таким образом, книга должна, с одной стороны, дать математический аппарат, а с другой — показать, что существенного для геологии получено с помощью этого аппарата. Для достижения указанной цели книгу пришлось разделить на две части. В первой части, предлагаемой ныне читателю, даны определение математической геологии и математический аппарат, на который опираются результаты, излагаемые во второй части.

Первая часть монографии делится на ряд глав. Первая глава посвящена определению предмета математической геологии. Хотя это определение в последние годы давалось и фигурирует в официальных справочных документах, вернуться к нему совершенно необходимо. Мы живем в период становления математической геологии и выработки на ее основе совершенно нового подхода к анализу геологических явлений. Специалистов в этой области очень мало, она бесконечно перегружена лицами, весьма далекими или от геологии, или от математики, иногда от них обеих, а подчас и от науки вообще. В этих условиях естествен хаос, существующий ныне в литературе по математической геологии. Хаос этот крайне отрицательно сказывается на развитии математической геологии и геологических наук в целом. Однако бороться с ошибочными тенденциями нужно не административными путями, а показом тех направлений, которые дают реальные геологические результаты. Для того чтобы найти эти направления, необходимо представлять, что такое математическая геология, почему она возникла и для чего нужна. Автор по мере сил стремился выяснить это в первой главе.

В результате анализа специфических особенностей геологических явлений в первой главе был сделан вывод, что для опи-

сания этих явлений исключительно большое значение имеет понятие «вероятность». Но для того чтобы получать с помощью этого понятия нетривиальные результаты, нужно знать его точное определение, важнейшие свойства вероятности и те методы, которые позволяют изучать эти свойства. Таким образом, геологу для занятий математической геологией нужно достаточно отчетливо представлять существо теории вероятностей и уметь использовать ее методы. Этим вопросам посвящена вторая глава. Она содержит сжатое изложение современной теории вероятностей. Особое внимание уделено наиболее важным для геологии проблемам, и в первую очередь характеристике случайных процессов. При изложении материала затронут более широкий круг вопросов, чем тот, который будет использован во второй части. Объясняется это тем, что сейчас невозможно точно предсказать детали развития математической геологии. Вместе с тем ясно, что использование теории, скажем, непрерывных случайных процессов — это вопрос только времени. Таким образом, вторая глава достаточно широко освещает главную массу фактов, охватываемых теорией вероятностей. При изложении автор стремился показать специфику методов, характерные постановки вопросов, а иногда и те достаточно тонкие моменты математического характера, которые могут возникнуть при построении моделей процессов. Так, автор коснулся некоторых свойств интеграла Лебега, которые, конечно, на практике могут быть обойдены, однако их показ демонстрирует характер точных математических построений. Кроме того, как уже упоминалось, эти деликатные математические вопросы могут возникнуть в будущем при разработке моделей процессов. Стремясь оживить по необходимости очень сжатое изложение, мы вводили примеры. Так, достаточно подробно был рассмотрен вопрос о преобразовании векторных случайных величин (процентный пересчет), который имеет большое значение в геохимии и петрологии, но который пока нигде с достаточной точностью не излагался.

Во второй главе показано, как на специфическом математическом языке выразить те или иные представления геолога о геологическом явлении, показан наиболее эффективный язык для изложения теории.

Однако на практике геологу, как и всякому естествоиспытателю, необходимо проверять по наблюдениям те теории, которые были им получены ранее. Это решается специальной математической наукой — математической статистикой. Она так же важна для геолога, как и теория вероятностей. Математическая статистика — обширная научная дисциплина, фрагменты которой уже публиковались в связи с постановкой специальных геологических задач, поэтому вряд ли рационально излагать математическую статистику в том плане, в каком излагалась теория вероятностей. Более целесообразно, кратко определив ее задачи, показать, как ее нужно применять на практике. Так, например, во всех геоло-

гических науках широко используются различные коэффициенты. Поэтому было бы рационально показать, каково соотношение между теоретически определенным коэффициентом и его аналогом, полученным из наблюдений (точечное оценивание), как изучаются вновь вводимые коэффициенты, какова связь между информацией, содержащейся во всех имеющихся в распоряжении исследователя наблюдениях, и коэффициентом, призванным передать эту информацию в сжатой форме. Наконец, следовало рассмотреть наиболее типичные методы проверки гипотез и на конкретном материале показать, как проверяется статистическая гипотеза. Требовалось также привести пример для иллюстрации того, как строится заключение о том или ином конкретном геологическом построении. Все эти вопросы в плане определений и демонстрации разработок даны в третьей главе. Эта глава должна показать, как те или иные теоретические построения проверяются через наблюдения и эксперимент.

Закончив изложение основ общей теории и показав, как она используется при операциях с наблюдениями, перейдем к систематическому изложению того аппарата, с помощью которого были получены важнейшие для математической геологии результаты. Этот аппарат — теория случайных последовательностей — в сколько-нибудь законченном виде в том плане, в каком он нужен геологам, нигде и никогда не излагался. В нашей книге его изложение составляет содержание четвертой, пятой и шестой глав. В этих главах не только ставится вопрос, но и подробно излагается его решение.

Четвертая глава, с которой начинается эта специфическая группа глав, содержит описание случайных последовательностей. Особенное внимание уделено в ней цепям Маркова и связанным с ними понятиям, из которых некоторые появились в результате обобщения опыта геологических исследований. Большое внимание уделено разбору техники анализа случайных последовательностей, в связи с чем подробно рассматриваются эффективные в этом случае матричные методы. Излагая те или иные вопросы, мы стремились их подробно иллюстрировать. Примеры подбирались так, чтобы вызвать интерес у специалистов тех разделов геологических наук, где развиваемые нами методы еще не использовались. Так, приводятся примеры из теории формирования нефтяных залежей и анализа соотношений в структурах силикатов. Нам кажется, что вопрос о внедрении стохастического моделирования в эти области достаточно назрел и не разрабатывается лишь из-за незнания того, что в математике уже несколько десятилетий существует весьма эффективный аппарат, способный дать минералогу важнейшие результаты (скажем, при изучении смешаннослойных минералов глин или минералогии слюд).

Пятая глава охватывает ряд вопросов о преобразованиях марковских цепей. В принципе вопрос в ней ставится так: анализ модели геологического явления показывает, что его реализация

есть цепь Маркова. Как изменятся свойства цепи, если ее реализации будут подвергнуты некоторому преобразованию? Например, какие-то состояния (скажем, зерна минералов, слои и т. п.) мы перестанем отличать друг от друга, какие-то состояния вообще исключим и т. п. Почти весь материал этой главы оригинален и полностью нигде ранее не публиковался. Все определения, теоремы и леммы, фигурирующие в этой главе, будут использованы при построении и анализе моделей.

Последняя — шестая — глава содержит сведения о статистике марковских цепей для тех моделей, которые будут использованы в дальнейшем. Материал этой главы в основном заимствован из более ранних публикаций.

В целом четвертая, пятая и шестая главы могут рассматриваться как сравнительно полное изложение теории марковских цепей с конечным числом состояний, с акцентом на те случаи, которые оказались важными в разработанных задачах геологии.

Текст книги иллюстрирован большим числом примеров. Эти примеры делятся на две категории. Одни из них — наиболее многочисленные — имеют своей целью пояснить то или иное определение, показать его специфический аспект на геологическом материале. Примеры второго типа введены для того, чтобы показать, с какого типа исследованиями — математическими и геологическими — сталкивается специалист, занимающийся математической геологией. Это по-существу как бы отдельные «статьи», включенные в текст с отмеченной целью. Они выделены специальными заголовками и имеют самостоятельное значение. Автор стремился к тому, чтобы в каждой главе был хотя бы один подобный пример. Между выходом первой части этой книги и публикацией второй неизбежен временной разрыв. Примеры в этой ситуации могут служить для разбора используемой техники до того, как будут опубликованы модели, дающие подробное изложение материала.

Из сказанного видно, что публикуемая ныне часть монографии является подсобной — это как бы методическое введение ко второй части.

При изложении материала автор руководствовался основным соображением — показать специалистам методы, позволяющие с помощью математики достигать реальных результатов, методы, которые могут привести к результатам, неизвестным ранее геологической науке и важным для ее развития. В соответствии со сказанным ясно, что книга не претендует на популярность излагаемого в ней материала.

В заключение хочется отметить одно обстоятельство — чтобы разобраться во всем изложенном в книге, требуется известная математическая подготовка и небанальное представление о путях развития геологии. Без понимания того, что геология нуждается в создании собственной математической дисциплины, читать эту книгу бессмысленно. Нам кажется, что ее читателями должны быть лица,

специализирующиеся по математической геологии; в равной степени она может быть отнесена сотрудиникам ныне многочисленных математических учреждений различных геологических организаций — от лабораторий математической геологии до вычислительных центров. Лучший способ чтения книги — знакомство с текстом одновременно двух специалистов: геолога и математика. Первая глава доступна геологу без каких-либо математических знаний. Она полезна всем тем, кто хочет понять в принципе, в чем задачи и методы математической геологии, каково ее положение в цикле геологических наук. Просмотр этими лицами других глав (без разбора математики) покажет, что в геологии существует множество задач, ждущих разработки и обещающих интересные решения, если исследователь сможет переформулировать их в математических терминах.

Что касается математики, то при изложении новых результатов доказательства приводились полностью, часто с сохранением промежуточных выкладок. Утверждения, заимствованные из литературы, иногда давались без доказательств.

Учитывая, что читателем книги может оказаться геолог, не сталкивавшийся ранее с математикой, или математик, мало знакомый с геологией, автор стремился дать достаточно полные библиографические ссылки, указывая читателю не только источник, но и страницы. Списки используемых работ даются по главам.

Итак, предположим, что читатель хочет получить в геологии нетривиальный результат с помощью математики. Он готов потратить время на то, чтобы выяснить, каким аппаратом для этого он должен овладеть. Мы надеемся, что, штудировав эту книгу, читатель заметно приблизится к поставленной им цели.

Глава I

МАТЕМАТИЧЕСКАЯ ГЕОЛОГИЯ И РАЗВИТИЕ ГЕОЛОГИЧЕСКИХ НАУК

Определение математической геологии. Анализ причин, вызывающих смену парадигмов в развитии геологии. Вопросы организации среды, реализации геологических явлений и структуры объектов, изучаемых геологией. Типы задач. Уточнение понятия «концептуальная вероятностная модель». Вопросы развития математической геологии.

Ключевые слова:

математическая геология, концептуальная вероятностная модель, организация среды, типичные структуры, парадигмы.

I.1. ВВЕДЕНИЕ

В самом общем виде математическая геология может быть определена как наука, занимающаяся построением, анализом и использованием в конкретной работе математических моделей геологических явлений. В конце главы мы уточним это определение, но пока ограничимся приведенным для того, чтобы не вводить некоторых понятий, относительно новых для геологов.

Вопрос о выделении научной дисциплины, которая решала бы задачи геологии математическим путем, насколько известно автору, был поставлен впервые около 35 лет тому назад (Вистелиус, 1944). Сама эта дисциплина, недостаточно точно определенная, была названа аналитической геологией. Ее появление было поддержано В. И. Вернадским, а затем английским журналом «Природа» (Tomkeieff, 1947). В начале 60-х годов по ряду причин автор заменил название «аналитическая геология» названием «математическая геология». Определение математической геологии и ее специфических задач было дано в 1962 г. (Вистелиус, 1962). Эта статья, вышедшая затем в переводах во Франции, США и ГДР, содержала развернутое изложение лекции, прочитанной ее автором весной 1962 г. в Институте геологии и геофизики СО АН СССР по предложению руководства этого института.

В 1967 г. по инициативе Р. Реймента (Швеция) был сформирован неофициальный международный комитет для создания международной ассоциации по математическим методам в геологии. Комитет принял решение называть новую ассоциацию Ассоциацией математической геологии. В 1968 г. ассоциация была официально утверждена XXIII сессией Международного геологического конгресса в Праге. В то же время было дано первое развернутое определение предмета математической геологии в СССР (Вистелиус, 1969; Vistelius, 1968). В 1969 г. был основан международный журнал математической геологии, название которого было определено после дискуссии. Оно было предложено в циркулярном письме президента вновь созданной ассоциации.

Официальные определения математической геологии были даны в СССР (Геологический словарь, 1973) и в США (Glossary, 1972), а также в адресе президента Ассоциации XXIV сессии Международного геологического конгресса (Монреаль, 1972; Vistelius, 1976). В настоящий момент определение математической геологии дано в целом ряде работ (Вистелиус, 1977; Vistelius, 1976; Whitten, 1977). Для того чтобы были ясны причины возникновения математической геологии, обратимся к истории геологических наук.

1.2. РАЗВИТИЕ ГЕОЛОГИИ И СМЕНА ПАРАДИГМОВ

В 1680 г. Лейбницом была предложена теория образования Земли, исходящая из ее огненно-жидкого начального состояния. Эта работа не оставила следа в развитии геологии, но за ней последовал примерно 100-летний период, в течение которого одно за другим появлялись построения такого же типа. Период этот закончился знаменитой публикацией Хаттона (Hutton, 1788), а затем комментариев к ней Плейфейера (Playfaier, 1802). Эти публикации завершили большой исторический период, в течение которого была проведена та подготовительная работа, после которой смогла появиться собственно геология как наука. Этот период был знаменателен близостью всех наук — математика обогащала геологию, а геология давала математике содержательные задачи.¹ Именно в этот период наметился колоссальный прогресс в области точных наук. Сложились основанные на детерминистическом принципе и понятии функции математические дисциплины. Зародился вероятностный подход к анализу случайных явлений. Возникли идеи эксперимента.

¹ При установлении участия представителей точных наук в решении геологических задач широко использованы ссылки, содержащиеся во фрагментах рукописи «Temperature of History» профессора Бруша (Brush), любезно присланных им автору.

Не имея возможности подробно остановиться на характеристике этого времени, отмечу некоторые его черты, особенно важные для наших целей.

1. Исследовательская мысль развивается чисто аксиоматическим путем на основе библейских представлений, являвшихся в то время парадигмом² естествознания. Для геологии принятие библейской аксиоматики означало, что вся геологическая история Земли должна быть уложена примерно в шесть тысяч лет. Принятие библейской догмы не позволяло ввести весьма продуктивный и естественный принцип актуализма. Как можно пользоваться принципом актуализма, писал виднейший геолог того времени, президент Ирландской Академии наук Кирван (Kirwan, 1799), когда через 14 дней после прекращения подъема вод, вызванного потопом, голубь принес Ною масличную ветвь? Ведь для падения уровня вод от отметки вершины Арарата до отметки, на которой растут оливы, за такой короткий срок нужно было, чтобы воды двигались со скоростью, большей, чем в любом водопаде. Но как же тогда могла сохраниться олива? Так как священное писание непогрешимо, то, значит, было совершенно чудо, а в этих условиях бессмысленно прибегать к униформизму. Мы уже отмечали (Vistelius, 1976), что аналогичная аргументация использовалась Геснером (Gesneri, 1758) при анализе скорости тектонических движений.

2. Появляются первые попытки использовать достижения точных наук для проверки реальности геологических построений. Нельзя сказать, чтобы эти попытки были удачны. Сначала И. Ньютон, а затем П. С. Лаплас категорически заявляют, что никаких изменений положения земной оси за геологическое время не происходило. На основе этих заключений «бракуется» одна из очередных теорий Земли (Барнета; цит. по: Лайель, 1866). Однако окончательного решения вопроса нет до настоящего времени, так как нет ясности в том, как воздействовали на планету внешние силы в течение ее существования.

3. Как отмечалось, в тот же период возникает мысль об использовании эксперимента. Для определения периода, прошедшего со времени огненно-жидкого до современного состояния Земли, изучалось охлаждение раскаленных ядер (Mairan, de, 1719; Buffon, de, 1774). Эксперименты показали, что Земля охладилась за 75 тыс. лет.

Однако, несмотря на обращение к математике и экспериментам, геология по существу не прогрессировала, будучи замкнутой библейским парадигмом. Так, догма, активно поддерживавшаяся властью и тяготевшая над умами даже крупнейших ученых, фак-

² Под термином «парадигм» понимается совокупность взглядов, принявшаяся за очевидную истину в том или ином разделе науки в течение определенного периода ее развития. Смена парадигмов приводит к научным революциям (Kuhn, 1970).

тически остановила прогресс важнейшей для развития производительных сил научной дисциплины.

В обстановке догматического бесплодия выходит уже цитированная нами работа Хаттона (Hutton, 1788). В этой работе, написанной, как и многие современные ему произведения, с глубоким религиозным чувством, отмечается, что для изучения геологии нет необходимости обязательно искать начало геологической истории или стремиться выяснить время наступления ее конца. Можно заниматься наблюдениями над самими геологическими объектами и по этим наблюдениям делать заключения о том, как они формировались в соответствующий момент времени или что произошло с ними позже. При этом естественно сравнивать то, что мы видим в геологических разрезах, с тем, что мы можем наблюдать, изучая современные геологические процессы. В подтверждение своих идей Хаттон исследует о-в Эран и разрезы в Шотландии³ и делает заключения, полные, с современной точки зрения, элементарного здравого смысла. Хаттона поддерживают физико-химические (в современной терминологии) построения Холла (Hall, 1794). Примерно в то же время Смит вообще без всякой теории, а путем прямых полевых наблюдений ранжирует слои, различающиеся по ископаемым остаткам. Иными словами, он показывает, что можно учитывать относительный возраст пород не в годах, а в некоторой условной шкале смены вышерших организмов.

Работа Смита, положившая начало современной стратиграфии, по-видимому, в первый момент вообще не обратила на себя внимания лиц, возглавлявших геологию, и не удостоилась их критики, так как в ней не было общих рассуждений. В первой редакции 1790 г. это была просто таблица (ныне она хранится в Лондонском геологическом обществе), огромное значение которой в то время было трудно предугадать. Такова же судьба исследований Холла, признанного ныне отцом экспериментальной петрографии. Другое дело развернутая концепция Хаттона, бросавшая вызов Вернеру и его общепринятым взглядам (Севастьянов, 1810), отлично отвечавшим существовавшей догме. Дальнейшее развитие событий носило трагический характер. Р. Кирван⁴ (Kirwan,

³ Следует отметить, что в работе 1788 г. Хаттон дал только основные положения своей теории. Как можно понять из текста его работы 1795 г., он хотел изложить эту теорию в четырех томах. Из этих четырех томов два были опубликованы в 1795 г., третий был найден в архивах Лондонского геологического общества и опубликован под редакцией А. Гейки в 1899 г.; о четвертом томе ничего не известно.

⁴ Р. Кирван — один из энергичнейших и влиятельнейших деятелей британской и вообще европейской геологии конца XVIII в. Пользовался огромной известностью. Иезуит по воспитанию, не имевший специального образования ни по геологии, ни по химии, он начал заниматься ими в возрасте более 33 лет. Публиковал работы, которые ныне могли бы быть отнесены к геохимии. Ревностный эпигон учения Вернера (Kirwan, 1799) и энергичный защитник представлений о флогистоне (Kirwan, 1788). Энергия Кирвана, по-видимому,

1799) — президент Академии наук и член многих иностранных академий — и Уильямс (Williams, 1789) — начальник горных работ Шотландии (в современной терминологии — министр геологии) — со всей энергией эпигонов-администраторов обрушиваются на учение Хаттона. Хаттон умирает до завершения публикации своего основного труда. Взгляды Хаттона защищаются Дж. Холлом (Hall, 1794), а затем Плейфейером (Playfair, 1802). Разгорается спор между нептунистами и плутонистами и, наконец, наступает полное крушение вернерианства, а следовательно, и построений Кирвана. Обычно этот спор рассматривается, как схоластическая риторика. Однако автору кажется, что он имел большее позитивное значение, так как способствовал выработке умения наблюдать природу и делать непосредственные умозаключения из этих наблюдений. Получают должную оценку и работы Смита. Шаг за шагом укрепляется представление о ценности фактов. Постепенно принимаются взгляды, что только описание фактов и их прямая интерпретация являются настоящей наукой. Дедуктивные выводы, не опирающиеся на прямые, желательны полевые, наблюдения резко порицаются (например, Sadgwick, Murchison, 1840). Во главе этого течения становится Лондонское геологическое общество с такими выдающимися учеными, как Седжвик и Марчиссон, а позже Лайель.

Вырабатывается описательный парадигм. Работы на его основе выявляют чисто опытным путем специфику изучаемого геологами материала. Эта специфика показывает, что выводы о свойствах изучаемых объектов можно делать только после исследования множества их представителей; так, для суждения о возрасте слоев пород нужны обильные сборы фауны, состав же типичной минеральной ассоциации выявляется изучением большого числа штUFFов. Отличный пример такого типа работы был дан еще Лайелем при разработке стратиграфии Парижского бассейна (Lyell, 1833). Такой подход отлично увязывается с методом, использованным позже Дарвином в происхождении видов.

В тот же период выясняется, что хотя заключение может быть сделано о том или ином явлении природы только на основе множества наблюдений, тем не менее несомненно существуют устойчивые типы структур. Разнообразие геологических объектов неограничено, и они подразделяются на устойчивые классы. Такие классы выделяются во всех геологических науках. К ним принадлежат устойчивые парагенезисы (скажем, ассоциации пегматитов, скарнов, фумаролл и т. п.), биоценозы, комплексы осадочных образований, дающие формации, и т. п. В каждом классе объекты индивидуальны и разнообразны, но сами классы с учетом разнообразия входящих в них индивидов устойчивы.

Влияние описательного парадигма ясно ощущается во всей геологии вплоть до нашего времени. Однако его абсолютное господство задержало развитие геологии, хотя его компиляция по минералогии сыграла положительную роль в нашей науке.

подство знаменательно для первой половины XIX в. Интересно, что дедуктивные построения в геологии, потерявшие интерес для самих геологов того времени, привлекали в то время математиков. Вопросом скорости охлаждения земного шара интересовался Ж. Фурье, собиравшийся таким способом рассчитать возраст Земли (Fourier, 1819, 1827). В итоге им была получена цифра, близкая к 200 млн. лет. Эта величина показалась ему обескураживающей; по его мнению, никакие геологические события не могли стимулироваться столь медленным процессом. В дальнейшем Фурье к геологическим задачам не возвращался. Это случай, когда задачи геологии были объектом длительной, систематической разработки одного из величайших математиков нового времени.

Накопление огромного числа описаний геологических объектов и совершенная недостаточность их прямой интерпретации без дедуктивных построений возродили в геологии интерес к проблеме происхождения геологических объектов. Одновременно возрастал интерес и к вопросу о возрасте Земли. Острота этой проблемы снова привлекает к ней внимание представителей точных наук. В 1868 г. Томсон (лорд Кельвин) докладывает в Геологическом обществе в Глазго результаты своих расчетов возраста Земли, снова, как и у Фурье, основывающиеся на скорости охлаждения Земли от температуры огненно-жидкого тела до современного состояния. Лорд Кельвин получает цифру порядка 100 млн. лет (Thomson, 1868). Однако к тому времени на основе работ Дарвина уже сложилось представление о том, что возраст Земли значительно больше. Хаксли призывает к осторожности (Huxley, 1869), однако Кельвин занимает поучительную позицию и обвиняет геологов в незнании современной науки, выражающемся в игнорировании достижений физики и термодинамики. Его поддерживает Тейт, заявляющий, что не обязательно учитывать специфику той или иной науки, так как совершенство различных наук различно: есть более фундаментальные и менее основательные. Нужно прежде всего помнить, что математика не оценима для полноты развития любой настоящей науки (Tait, 1869).

Отметим, что цифра возраста, приведенная лордом Кельвином, фигурировала в геологии до первой половины XX в. Казалось бы, что современная геохронология должна была бы убедить геологов, что они должны верить своей интуиции больше, чем любым точным методам, если нет уверенности, что эти методы опираются в данной задаче на проверенную аксиоматику и что такая аксиоматика может быть построена. Именно этого не было в распоряжении лорда Кельвина, а до этого — Ж. Фурье.

Число проблем, требующих объяснения, с середины прошлого столетия стремительно росло. Объяснение происхождения эратических валунов и остатков теплолюбивых животных в арктических областях, по-видимому, впервые привело к трудностям. При этом появлялись и продолжали существовать диаметрально

противоположные взгляды, скажем, на отложение валунов (Лайель, 1866; Кропоткин, 1876). Генетические вопросы все более и более проникали во все области геологии. Латеральная секреция (Sandberger, 1882) и гидротермальные ювенильные воды (Lindgren, 1927; Emmons, 1927), ликвационное образование магматических пород⁵ (Дели, 1936; Dugocher, 1857) и возникновение их за счет гравитационного разделения в процессе фракционной кристаллизации (Боуэн, 1934), пегматиты как остаточные эвтектические расплавы (Niggli, 1920) и как продукты замещения (Landes, 1925). На наших глазах в одном и том же институте один крупнейший советский нефтяник приходил к выводу, что нефть — продукт переработки водородослей на отменях при полном отсутствии миграции материала (Калицкий, 1944), а другой — что нефть магматическое образование, свободно мигрирующее на большие расстояния (Кудрявцев, 1973). В конце концов все эти задачи начинают решаться по схеме, в общем виде сформулированной еще Бюффоном. Согласно Бюффону, цель науки — описание объекта и объяснение его истории (Buffon, 1785). Это положение постепенно вытесняет описательный парадигм и само становится парадигмой геологии, дошедшим до наших дней.

Исследование возраста Земли, проведенное лордом Кельвином, было, видимо, последней попыткой крупных представителей точных наук решать сколько-нибудь общие задачи геологии. С середины XIX в. начинается бурное развитие математики. После Кантора большое число математиков было поглощено устранением противоречий в собственной науке. Физики обращаются к фундаментальным проблемам строения вещества. Одновременно в физике растет понимание важности вероятностного подхода к анализу экспериментальных фактов, который к настоящему времени оснащен мощным техническим аппаратом. Появляются и становятся доступными ЭВМ.

В этих условиях в геологических науках появляется тенденция к внедрению в них техники и методов точных наук со стороны самих деятелей геологии. Наиболее полно это проявляется в использовании термодинамики, физико-химического эксперимента и в применении современной физической аппаратуры. Это направление берет начало с впечатляющих работ Гольдшмидта (Goldschmidt, 1912), Боуэна (Bowen, 1912) и Ниггли (Niggli, 1920). Именно отмеченные области точного знания глубоко вошли в настоящее время в геологические науки. Однако в то же время остается неясным — нет ли резервов внутри самой геологии для того, чтобы на их основе попытаться перестроить эту науку на математической основе.

Для того чтобы выяснить вопрос о резервах, посмотрим, чему учит нас история геологии и изучение тех или иных частных объ-

⁵ Новейшие эксперименты снова привлекают внимание к проблеме несовместимости расплавов (Naslund, 1976).

ектов, имеющих народнохозяйственный или другой специальный интерес.

1. Аксиоматический метод является ведущим методом в современных точных науках. Этот метод широко применялся в геологии по крайней мере в течение 120 лет. Однако его применение дало отрицательные результаты. В итоге аксиоматический метод и тесно связанный с ним дедуктивный метод были дискредитированы и около 150 лет в явном виде в геологии фактически не применялись. Анализ исторических фактов показывает, что крушение аксиоматического метода произошло не из-за внутренних недостатков метода, а вследствие отсутствия в то время аксиоматики, адекватной сущности явлений, а также отсутствия развитого аппарата для дедуктивных построений, адекватного существованию геологических явлений.

2. Индуктивный метод, использовавшийся в описательном плане почти 150 лет, показал, что он не в состоянии дать непротиворечивую конструктивную схему геологических явлений. Систематическое применение этого метода привело к возникновению специфических требований, предъявляемых к материалу наблюдений в геологии; единичное наблюдение, как правило, имеет малое значение; для далеко идущих выводов нужно иметь много наблюдений. При этом не следует смущаться тем, что результаты некоторых наблюдений очень далеко уклоняются от того, что мы в конечном счете признаем за типичное.

3. В геологической науке почти не проявляется интерес к тому, чем отличаются объекты, изучаемые геологией, от объектов, изучаемых другими науками. В геологических науках до сих пор нет разработанного представления о том, какова должна быть техника для построения заключения о происхождении изучаемых объектов.

4. В геологии очень высоко ценились идеи, подававшиеся геологам представителями точных наук, особенно математиками. Существует даже понятие «это математически доказано», что воспринимается как адекватное понятию «это утверждение истинно». Между тем утверждение это «математически доказано» в геологических науках и вообще в естествознании имеет не тот смысл, который ему обычно придается. В этих науках математически доказуемых истин не существует. В них выражение «математически доказано» осмысленно, если под ним понимается безошибочность дедукции из формализованных представлений. Если аксиоматика, подвергшаяся формализации, адекватна существованию явления, то математические выводы, полученные дедукцией из этой формализации, отражают суть явления. Если формализация не адекватна аксиоматике или аксиоматика не отражает специфики явления, то самая безукоризненная математическая дедукция ничего, кроме ошибок, дать не сможет. И. Ньютон и П. С. Лаплас, Ж. Фурье и лорд Кельвин пытались решать задачи геологии. Однако все их попытки оказались тщетными. Причина этого

ясна — задачи решались при игнорировании геологических данных. Вследствие этого принятая аксиоматика не отвечала существу вопроса, а поэтому даже безукоризненная математическая дедукция ничего не могла дать. В настоящее время это ясно не только геологам, но и всем естественникам вообще. Поэтому следует возлагать надежды не на «математическую обработку наблюдений», а на создание специфических научных дисциплин, скажем, на границе математики и геологии.

Эта книга должна содержать основы математической дисциплины в цикле геологических наук. Такая дисциплина будет иметь право на существование, если в конкретных задачах она будет вскрывать больше, чем это позволяют существующие в геологии методы. Новая дисциплина должна эффективно помогать установлению закономерностей в геологических явлениях, так как поиски закономерностей являются важнейшей частью геологических исследований. В связи со сказанным необходимо рассмотреть, чем же являются геологические объекты и каковы характерные для них структуры.

1.3. ОРГАНИЗАЦИЯ СРЕДЫ И ТИПИЧНЫЕ СТРУКТУРЫ

Одна из характерных черт геологии — обилие точек зрения на особенности происхождения того или иного объекта. Много времени и сил было потрачено на выяснение, в каком направлении действуют силы, формирующие тектонические структуры. Более 150 лет продолжают бурные дискуссии по вопросу о происхождении гранита. Источником вод Земли одни считают мантию, другие — кору.

Однако, как правило, все дискуссии велись и ведутся по частным вопросам. Вопрос же о том, как в целом организована среда, в которой протекают геологические процессы, оказался практически не разработанным. Между тем без внесения ясности в этот вопрос невозможно корректно решать частные задачи, интересующие геологов.

Обращаясь к вопросу организации среды с точки зрения геологии, нужно в первую очередь определить, с чего начинается специфика геологии. Как кажется автору, огромное число локальных процессов в земной коре проходит без всякой геологической спецификации. Они протекают по законам физики, химии или физической химии. Характернейшим из этих процессов является физико-химическое равновесие. Правило фаз для объектов, где оно приложимо или не замаскировано другими явлениями, реализуется совершенно одинаково в лаборатории, в мантии Земли и в доменной печи. Во всех случаях соблюдаются соотношения между компонентами, фазами и степенями свободы, обнаруженные В. Гиббсом. Однако, как было ярко оттенено Д. С. Коржинским (1962), равновесия реализуются локальными процессами в отдельных точках. Вся же среда в целом представляет собой мозаику

локальных равновесий. Выяснение же того, чем управляется эта мозаика и каковы свойственные ей тенденции развития, и является задачей геологической науки. Установив поведение объекта в элементарном объеме с помощью той или иной точной науки, необходимо показать, как сочетаются объемы, в которых протекали локальные процессы.

Аналогичное положение и в седиментологии. Здесь задача снова расчлениется на задачу о локальном поведении, скажем, движении частицы в определенной среде в гравитационном поле и на задачу об ее интегральном поведении. Как сочетаются частицы, почему здесь появляется та или иная среда, чем объясняются вариации среды?

Локальная задача — задача использования в геологии различных точных наук. Интегральная задача — задача собственно геологии. Точные науки сами по себе не решают геологических задач.

Если обратиться к состоянию поставленных вопросов, то окажется, что по вопросу о том, как протекают локальные процессы, сделано достаточно много. Одна физико-химическая петрография дает впечатляющую картину. Однако где проходит граница между локальными и интегральными процессами, известно очень мало. Еще меньше определенности в вопросе об интегральных процессах. Часто результаты, полученные о локальных процессах, без достаточных оснований распространяются на интегральные. Иногда вообще вопрос о точном описании явлений не ставится. Иными словами, эта важнейшая проблема требует постановки и изучения.

Насколько известно автору, единственным геологом, много раз ставившим вопрос об организации среды, был В. И. Вернадский (1926, 1931, 1965). При этом, отдавая дань традиционному направлению XIX в., он занимал вначале твердую детерминистическую позицию. Организация среды, по В. И. Вернадскому, определялась термином «механизм», который в принципе действовал, якобы, с той же точностью, что и механизм часов. Однако со временем В. И. Вернадский пришел к мысли, что представление о такой организации среды неадекватно природным явлениям. Эти явления протекают по значительно менее определенной схеме. Придя к такому выводу, В. И. Вернадский ввел свой термин «организация». Смысл этого термина очень близок к понятию сильно зависимых случайных событий или случайных величин. Иными словами, В. И. Вернадский определил организацию среды как организацию вероятностного типа. Этот вывод В. И. Вернадского резюмирует длинный путь постепенного осмысливания того, с каким объектом в отношении организации имеет дело геолог. Сначала геологи безо всякой теории использовали статистический принцип, стремясь обосновать выводы возможно более обильными наблюдениями. Уже Лайель (Lyell, 1833), как это отметил Р. Фишер (Fisher, 1953), изучая Парижский бассейн, действовал как статистик.

Однако в выводах и в изложении представлений о ходе процесса геолог, как правило, детерминист. Таким образом, возникает внутреннее противоречие между действиями при сборе материала (полевого и лабораторного), интуитивно и часто неосознанно отражающими сущность явлений, и детерминизмом выводов, навеянным поверхностной образцованностью, аппелирующей к общеизвестной школьной математике. Так или иначе, но специфика сбора материала — в принципе статистический подход — постепенно порождает желание количественно исследовать колеблемость в проявлении наблюдаемых объектов.

Видимо, одним из первых геологов, понявших необходимость количественного учета колеблемости изучаемых геологией признаков, был Рейер (Reyer, 1877). Затем появились работы Харкера (Harker, 1909), Ричардсона и Снизби (Richardson, Sneesby, 1922), Ниггли (Niggli, 1923), Ф. Ю. Левинсон-Лессинга (1924) и, наконец, Крамбайна (Krambein, 1934) и Н. К. Разумовского (1940). Эта линия исследования наметила пути количественного изучения колеблемости. Она подготовила почву для дальнейших шагов по изучению вероятностных закономерностей.

Работы по количественному изучению изменчивости показали огромную важность для решения задач геологии понятия «вероятность». Устойчиво не то или иное значение исследуемого объекта, а вероятность, с которой появляется это значение. При этом особое значение для практики, скажем, принятия решения, имеют большие вероятности, практически гарантирующие появление объекта, и малые, позволяющие игнорировать это появление (см. гл. III).

Введение в сферу исследований геолога вероятности вынуждает уточнить этот термин в познавательном плане (точное математическое определение его мы дадим в гл. II). При этом следует отметить, что смысл этого понятия трактуется двояко. Одними допускается, что в ряде ситуаций имеется такое разнообразие явлений, каждое из которых протекает детерминированно, что обрисовать действие каждой причины невозможно, нужно подходить к анализу причин в среднем. Такое понимание вероятности носит название эпистемологического и имеет прагматический характер. Однако другими специалистами высказывается мнение, что вероятность присуща явлению. Это не результат нашей ограниченности временем и неспособности рассмотреть детали, а сущность явления, не сводящегося к совокупности детерминированных действий, вызываемых известными причинами. Такое понимание вероятности носит название онтологического. Однако в геологии еще не накоплен материал, позволяющий занять ту или иную точку зрения на сущность вероятности. Мы находимся в такой стадии развития науки, что можем говорить о продуктивности описания явлений на вероятностном языке. Таким образом, вероятность в нашем понимании есть форма описания явления. На данной стадии для математической геологии этого нам кажется достаточно. Говоря, что среда, в которой протекают геологические явления, имеет сто-

хастическую организацию, мы будем подразумевать, что эта среда хорошо описывается языком наук, в основе которых лежит формальное понятие вероятности.

Привлечение вероятностного языка позволяет подойти к выработке принципов для поисков закономерностей в геологических явлениях, а нахождение закономерностей является основной задачей геологии при любой постановке вопроса. Обращаясь к задачам собственно геологии, нужно подчеркнуть, что для геологии, как для каждой естественной науки, цель исследования — выявление того, как протекает процесс, породивший данный объект. Изучая объект — биогермы или лакколлиты, вицинальные грани или спектральные яркости песков, — мы должны поставить вопрос, каковы закономерности, благодаря которым эти объекты возникли? Выявив соответствующие закономерности, мы можем их в дальнейшем использовать для поисков и разведки тех или иных конкретных объектов (месторождений). При такой постановке вопроса конкретный объект рассматривается как частное проявление общей закономерности. Общая закономерность определяет некоторые границы, в пределах которых могут флюктуировать характеристики конкретного объекта. Конкретный объект в частностях отклоняется от того, что дает закономерность. Закономерность — это среднее из большого числа частных проявлений. Частное проявление — единичная реализация закономерности. Принимая вероятностный характер организации среды, в которой протекают геологические процессы, мы должны ввести классификацию типов процессов, протекающих в соответствующей среде. При этом мы должны выделить два крупных класса процессов и отражающих их структуру объектов.

Первый класс образуют детерминированные процессы. Случайность в них не играет никакой роли. Эти процессы изучаются в основном точными науками. Классическим примером такого типа процессов является гетерогенное равновесие, протекающее по правилу фаз. В структуре объектов в чистом виде детерминированные процессы проявляются редко. Однако соответствующие приближения могут быть весьма полезны. Так, на рис. 1.1 (см. вкл.) приведена схема, показывающая в изолиниях, как изменяется возраст гранитоидных интрузий в Восточной Австралии в зависимости от географического положения интрузии. На схеме видно, что, начиная с кембрийского времени, докембрийский щит Австралии постепенно наращивался гранитными интрузиями по направлению к востоку. Происходило приращение гранитного вещества к древнему ядру континента. Картина эта получена расчетом тренда, т. е. путем выявления некоторой детерминированной составляющей.

Детальное изучение показывает, что стохастические явления очень важны для понимания структур кристаллов (проблема упорядоченности). Однако, когда речь идет о выявлении основных черт строения кристалла, случайные явления можно игнорировать.

При этом чисто детерминистический подход оказывается весьма продуктивным. Таким образом, есть случаи, когда имеет смысл пожертвовать частью информации для того, чтобы упростить решение. Детерминистический подход в таких случаях может быть оправдан.

Второй класс процессов (структур), с которыми имеет дело геология, носит стохастический характер. При этом выделяются два типа таких процессов (структур).

В первом типе знание предыстории процесса (характеристик структуры) и координат точки наблюдений не дает никакой информации о вероятностном поведении в данной точке. В этом случае можно говорить о стохастической структуре с независимыми элементами. Свойство стохастической независимости характеризует всю совокупность однородных объектов в целом и не может быть установлено для единичной пары объектов. Об этом мы будем говорить в следующих главах. Стохастическая независимость встречается очень редко. Обычно, но не всегда, за ней стоит причинная независимость. В качестве примера структур с независимыми элементами можно привести последовательность мощностей слоев во флише после исключения влияния состава слоя. В гранитных массивах в некоторых условиях последовательности зерен кварца, калиевого полевого шпата и плагиоклаза также образуют структуру с независимыми элементами (Вистелиус, Романова, 1976). В последнем случае появлению независимых элементов можно дать петрологические объяснения. Мы будем изучать такие структуры как последовательности независимых испытаний (последовательности Бернулли).

Второй тип стохастических структур образуют такие объекты, как, скажем, чередование слоев в разрезе моласс, где чередуются глины, пески и алевроиты; чередование зерен пироксенов, оливина и плагиоклаза в никеленосных трапзах и т. п.. В этих случаях элементы стохастической структуры (слои, зерна и т. п.) как бы полузависимы. С одной стороны, последовательность случайна, а с другой, — зная предыдущие элементы структуры, — можно делать некоторый прогноз о появлении последующих. Такой прогноз неточен (он носит характер вероятностного предсказания), но в длинной последовательности испытаний он в среднем оправдывается. Такие стохастические структуры мы будем называть структурами с зависимыми элементами. Одна из таких стохастических (случайных) структур в математике носит название марковской цепи. О такого типа структурах мы будем подробно говорить в IV, V и VI главах этой книги, некоторые общие характеристики дадим во второй главе.

На рис. 1.2 приведена реализация структуры с зависимыми элементами. Она реально наблюдалась в границах Центрального Казахстана. Расчеты продемонстрировали, что последовательности зерен в этой структуре вдоль каждой из линий U , M и L не от-

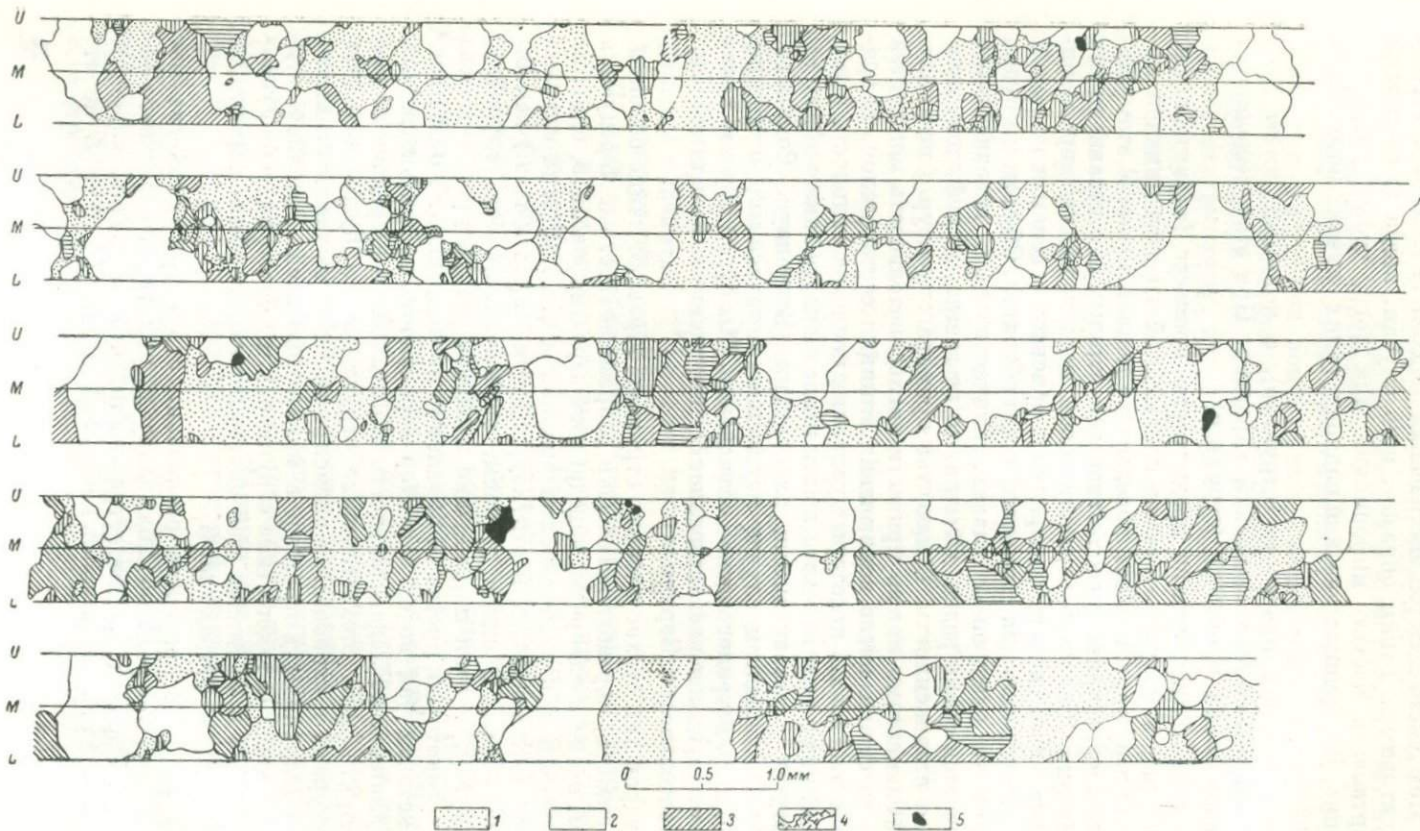


Рис. 1.2. Чередование зерен калиевого полевого шпата (1), кварца (2), плагиоклаза (3), биотита (4) и непрозрачных минералов (5) в мелкозернистом граните Кызыл-Тас (Казахстан; зарисовка Д. Н. Иванова).

Последовательность зерен минералов 1, 2, 3 не отличима от простой цепи Маркова вдоль любой из линий U, M или L.

личимы от цепи Маркова. Материал наблюдений показывает, что такие структуры характерны для гранитов магматического происхождения.

1.4. ПОСТАНОВКА ЗАДАЧ, РОЛЬ МОДЕЛИ В ПОИСКАХ РЕШЕНИЯ

Основная задача геологии — выяснение механизма процесса, породившего изучаемый ею объект, восстановление истории его формирования. Задача эта может решаться в двух принципиально различных постановках.

Если для решения задачи вводятся различные априорные схемы явления (модели), из которых одна принимается на основе проверки согласия этих схем (моделей) с наблюдениями (остальные бракуются), то говорят о постановке прямой задачи.

В геологии обычна другая постановка задачи, когда при изучении объекта на основе ряда признаков шаг за шагом воссоздается история его формирования. Так, если раковина изучаемого моллюска толстая, то мы говорим, что он жил в зоне интенсивного волнения, если раковина его тонка, то мы говорим, что моллюск жил в застойной зоне. Если минеральная ассоциация однородна, мы склонны интерпретировать ее как магматическое образование, неоднородность в каком-то смысле указывает на метасоматическое происхождение породы. Подобная постановка вопроса — наблюдение и его интерпретация — называется постановкой обратной задачи.

Обращаясь к современным публикациям по математическим методам в геологии, мы легко убеждаемся, что подавляющая часть их содержит постановку обратной задачи. Схема большинства таких публикаций состоит в следующем. Имеются наблюденные частоты (или какие-либо другие результаты наблюдений), по этим частотам строится некоторая гладкая функция (обычно идет речь о подборе функции распределения). Как правило, не пользуются никакой теорией, ограничивающей выбор таких функций. Просто в распоряжении исследователя находится некоторый набор ему известных функций, и одна из них после сравнения с наблюдениями по заданным критериям оказывается не противоречащей наблюдениям. Таким образом, исследователь «открывает „некоторую“ математическую закономерность».

Подобные исследования реализуются на основе описательного парадигма, характерного для первой половины XIX в. Здесь все происходит так, как, скажем, в классических трудах Седжвика или Марчиссона, — есть исходный, как думает исследователь, твердо установленный факт (в главе III этой книги будет показано, что этого на самом деле обычно нет). Без всяких «спекуляций» этот факт математически выражается. Это математическое выражение принимается за реально существующий закон, из чего делаются соответствующие выводы. Пример такого подхода дает представ-

ление от логнормального распределения как об основном законе геохимии (Ahrens, 1953).

Анализируя решение обратных задач в стиле, изложенном выше, мы отметили, что подобные решения ничем не отличаются от заключений геологов первой половины XIX в. Рассмотрим этот вопрос более подробно.

Значения характеристик, которыми мы оперируем и которые относятся к наблюдаемому объекту, от испытания к испытанию дают различающиеся между собой результаты без видимого изменения исходных условий. При этом результат каждого единичного наблюдения, будь то перидотитовый массив или структура конкретного рудного поля, характеризуется следующими чертами.

а) Измерения элементов изучаемого объекта неточны. В некоторых случаях эта неточность невелика, но иногда (скажем, при определении положения отражающего горизонта сейсмическими методами при неблагоприятных структуре и разрезе) может достигать значительной величины.

б) Измеряемые элементы в силу случайного механизма явления могут в конкретном случае очень сильно отклоняться от той типичной картины, которая дает представление о явлении.

в) Наблюдения выполнены над какой-то небольшой частью объекта (во всяком случае они относятся к конечному интервалу). Совершенно неясно, насколько свойства небольшого участка, даже при однородной структуре, могут быть распространены на весь объект, для которого развивается теория.

Следует, конечно, иметь в виду, что как бы тщательно не выполнялось описание наблюдаемого объекта, оно всегда искажает его действительный вид. Кроме того, при описании данной конкретной картины могут оказаться нефиксированными черты, крайне важные для понимания смысла явления. И, наконец, если даже по большому числу наблюдений нам удастся с большой точностью воспроизвести функцию распределения некоторой случайной величины, то эта функция распределения мало скажет нам о механизме явления. Дело в том, что эмпирические аналоги распределений в ряде случаев очень близки для совершенно различных распределений, а эти распределения могут порождаться различными механизмами, т. е. обратная задача будет решаться неоднозначно. Для того чтобы построить теорию, необходимо опираться не только на картину наблюдений и свой личный опыт, но и на всю совокупность данных.

Проследившая развитие некоторых геологических идей, можно оценить прямую и обратную постановку задач даже в тех случаях, когда авторам была совершенно чужда подобная терминология. Рассмотрим, например, проблему коннексии. Если обратиться к этой проблеме в аспекте сказанного, то будет ясно следующее. Де Геер, разрабатывая проблему коннексии четвертичных отложений, имел дело с прямой задачей. У ледника имеется озеро. Озеро замерзает зимой и оттаивает летом. Ежегодно образуются

два слоя — глинистый зимой и песчаный летом (варв). Последовательность слоев детерминированная — смена зимы и лета детерминированные события.

Методика де Геера была применена при сопоставлении разрезов флиша. Была взята внешняя сторона методики — коннексионные диаграммы. Постановка же вопроса, принципиальная сторона проблемы, независимо от авторов соответствующих работ, оказались совершенно отличными от прямой задачи в трактовке де Геера. Действительно, за сколько времени образуется флишевый ритм (аналог варва), не ясно. Не ясна и другая сторона проблемы: детерминирована или случайна последовательность слоев во флише? Раз строятся коннексионные диаграммы и вводится типизированный состав слоя, то, очевидно, разрез приводится к детерминированной последовательности слоев. Но как в этом случае расценивать выпадение отдельных слоев? Ведь размыв их не документируется! Почему кордильера порождает такие осадки, из которых при захоронении после переноса образуется трехчленный «ритм». Не разумнее ли отказаться от детерминистического происхождения ритмов, считать последовательность слоев стохастической, отбросить типизацию состава слоев и верить не схеме, а своим глазам, фиксирующим реальные составы пород. Иными словами, коннексионный подход при анализе флиша был реализован на основе обратной постановки задачи, но ясного понимания сути явления получить не удалось. Только работы Кьюнена (Kuenen, Migliorini, 1950), поставившие эту проблему в плоскости прямой задачи, внесли в нее ясность и послужили толчком для ее развития.

Изложенное показывает, что в принципе, когда это возможно, математика в геологию должна вводиться на основе прямой задачи. Математическая геология должна сконцентрироваться на решении прямых задач. Задача формулируется как представление о явлении, исходящем не только из личных наблюдений, но и на основе всего имеющегося по этому вопросу материала. По совокупности этих сведений создается некоторая математическая конструкция, которая называется концептуальной моделью. У этой конструкции фиксируются типичные черты (их совокупность называется статистической гипотезой), и эти черты сравниваются с наблюдениями. При этом, как мы отмечали, модель либо принимается, либо отбрасывается. Таким образом, все зависит от умения построить содержательную концептуальную модель. Остановимся на понятии «модель» подробнее.

Впервые идея модели в геологических науках, насколько нам известно, была выражена шотландским литологом Мекки (Masky, 1899), определившим это понятие при исследовании окатанности обломочных частиц. Модель носила детерминистический характер и утверждала в общем виде, что окатанность обломочной частицы является функцией небольшого числа характеристик. Она была настолько неконкретна, что не давала никакого метода решения вопроса в частных задачах. Сомнительно, чтобы и сам автор модели

четко представлял, как следует действовать в конкретных случаях. Через 40 лет В. С. Крамбайн при исследовании галечников каньона Сан-Габриел возродил детерминистическую модель Мекки в виде конкретного дифференциального уравнения (Krumbein, 1940). В 1941 г. он уточнил уравнение (Krumbein, 1941). В 1948 г. идеи Крамбайна реализовал Пламли, исследовавший перенос галек (Plumley, 1948). Дальнейшие попытки развития детерминистических моделей были неоднократны. Так, например, они использовались при построении модели формирования турбидитов (Hinze, 1960; Plapp, Mitchell, 1960); проводились опыты моделирования процессов рудообразования (Голубев, 1975) и т. д. Эта линия моделирования полезна на предварительной стадии, когда необходимо выявить основное содержание задачи. При построении таких моделей не учитывался стохастический характер явлений, что естественно приводило к трудностям.

В 1941 г. А. Н. Колмогоров⁶ построил первую в геологии стохастическую модель (Колмогоров, 1941). Она была посвящена выводу логарифмически нормального распределения из схемы дробления. Эта модель, в принципе имевшая огромное значение как реализация исключительно продуктивного нового подхода, не вызвала в момент появления никакого интереса у геологов. Она опиралась на безукоризненный математический аппарат, но в ее основу была положена примитивная, неадекватная природе явления аксиоматика (Миддлтон, 1968). Не было также проверки согласия с наблюдениями.

В 1945 г. автор этих строк, не зная в условиях военного времени об исследовании А. Н. Колмогорова, ввел вероятностную модель процессов формирования пористости в карбонатных породах (Вистелиус, 1945). Модель опиралась на тщательно разработанную и проверенную наблюдениями в шлифах аксиоматику. Математический аппарат был примитивен, согласие модели с наблюдениями не проверялось. Убежденный в большом значении найденного подхода, автор установил связи с математиками. В итоге в 1947 г. в геологической литературе появилась первая статья, в которой содержались геологическая аксиоматика, математический вывод функции распределения и использовался критерий согласия между наблюдениями и выведенной функцией (Вистелиус, Сарманов, 1947). Возник метод решения геологических задач с помощью вероятностных моделей, метод прямых задач геологии. В дальнейшем наиболее полно этот подход был реализован при изучении механизма слоеобразования и задач, связанных с формированием

⁶ Работа А. Н. Колмогорова является первой в отмеченном направлении, если рассматривать исследования геологических объектов. В экспериментах по транспортировке осадков, скорее относящихся к гидравлике, но важных и для седиментологии, по-видимому, первой была стохастическая модель, предложенная Эйнштейном (Einstein, 1937). Содержательные данные по состоянию вопроса в этой области имеются в работе (Lee, Jobson, 1977). На последнюю работу внимание автора обратил В. Н. Подковыров.

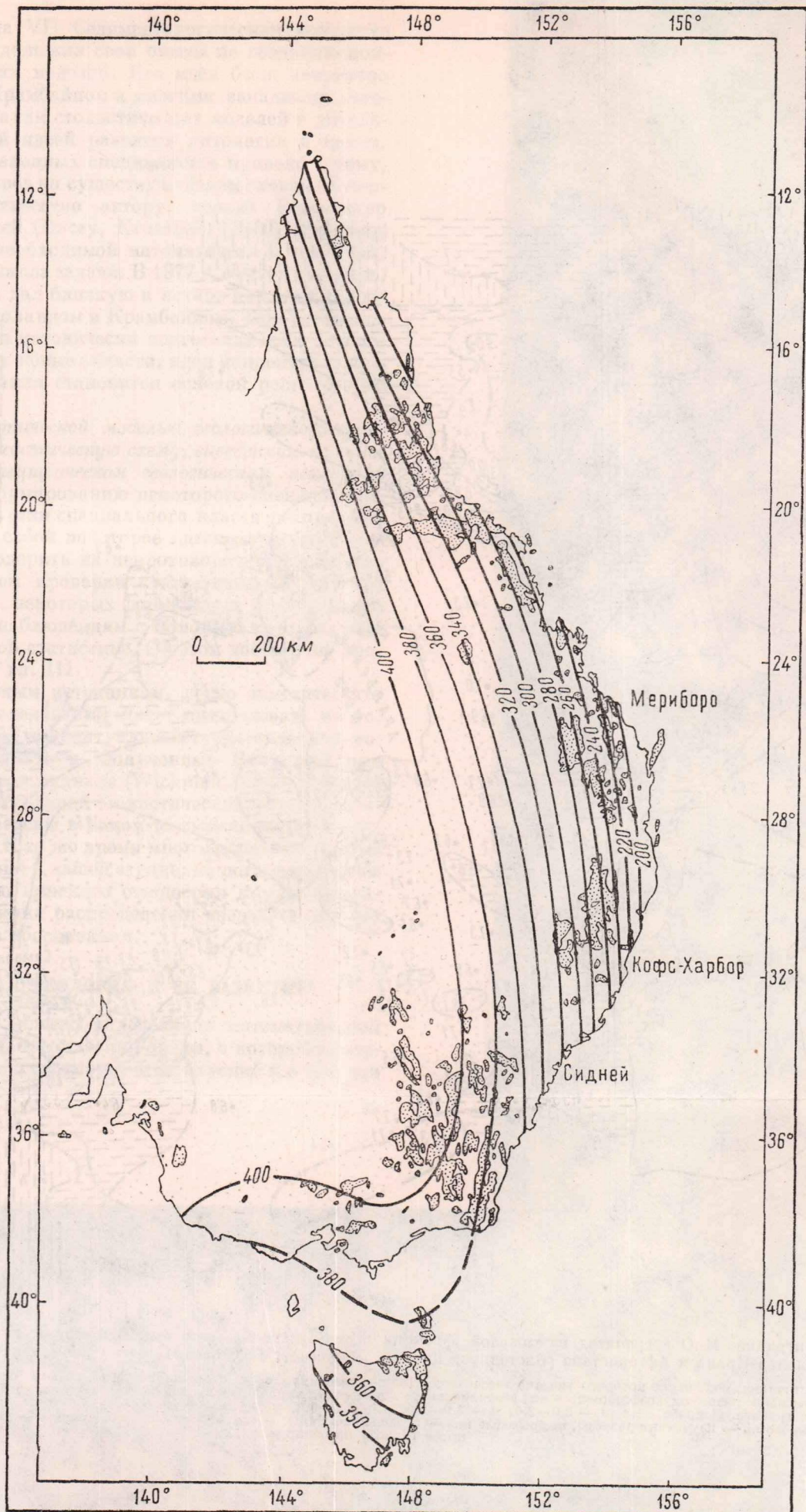


Рис. 1.1. Основная тенденция изменения возраста фанерозойских гранитоидов Восточной Австралии.

Изолинии отвечают линиям уровня функции $z = \text{схр } P_3(\varphi, \lambda)$, где P_3 — полином третьего порядка от географических координат; цифры — миллионы лет, точки — гранитные интрузии.

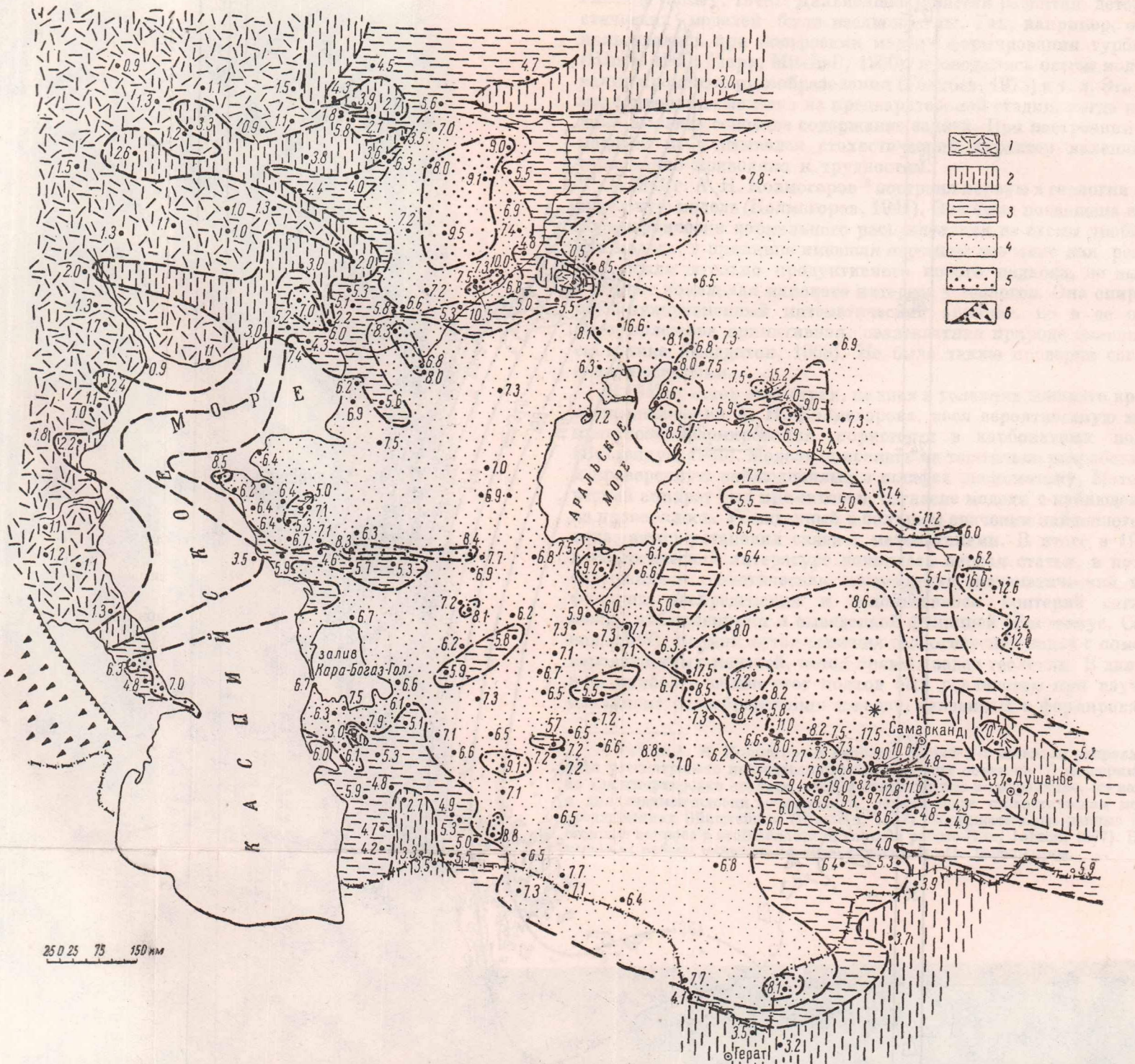


Рис. П.8. Поле концентраций MnO в гранатах из тяжелой фракции антсеноманских песчаных отложений юга СССР и прилегающих районов Ирана и Афганистана (Вистелюс, Демина, Харламов, 1976 с некоторыми дополнениями).

Спессартит поступает из источника, положение которого указано звездочкой. Наличие максимумов в содержании MnO вдали от предполагаемого источника питания может быть объяснено разрывом коэффициентов дифференциального уравнения, управляющего переносом терригенного материала.
 1-5 — содержания MnO в гранате в пробе, %: 1 — 0.0-2.0, 2 — 2.0-4.0, 3 — 4.0-6.0, 4 — 6.0-8.0, 5 — > 8; 6 — границы распространения тяжелых фракций с гранатом; жирные точки — изученные разрезы, цифры около них — содержание MnO в пробе, отсутствие цифр — указание на то, что в этой пробе граната не обнаружено.

гранитоидов. В 1967 г. на VII Седиментологическом конгрессе в Рединге (Англия) автор доложил свои опыты по созданию концептуальных стохастических моделей. Его идеи были энергично подхвачены в общем виде Крамбайном и многими западными литологами. Идея об использовании стохастических моделей в литологии оказалась центральной идеей развития литологии в целом. Однако отсутствие подготовленных специалистов привело к тому, что в основном все дело свелось по существу к общим словам (Krumbein, 1969). Насколько известно автору, только Шварцзахер (Schwarzacher, 1975) и Дасей (Dacey, Krumbein, 1976) оказались в состоянии справиться с необходимой математической техникой, не теряя геологического смысла задачи. В 1977 г. Уиттен (Whitten, 1977) в обзоре работ автора дал близкую к истине картину, согласовав свою публикацию с Холандом и Крамбайном. Так, оставаясь с 1967 г. в центре внимания методически подготовленных литологов и завоевывая все новые и новые области, идея концептуального стохастического моделирования становится основой роста нового парадигма геологии.

Концептуальной стохастической моделью геологического явления мы будем называть стохастическую схему, выведенную на основании предположений о специфическом геологическом механизме явления и приводящую к формированию некоторого специального распределения вероятностей (или специального класса распределений). Модель представляет собой некоторое математическое выражение. Его необходимо проверить на непротиворечивость наблюдениям. Это делается путем проверки статистических гипотез, вытекающих из модели, т. е. некоторых формальных утверждений, поддающихся проверке по наблюдениям с помощью методов, разработанных в математической статистике. Об этом достаточно подробно мы будем говорить в гл. III.

Обращаясь к литературным источникам, легко заметить, что только небольшая часть исследований может претендовать на то, что они проведены на основе концептуальной стохастической модели. Так, например, результаты, полученные Викманом при исследовании им работы жерла вулкана (Wickman, 1966), действительно построены на концептуальной стохастической модели. Приводимые им выражения выведены в каком-то смысле из реальных геологических наблюдений. В то же время многочисленные работы Аренса, почему-то популярные в нашей стране, по логарифмическим нормальному распределению никакого отношения к моделированию не имеют, так как функция распределения введена в них без какого-либо геохимического обоснования.

1.5. МАТЕМАТИЧЕСКАЯ ГЕОЛОГИЯ И ЕЕ РАЗВИТИЕ

Мы начали эту главу с общего определения математической геологии. Выяснив вопрос об организации среды, в которой протекают геологические явления, о структуре этих явлений и о том, что

такое модель, мы можем уточнить приведенное определение. *Математической геологией мы будем называть научную дисциплину, занимающуюся построением, анализом и использованием при решении конкретных задач концептуальных вероятностных моделей геологических явлений.*

Модель должна быть концептуальной, т. е. должна отражать главное содержание геологической идеи, подлежащей проверке. Работа заключается в выдвижении некоторых геологических предположений. Эти предположения, выраженные в математической форме, проверяются математическими средствами. В такой форме постановка задачи отражает специфику математики.

Модель должна быть вероятностной, так как объекты, изучаемые геологией, порождаются средой, для которой типичными являются закономерности, выражаемые вероятностным языком. Конечно, в отдельных случаях можно представить детерминированную модель, дающую хорошее приближение. Как показывает опыт, это достаточно редкое исключение, которое может быть интерпретировано как вырожденный случай вероятностной модели. Такие детерминированные модели, как только что отмечалось, иногда очень полезны в качестве первого шага в выработке вероятностного подхода, когда стохастический смысл задачи еще не уловлен, а схематизация концептуальной стороны проблемы уже нащупана.

Об особенностях построения, анализа и использования моделей было сказано в предыдущем параграфе.

Вопрос о развитии математической геологии носит скорее организационный, а не научный характер, поэтому он не может входить в основы этой науки. Однако кратко на нем все же надо остановиться, так как в настоящее время важно добиться понимания ряда положений, без чего развитие этой науки сильно замедлится.

Прежде всего, нужно со всей энергией подчеркнуть, что математическая геология — весьма специфическая самостоятельная наука. Может быть именно к ней особенно хорошо подходят слова Бюффона о том, что ее трудность в широте объекта исследования и в необходимости изучения его чрезвычайно детальными методами.

Математическая геология не может развиваться без глубокой геологической основы при постановке любых ее задач. Поэтому математическую геологию должны развивать геологи, а не геофизики, лингвисты, теплотехники или другие специалисты, знакомые с математическими методами. Необходимы оригинальные первичные наблюдения, выполняемые по специфическим схемам.

Математическая геология — математическая дисциплина в группе геологических наук, поэтому для ее развития и оценки ее результатов необходимо по меньшей мере отчетливое знание математических идей, лежащих в основе используемого ею аппарата. Без этого невозможно ни оценивать полученные результаты, ни развивать саму математическую геологию. Дело не в том, просты или сложны методы, — они должны быть адекватны задаче.

В этом отношении очень показательны применение статистических методов в том виде, в каком они изложены в упрощенных руководствах. Успех их применения целиком зависит от того, адекватна ли задаче модель, лежащая в основе соответствующего статистического метода. Это недавно рассматривалось на конкретном материале (Вистелиус, 1978). Любой статистический метод исходит из модели, хотя в популярных учебниках этот момент обычно не акцентируется. Успех зависит от опробования, что также не отменяется. В каждодневной практике имеется множество примеров, когда модель игнорировалась и результаты расчета только дезориентировали исследователей. Например, вычислялись коэффициенты корреляции между гранулометрическим составом пород и содержанием алмазов на всем диапазоне составов обломочных пород. Такой подсчет заведомо лишен смысла, так как алмазы сконцентрированы только в грубообломочных фракциях. Таким образом, даже простейшие статистические методы нельзя вводить механически, не оценивая геологического смысла задачи.

Действительно эффективные решения достигаются только при совместном ведении работ геологом и математиком, при творческом отношении обоих к решаемой задаче.

Все это показывает, что мнения некоторых лиц о том, что задачи математической геологии должны решаться простыми методами, несостоятельны. Как было только что отмечено, дело не в сложности или простоте, а в адекватности метода задаче (Миддлтон, 1968). Как правило, задачи сложны, и поэтому должен быть использован развитый математический аппарат. Для того чтобы сделать хороший химический анализ, нужно иметь химически чистые реактивы, хотя получить их трудно. Для того чтобы провести содержательную дедукцию математическими методами, нужно иметь эффективный математический аппарат, хотя освоение его требует большой работы.

Существует еще одна сторона, требующая оговорки. Как отмечалось, применять математику значит в явном или скрытом виде использовать модели. Использование готовых популярных руководств иногда означает привлечение слишком общих моделей. Это не стимулирует открытия новых областей исследования обычными геологическими методами, как это бывает при удачном моделировании. В итоге при работе с упрощенными моделями не происходит развития науки, открытия новых ее разделов. Все дело сводится только к чисто технологическому усовершенствованию анализа наблюдений. Яркий пример тому — использование в геологии дискриминантного анализа. Очень полезный метод сам по себе, он был низведен до описательного. Все внимание исследователей было обращено на использование новых оценочных средств, но принципиального развития, специфического для геологии, не произошло. Все осталось на техническом уровне. Внедрение вычислительных методов без построения концептуальных моделей («вычислительная геология» некоторых авторов) — яркий

пример вульгаризации математической геологии (Vistelius, 1974).

В заключение несколько слов о так называемых фундаментальных и нефундаментальных науках. Этой теме уже касался Бреш (Brush, 1976, 1978), мы же немного можем добавить к сказанному им. Фундаментальность науки определяется тем, что получаемые результаты широко используются в других науках. Именно это характерно для математики. Однако это обстоятельство совершенно не означает, что науки, не относимые к фундаментальным, менее содержательны или более поверхностны. Создается впечатление, что роль фундаментальных приобрели те науки, которые в начале развития имели исключительно простую аксиоматику. Это позволило сконцентрироваться на развитии аппарата дедукции. Понять и принять то, что прямая — есть кратчайшее расстояние между двумя точками на плоскости, кажется, было не так уж и сложно. Положить в основу рассуждение о том, что тепло распространяется от теплого тела к холодному и убедиться в том, что обратный процесс невозможен, уже потребовало более столетия (от Бернулли до Карно). Что же говорить об основных принципах геологии? Мы сознательно рассмотрели вопрос об определении возраста Земли по скорости ее охлаждения. По-видимому, никто из крупных ученых до конца XIX в. не сомневался в том, что Земля образовалась из огненно-жидкой массы. По меньшей мере странным и абсолютно ненаучным было бы в то время допущение, что внутри Земли скрыты источники для ее разогревания. Поэтому исходная аксиоматика Фурье и Томсона была для своего времени абсолютно корректна. Сейчас выяснены такие факты, которые ничего не оставляют от этой аксиоматики. В подобной ситуации требуется огромное напряжение, чтобы создать необходимую аксиоматику. И силы ученых уходят на поиски такой аксиоматики. Методов разработки аксиоматики, по-видимому, не существует. Имеются методы проверки ее формально корректного построения, но это только небольшая часть того, что требуется для аксиоматики естествознания.

Сейчас для математической геологии имеет решающее значение выбор простой и ясной аксиоматики. Кажется, что геология имеет в этом плане большие и неразработанные резервы. Понимание исследователями, что точная дедукция из этой простой аксиоматики способна вскрыть важнейшие факты, может совершить научную революцию в нашей науке.

Путем изучения истории развития геологии мы пришли к заключению, что появление математической геологии — естественный этап развития всего цикла геологических наук. Элементы, на которых построена математическая геология, — аксиоматический подход, развитие представления о генезисе геологических объектов, точные наблюдения и достаточно мощный формальный

аппарат для дедукции — в той или иной мере уже были испробованы геологами, вскрыты их слабые и сильные стороны и накоплен достаточный опыт для того, чтобы все это могло быть критически освоено.

Реализуя аксиоматический подход, математическая геология наталкивает исследователя на поиски у изучаемых объектов информативных признаков, тем самым расширяя области использования традиционных методов геологии. Решение достигается не путем интерпретации заранее намеченных признаков, а как итог сравнения всей совокупности данных об объекте в точно фиксированной ситуации с теми наблюдениями, которые требуются для проверки соответствующей аксиоматики. Проверка согласия теории с наблюдениями достигается с помощью специального формального аппарата. Основные особенности этого аппарата будут рассмотрены в следующих главах.

Литература

- Боуэн Н. Л. Эволюция изверженных пород. М.—Л.—Новосибирск, Гос. научн. техн. горно-геолого-нефтяное изд-во, 1934. 324 с.
- Вернадский В. И. Биосфера. Л., Научное хим.-техн. изд-во, 1926. 146 с.
- Вернадский В. И. Об условиях появления жизни на Земле. — Изв. АН СССР, сер. 7, математ. и естеств. науки, 1931, № 5, с. 633—653.
- Вернадский В. И. Химическая организация биосферы Земли и ее окружения. М., «Наука», 1965. 344 с.
- Вистелиус А. Б. Заметки по аналитической геологии. — Докл. АН СССР, 1944, т. 44, № 4, с. 27—31.
- Вистелиус А. Б. Распределение частот коэффициентов пористости и эпигенетические процессы в спириферовых слоях Бугурусланского нефтеносного района. — Докл. АН СССР, 1945, т. 49, № 1, с. 44—47.
- Вистелиус А. Б. Проблемы математической геологии. К истории вопроса. — Геология и геофизика, 1962, № 12, с. 3—9.
- Вистелиус А. Б. Математическая геология (состояние, перспективы). — В кн.: Математическая геология, реферативный систематический указатель. Л., изд. Библ. АН СССР, 1969, с. 11—56.
- Вистелиус А. Б. Математическая геология — ее основные направления и задачи. — Советская геология, 1977, № 1, с. 11—34.
- Вистелиус А. Б. Платиноиды и золото в силикатной, троилитовой и металлических фазах хондритов. — Зап. ВМО, 1978, № 3, с. 257—270.
- Вистелиус А. Б., Романова М. А. О вырожденном случае модели кристаллизации идеальных гранитов. — Докл. АН СССР, 1976, т. 228, № 1, с. 170—173.
- Вистелиус А. Б., Сарманов О. В. Стохастическое обоснование одного геологически важного распределения вероятностей. — Докл. АН СССР, 1947, т. 58, № 4, с. 631—634.
- Геологический словарь. Т. I. М., «Недра», 1973. 415 с.
- Голубев В. С. Динамика формирования инфильтрационных эпигенетических месторождений. — В кн.: Применение математических методов и ЭВМ в рудной геологии. Иркутск, Изд-во Иркутск. гос. ун-та, 1975, с. 36—50.
- Делл Р. О. Изверженные породы и глубины Земли. Л.—М., ОНТИ, 1936. 591 с.
- Калицкий К. П. Научные основания поисков нефти. М.—Л., Гостехиздат, 1944. 224 с.

- Колмогоров А. Н. О логарифмически-нормальном законе распределения размеров частиц при дроблении. — Докл. АН СССР, 1941, т. 31, № 2, с. 99—101.
- Коржинский Д. С. Теория процессов минералообразования. М., Изд-во АН СССР, 1962. 24 с.
- Кропоткин П. А. Исследования о ледниковом периоде. — Зап. РГО по общ. геогр., 1876, т. 7. 840 с. с прил.
- Кудрявцев Н. А. Происхождение нефти и газа. — Тр. ВНИГРИ, вып. 319. Л., «Недра», 1973. 214 с.
- Левинсон-Лессинг Ф. Ю. О пределах и подразделениях семейства андезитов. — Вестн. Геол. ком., 1924, № 6, с. 723—735.
- Лайель Ч. Основные начала геологии (перевод А. Мин). М., Изд-во А. И. Глазунова, 1866, т. I. 399 с.
- Миддлтон Г. В. Возникновение логнормального распределения частот в осадках. — В кн.: Вопросы математической геологии. Л., «Наука», 1968, с. 37—44.
- Разумовский Н. К. О распределении концентраций металлов в рудных месторождениях. — Докл. АН СССР, 1940, т. 28, № 9, с. 815—817.
- Романова М. А. Белые граниты Арга-Ышных-Хая (Якутия) и механизм их образования. — В кн.: Исследования по математической геологии. Л., «Наука», 1978, с. 25—40.
- Севастьянов А. Геогнозия, или наука о горах и породах (лекции Вернера, записанные Нассе). СПб., 1810. 350 с.
- Ahrens L. N. A fundamental law of geochemistry. — Nature, 1953, v. 172 (4390), p. 1148.
- Bowen N. L. The Binary System: (Nephelite—Carnogeite)—Anorthite. — Amer. J. Sci., 1912, v. 33, No. 198, p. 551—573.
- Brush S. G. Irreversibility and indeterminism: Fourier to Heisenberg. — J. History Ideas, 1976, v. 37, No. 4, p. 603—630.
- Buffon, de. Histoire Naturelle, générale et particulière, Servant de suite à la Théorie de la Terre a. d'introduction à l'histoire de Minéraux. A Paris, de l'imprimerie Royale, 1774. 422 p.
- Buffon, de. Histoire Naturelle, générale et particulière. V. I, II. Aux Deux-Ponts. Sanson a. Compagnie, 1785, p. 7—332.
- Dasey M. F., Krumbain W. C. Topological Properties of Disjoint Channel Networks Within Enclosed Regions. — J. Intern. Assoc. Math. Geol., 1976, v. 8, No. 4, p. 429—462.
- Durocher J. Essai de pétrologie comparée, on recherches sur la composition chimique et mineralogique des roches ignées, sur les phénomènes de leur émission et sur leur classification. — Ann. des Mines, 1857, v. 11, 5 ser., p. 217—259.
- Einstein H. A. Der Geschiebetrieb als Wahrscheinlichkeitsproblem. — Mitteilungen der Verschanstalt für Wasserbau, an der Eidgenössische Technische Hochschule in Zürich, Verlag Rascher und Co., 1937. 110 S.
- Eisenhart C. A test for the significance of lithological variations. — J. Sediment. Petrol., 1935, v. 5 (3), p. 137—145.
- Emmons W. H. Relations of metalliferous lode systems to igneous intrusives. — Trans. Amer. Inst. Min. and Met. Engrs., 1927, v. 74, p. 29—70.
- Fisher R. A. The expansion of statistics. — J. Roy. Statist. Soc., 1953, ser. A, v. 116, No. 1, p. 1—6.
- Fourier J. B. J. Memoire sur le refroidissement sécubaire du globe terrestre. — Ann. Chimie et de physique, 1819, v. 13, p. 418—438.
- Fourier J. B. J. Memoire sur les temperatures du Globe terrestre et des espaces planétaires. — Mem. de l'Acad. Royale Sci. de l'Institut de France, 1827, v. 7, p. 570—604.
- Gesneri J. Tractatus Physicus de Petrificatis. Theodorum Haak, 1758, 136 p.
- Glossary of Geology. Amer. Geol. Inst. Washington D. C. 1972, p. 434.
- Goldschmidt V. M., Die Gesetze der Gesteinsmetamorphose, mit Bei-

- spielen aus der Geologie des südlichen Norwegens. — Vidensk. Skrift. 1912, v. 1, Math.-Naturv., No. 22, Kristiania, S. 16.
- Hall J. On granite. — Trans. Roy. Soc. Edinburgh, 1794, v.3, p. 8—12.
- Harker A. The Natural History of Igneous Rocks. London, Methuen & Co, 1909. 384 p.
- Hinze J. O. On the hydrodynamics of turbidity currents. — Geol. Mijnbouw, 1960, v. 39, No. 1, p. 18—25.
- Hutton J. Theory of the Earth or an Investigation of the Laws observable and the Composition, Dissolution and Restoration of Land upon the Globe. — Trans. Roy. Soc. Edinburgh, 1788, v. 1, pt II, p. 209—304.
- Hutton J. Theory of the Earth with Proofs and Illustrations. V. 1, 1795, v. 2, 1795, Edinburgh; v. 3, 1899, London.
- Huxley T. H. Anniversary Address of the President. — Quart. J. Geol. Soc. of London, 1869, v. 25, pt I, p. 28—53.
- Kirwan R. Essai sur le phlogistique, et sur la constitution de Acides. Paris, Rue et Hotel Serpente, 1788. 344 p.
- Kirwan R. Geological Essays. London, Bensley, 1799. 502 p.
- Kirwan R. Observations on the Proofs of the Huttonian theory of the Earth, adduced by Sir James Hall Bart. — Trans. Roy. Irish Acad., 1802, v. 8, p. 3—27.
- Krumbein W. C. The probable error of sampling sediments for mechanical analysis. — Amer. J. Sci., 1934, v. 227, № 159, p. 204—214.
- Krumbein W. C. Flood gravel of San Gabriel Canyon. — Geol. Soc. Amer. Bull., 1940, v. 51, № 5, p. 639—676.
- Krumbein W. C. The effects of abrasion on the size, shape, and roundness of rock fragments. — J. Geol., 1941, v. 49, № 5, p. 432—521.
- Krumbein W. C. Deterministic and Probabilistic Models in Geology. — In: Models of Geologic Processes. Washington, Amer. Geol. Inst. 1969, WCK-B-1—W-B-14a.
- Kuenen Ph. H., Migliorini C. J. Turbidity currents as a case of Graded Bedding. — J. Geol., 1950, v. 58, No 2, p. 91—127.
- Kuhn T. S. The Structure of Scientific Revolution. Chicago, Univ. of Chicago Press, 1970. 210 p.
- Landes K. K. The paragenesis of the granite pegmatites of central Maine. — Amer. Miner., 1925, v. 10, No. 11, p. 355—411.
- Lee B. K., Johnson H. E. Stochastic analysis of particle movement over a dune bed. — Prof. Paper 1040, Geological Survey. Washington, Governmental Printing Office, 1977. 72 p.
- Leibnitii G. G. Protogaea. 1680. цит. по: Leibnitii G. G. Protogaea in Lucem edita Ch. L. Scheidio, Goettingae, J. G. Schmidii, 1749. 86 p.
- Lindgren W. Magmas, dikes and veins. — Trans. Amer. Inst. Min. Met. Engrs., 1927, v. 74, p. 71—126.
- Lyell Ch. Principles of Geol., v. 1, 2, 3. London, J. Murray, 1833—1837.
- Mackay Wm. On the laws that govern the rounding of particles of sand. — Trans. Edinb. Geol. Soc., 1899, v. 7, p. 298—312.
- Mairan, de. Memoire sur la cause generale du froid en hyver, & de la chaleur en Ete. Histoire de L'Academie Royale des Sciences, 1719. 31 p.
- Naslund H. R. Liquid Immiscibility in the System. . . and its Application to Natural Magmas. — Ann. Rept. Director Geophys. Labor., Carneg. Inst. Washington, Year Book 75, 1976, p. 592—597.
- Niggli P. Die Leichtflüchtigen Bestandteile im Magma. Leipzig, B. G. Teubner, 1920. 272 p.
- Niggli P. Anwendungen der mathematischen Statistik auf Probleme der Mineralogie und Petrologie. — Neues Jahrbuch Miner, 1923, v. 48, p. 167—212.
- Pfaff F. Grundriss der Geologie. Leipzig, Engelmann, 1876. 399 p.
- Plapp J. C., Mitchell J. P. A hydrodynamic theory of turbidity currents. — J. Geoph. Res., 1960, v. 65, No. 3, p. 983—992.

- Playfair J. Illustrations of the Huttonian Theory of the Earth. Edinburgh. 1802. 528 p.
- Plumley W. J. Black Hills Terrace gravers: a study in sediment transport. — *J. Geol.*, 1948, v. 56, No. 6, p. 578—591.
- Reyer E. Beiträge zur Physik der Eruptionen und der Eruptivgesteine. Vienna, 1877. 225 S.
- Richardson W. A., Sneesby G. The frequency distribution of igneous rocks. — *Miner. Magaz. a. Journ. Miner. Soc.*, 1922, v. 19, No. 97; v. 20, No. 100.
- Sandberger F. Untersuchungen über Ergänge. Wiesbaden, Bd I. C. W. Kreidel, 1882, Hft 1, S. 1—158; 1885, Hft 2, S. 159—430.
- Schwarzacher W. Sedimentation models and quantitative stratigraphy. Amsterdam, Elsevier, 1975. 382 p.
- Sedgwick A., Murchison R. I. On the Distribution and Classification of the older or Palaeozoic Deposits of the North of Germany and Belgium, and their comparison with Formations of the same age in the British Isles. — *Geol. Trans.*, 1840, v. 6, 2 ser., p. 221—301.
- Smith W. Tubular View of the British Strata. London, 1790.
- Tait P. G. Geological time. — *North British Rev.*, 1869, v. 50, p. 406—439.
- Thomson W. On geological time. — *Trans. Soc. Glasgow*, 1868, v. 3, pt 1, p. 1—28.
- Tomkeieff S. G. Analytical Geology. — *Nature*, 1947, v. 160, No. 4076, p. 846, 847.
- Vistelius A. B. Mathematical Geology: report of progress. — *Geocom. Bul.*, 1968, v. 1, No. 8, p. 229—269.
- Vistelius A. B. Proper Function and Role of the International Association For Mathematical Geology in the revolution in Geological Sciences. — *Intern. Assoc. Math. Geol. Kansas serie, News Lett.*, 1974, No. 3, p. 1.
- Vistelius A. B. Mathematical Geology and Development of Geological Sciences. — *J. Intern. Assoc. Math. Geol.*, 1976, v. 8, No. 1, p. 3—8.
- Vistelius A. B. Mathematical Geology and the Progress of Geological Sciences. — *J. Geol.*, 1976, v. 84, No. 6, p. 629—653.
- Wickman F. E. Repose period patterns of volcanoes. V. General Discussion and tentative stochastic model. — *Arkiv Miner. Geol.*, 1966, Bd 4, Hft 5, No. 11, S. 351—367.
- Williams J. Natural History of Mineral Kingdom in three parts. Edinburgh, v. 1, 1789, 451 p.; v. 2, 1789, 532 p.
- Whitten E. H. T. Vistelius' views on conceptual stochastic models. — *J. Geol.*, 1977, v. 85, No. 3, p. 326—328.

Глава II

ВЕРОЯТНОСТНОЕ ПРОСТРАНСТВО И СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Основные сведения из теории вероятностей — важнейшие определения и элементы техники исследований. Общие свойства вероятностей, случайных величин и случайных функций. Векторная случайная величина, преобразование случайных величин, случайные процессы, геометрические вероятности.

Ключевые слова:

вероятность, распределение, случайная величина, математическое ожидание, случайный процесс, геометрические вероятности.

II.1. ВВЕДЕНИЕ

В первой главе мы достаточно подробно рассмотрели важность для геологических построений знания вероятности появления в типичной ассоциации того или иного геологического объекта (или значения характеристики объекта). Однако до сих пор, говоря о вероятности, мы не дали точного определения этого понятия. В настоящей главе даются основные определения и приводятся элементы специфических методов исследования в математической теории вероятностей. Используемое в математике и вводимое в настоящей главе понятие вероятности (вероятностной меры) не адекватно житейскому понятию возможности или вероятности, используемому в обиходном языке или в языке неформализованных научных дисциплин. Вероятность в математике — это всего лишь абстракция свойства статистического ансамбля, приписываемая отдельному элементу этого ансамбля. Эта абстракция оказалась исключительно плодотворной для построения математической теории вероятностей и для выражения в терминах последней геологических явлений, что постепенно будет раскрываться в этой книге. Практическая неприложимость понятия «вероятность» к единичным событиям вне статистического ансамбля, конечно, ограничивает его применение. Однако во многих случаях в геологии эта трудность обходится путем простой переформулировки задачи.

Итак, мы переходим к ознакомлению с основами теории вероятностей. Они даны в той полноте, какая доступна при данном объеме книги и достаточна для сознательного построения значительного числа вероятностных моделей геологических явлений.

II.2. ДИСКРЕТНОЕ ПРОСТРАНСТВО ЭЛЕМЕНТАРНЫХ СОБЫТИЙ

Рассматривается дискретное пространство элементарных событий, при котором не возникает трудностей, связанных с измеримостью и интегрированием. Вводятся основные понятия теории вероятностей для частного случая дискретного вероятностного пространства, которые в большинстве своем будут заново сформулированы в других разделах при использовании вероятностного пространства общего вида.

II.2.1. Вероятностное пространство

Для того чтобы ввести читателя в круг идей теории вероятностей, не усложняя первого знакомства с предметом техническими трудностями, рассмотрим сначала частный случай *дискретного вероятностного пространства*. Это случай, когда в результате испытания или эксперимента возможно только *конечное* или *счетное множество* исходов. Примеры испытаний с конечным множеством исходов может привести каждый читатель. Классический пример — это бросание монеты, в результате которого она падает той или иной стороной вверх. Простой пример бесконечного множества исходов дает испытание, состоящее в том, что монета бросается до первого выпадания «герба». Исход этого испытания — натуральные числа по числу бросаний до первого «герба». Эти примеры носят искусственный характер. Они отражают историю развития теории вероятностей, результаты которой сначала применялись для подсчета шансов игроков при игре в азартные игры. Но и на практике встречаются похожие ситуации. Каждому геологу ясно, например, что находка интересующей стратиграфа формы в том или ином ископаемом биоценозе в принципе аналогична появлению герба при бросании монеты. Некоторые могут даже указать вероятность встречи такой формы, которая не обязательно равна $1/2$, как в случае с монетой. Очевидно, что к такой же категории явлений относятся и случайные последовательности одинаковых слоев в осадочных толщах или располагающиеся рядом зерна одного и того же минерала в линейном сечении порита — это те же серии до первого выпадания «герба». Только вероятности будут иные, и их подсчет будет происходить по другим правилам, нежели в вероятностной модели бросания монеты. Рассмотрим следующий пример.

Пример II.1. Пусть у нас имеется механизм выбора одного зерна из гранита, которое может оказаться зерном одного из n

минералов. В этом испытании естественно выбрать пространство элементарных событий, состоящее из n точек. Пусть при этом $n=3$, а минералами являются кварц (Q), калиевый полевой шпат (KFsp) и плагиоклаз (Pl). Тогда $\Omega = \{Q, KFsp, Pl\}$. При более детальном описании, когда определяется тип калиевого полевого шпата и различаются ортоклаз (Or), анортклаз (An) и микроклин (Mi), пространство элементарных событий будет содержать уже пять точек и $\Omega = \{Q, Or, An, Mi, Pl\}$. Если мы отбираем m зерен гранита, то с учетом их порядка возможно уже $m \cdot n$ различных исходов. Пространство Ω состоит из последовательностей длиной m , причем каждое зерно в такой последовательности может быть одного из n минералов. Вопрос о вероятностях тех или иных сочетаний зерен мы пока оставляем в стороне. Обратим внимание читателя на то обстоятельство, что в последнем случае испытанием называется выбор всей последовательности из m зерен, а событие, состоящее в том, что первым зерном в последовательности оказался, скажем, кварц, — это сложное событие, состоящее из таких рассмотренных элементарных событий, т. е. из последовательностей тех составов, которые начинаются с зерна кварца.

Пусть $\Omega = \{\omega_1, \omega_2, \dots\}$ — множество исходов. Мы будем называть Ω *пространством элементарных событий*, а ω_i — *элементарными событиями*. Пусть каждому $\omega \in \Omega$ соответствует число $P(\omega)$, которое называется вероятностью элементарного события ω . При этом предполагается, что:

$$1a) 0 \leq P(\omega) \leq 1 \text{ для любого } \omega \in \Omega,$$

$$2a) \sum_{\omega \in \Omega} P(\omega) = 1.$$

Любое подмножество множества элементарных событий называется *событием*. В некоторых случаях из данного множества можно выделить подмножества, указав общее свойство, присущее всем элементам этого подмножества.

Пусть A — множество всех тех и только тех элементов множества Ω , которые обладают свойством α , а B — множество элементов, обладающих свойством β .

Тогда событие $A \cap B$ (пересечение множеств, которое иногда обозначается также AB) состоит в том, что выполняется одновременно свойство α и β , $A \cup B$ (объединение) — выполняется хотя бы одно из двух свойств α или β , $B \setminus A$ (разность) — выполняется свойство β , но не свойство α , $B = \Omega \setminus B$ (дополнение B) — свойство β не выполняется.

Вероятность события A , где $A \subset \Omega$, определяется по правилу

$$P(A) = \sum_{\omega \in A} P(\omega). \quad (II. 2. 1)$$

При этом вероятность события «выполняется свойство α » равна $P(A)$, где $A = \{\omega \in \Omega : \omega \text{ обладает свойством } \alpha\}$.

Таким образом, *вероятность — это числовая функция P , областью определения которой является множество Ω всех под-*

множестве множества A . Заметим, что мы употребляем слово «вероятность» как для функции P , так и для ее значения $P(A)$ для некоторого $A \in \mathfrak{A}$ (ситуация, сходная с употреблением слов «мера» и «расстояние»). В том случае, если мы хотим подчеркнуть, что речь идет о функции, а не о ее значении, мы будем говорить «вероятностное распределение» и, наоборот, «значение вероятности», если речь идет о значении этой функции. Функция эта не произвольна. Она подчиняется правилам, которые в данном случае вытекают из формулы (II.2.1).

1б) $0 \leq P(A) \leq 1$ для любых $A \in \mathfrak{A}$, при этом $P(\Omega) = 1$ (вероятность достоверного события равна 1).

2б) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ — вероятность объединения счетного числа попарно непересекающихся событий равна сумме вероятностей этих событий, если $A_1, A_2, \dots \in \mathfrak{A}$ и $A_i \cap A_j = \emptyset$ (где через \emptyset обозначено пустое множество и $i \neq j$).

Из 1б) и 2б) следует, что $P(\emptyset) = 0$ (вероятность невозможного события равна 0), а также, что $P(\bar{B}) = 1 - P(B)$.

Из формулы (II.2.1) также вытекает формула сложения вероятностей для пересекающихся событий

$$P(A \cup B) = P(A) + P(B) - P(AB), \quad (\text{II.2.2})$$

которую легко обобщить на случай произвольного числа слагаемых, используя первый распределительный закон для операций над множествами

$$(A \cup B)C = AC \cup BC$$

или двойственный ему второй распределительный закон

$$AB \cup C = (A \cup C)(B \cup C).$$

Пусть $A, B \in \mathfrak{A}$ и $P(B) > 0$. Отношение $P(A|B) = \frac{P(AB)}{P(B)}$ называется *условной вероятностью* A относительно B (при фиксированном B). Переписанная в виде

$$P(AB) = P(B)P(A|B), \quad (\text{II.2.3})$$

эта формула носит название формулы *умножения вероятностей*.

События A и B называются *независимыми*, если $P(AB) = P(A)P(B)$. В этом случае $P(A|B) = P(A)$, условная вероятность (если она имеет смысл) равна безусловной. Легко показать, что если A и B независимы, то A и \bar{B} , \bar{A} и B , \bar{A} и \bar{B} также независимы.

Пусть

$$A \subset \bigcup_{i=1}^{\infty} B_i,$$

где $B_i \cap B_j = \emptyset$ ($i \neq j$).

Тогда

$$A = \bigcup_{i=1}^{\infty} (AB_i),$$

где $AB_i \cap AB_j = \emptyset$ ($i \neq j$), и, следовательно,

$$P(A) = \sum_{i=1}^{\infty} P(AB_i) = \sum_{i=1}^{\infty} P(B_i) P(A | B_i). \quad (\text{II. 2. 4})$$

Формула (II.2.4) носит название *формулы полной вероятности*. С помощью (II.2.4) мы определяем вероятность события, про которое известно, что оно происходит только одновременно с одним из событий B_i , принадлежащим системе несовместных (непересекающихся) событий (B_1, B_2, \dots) . При этом считаются известными вероятности $P(B_i)$ и условные вероятности $P(A | B_i)$. Переписанная в виде

$$P(B_i | A) = \frac{P(B_i) P(A | B_i)}{\sum_{i=1}^{\infty} P(B_i) P(A | B_i)}, \quad (\text{II. 2. 5})$$

эта формула носит название *формулы Байеса* и служит для переоценки *априорных вероятностей* $P(B_i)$ в предположении, что событие A произошло. В связи с этим $P(B_i | A)$ называется *апостериорной вероятностью* события (гипотезы) B_i .

II.2.2. Случайные величины

Случайной величиной ξ называется функция $\xi(\omega)$, определяемая на множестве Ω и принимающая вещественное значение. В нашем случае этих значений не более чем счетное множество.

Пусть $\{a_1, a_2, \dots\}$ — значения функции $\xi(\omega)$ и $p_i = P(\xi = a_i) = \sum_{\omega: \xi(\omega)=a_i} P(\omega)$ (сумма по всем ω , для которых $\xi(\omega) = a_i$). Таблица

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix} \quad (\text{II. 2. 6})$$

называется *распределением случайной величины* ξ , а функция $F_{\xi}(x)$, задаваемая равенствами

$$F_{\xi}(x) = P(\xi < x) = \sum_{i: a_i < x} p_i, \quad (-\infty < x < \infty), \quad (\text{II. 2. 7})$$

называется *функцией распределения*.

Если ряд $\sum_{\omega \in \Omega} |\xi(\omega)| P(\omega)$ сходится, то сумма

$$E\xi = \sum_{\omega \in \Omega} \xi(\omega) P(\omega) \quad (\text{II. 2. 8})$$

называется *математическим ожиданием случайной величины* ξ .

Если ξ имеет распределение в виде таблицы (II. 2. 6), то

$$E\xi = \sum_{i=1}^{\infty} a_i p_i, \quad (\text{II. 2. 9})$$

где ряд также абсолютно сходится.

Действительно,

$$\sum_{\omega \in \Omega} \xi(\omega) P(\omega) = \sum_{i=1}^{\infty} \left(\sum_{\omega: \xi(\omega)=a_i} \xi(\omega) P(\omega) \right) = \sum_{i=1}^{\infty} \left(a_i \sum_{\omega: \xi(\omega)=a_i} P(\omega) \right) = \sum_{i=1}^{\infty} a_i p_i$$

и аналогично

$$\sum_{\omega \in \Omega} |\xi(\omega)| P(\omega) = \sum_{i=1}^{\infty} |a_i| p_i.$$

Из определения математического ожидания (II. 2. 8) легко следует его линейность: если $a, b \in R^1$ (R^1 — множество всех вещественных чисел) и ξ, η — две случайные величины, математические ожидания которых существуют, то

$$E(a\xi + b\eta) = aE\xi + bE\eta. \quad (\text{II. 2. 10})$$

Пусть $f(x)$ — любая вещественная функция вещественного переменного и ξ — случайная величина с распределением (II. 2. 6). Тогда $f(\xi)$ — это также случайная величина, для которой

$$Ef(\xi) = \sum_{i=1}^{\infty} f(a_i) p_i, \quad (\text{II. 2. 11})$$

если ряд абсолютно сходится. Для доказательства достаточно применить группировку слагаемых так, как это делалось при выводе формулы (II. 2. 9).

Дисперсией $D\xi$ *случайной величины* ξ *называется математическое ожидание случайной величины* $(\xi - E\xi)^2$:

$$D\xi = E(\xi - E\xi)^2. \quad (\text{II. 2. 12})$$

Пользуясь формулой (II. 2. 11), получаем

$$D\xi = \sum_{i=1}^{\infty} (a_i - E\xi)^2 p_i. \quad (\text{II. 2. 13})$$

С другой стороны,

$$D\xi = E(\xi^2 - 2\xi E\xi + (E\xi)^2) = E\xi^2 - 2E\xi \cdot E\xi + (E\xi)^2 = E\xi^2 - (E\xi)^2, \quad (\text{II. 2. 14})$$

где $E\xi$ — постоянная величина, которую мы интерпретируем как случайную величину, т. е. как функцию на Ω , равную $E\xi$ при всех $\omega \in \Omega$.

Число $E\xi$ является центром, вокруг которого группируются значения случайной величины ξ , а $D\xi$ — мера рассеяния значений случайной величины ξ относительно этого центра. Это наиболее

важные характеристики места и рассеяния для одномерных случайных величин.

Случайные величины ξ и η называются независимыми, если для любых $a, b \in R^1$ события $\{\xi < a\}$ и $\{\eta < b\}$ независимы. Нетрудно показать, что для любых множеств $A, B \subset R^1$ события $\{\xi \in A\}$ и $\{\eta \in B\}$ независимы, откуда следует независимость случайных величин $f(\xi)$ и $g(\eta)$ при любых вещественных функциях f и g .

Действительно,

$$P(\{f(\xi) < a\} \cap \{g(\eta) < b\}) = P(\{\xi \in f^{-1}(-\infty, a)\} \cap \{\eta \in g^{-1}(-\infty, b)\}) = \\ = P(\xi \in f^{-1}(-\infty, a)) P(\eta \in g^{-1}(-\infty, b)) = P(f(\xi) < a) P(g(\eta) < b),$$

где f^{-1} и g^{-1} — операции взятия полного прообраза, например $f^{-1}(A) = \{x \in R^1 : f(x) \in A\}$, при этом $f^{-1}(-\infty, a)$, $g^{-1}(-\infty, b) \subset R^1$. Определение независимости двух случайных величин просто обобщается на любое конечное их число. Случайные величины ξ_1, \dots, ξ_n называются взаимно независимыми, если для любых чисел $a_1, \dots, a_n \in R^1$

$$P\left(\bigcap_{i=1}^n \{\xi_i < a_i\}\right) = \prod_{i=1}^n P(\xi_i < a_i). \quad (\text{II. 2. 15})$$

Отсюда проще всего определить взаимную независимость n событий A_1, \dots, A_n .

Пусть $I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \in \bar{A} \end{cases}$ — индикаторная функция множества A .

События A_1, \dots, A_n называются взаимно независимыми (или независимыми в совокупности), если случайные величины, представленные индикаторными функциями I_{A_1}, \dots, I_{A_n} соответствующих множеств, взаимно независимы.

Если случайные величины ξ и η независимы и существуют $E\xi$ и $E\eta$, то существует математическое ожидание произведения $\xi\eta$ и

$$E\xi\eta = E\xi E\eta. \quad (\text{II. 2. 16})$$

Доказательство. Пусть $\{a_1, a_2, \dots\}$ и $\{b_1, b_2, \dots\}$ — значения, принимаемые соответственно случайными величинами ξ и η . Тогда

$$E\xi\eta = \sum_{\omega \in \Omega} \xi(\omega) \eta(\omega) P(\omega) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \left(\sum_{\omega: \xi(\omega)=a_i, \eta(\omega)=b_j} a_i b_j P(\omega) \right) = \\ = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i b_j P(\{\xi = a_i\} \cap \{\eta = b_j\}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i b_j P(\xi = a_i) P(\eta = b_j) = \\ = \sum_{i=1}^{\infty} a_i P(\xi = a_i) \sum_{j=1}^{\infty} b_j P(\eta = b_j) = E\xi \cdot E\eta.$$

Формула (II. 2. 16) доказана.

Очевидно следующее свойство дисперсии: $D(c\xi) = c^2 D\xi$.

Рассмотрим дисперсию суммы двух случайных величин ξ и η :

$$D(\xi + \eta) = E(\xi + \eta - E(\xi + \eta))^2 = E((\xi - E\xi) + (\eta - E\eta))^2 = \\ = E(\xi - E\xi)^2 + E(\eta - E\eta)^2 + 2E(\xi - E\xi)(\eta - E\eta).$$

Величина $E(\xi - E\xi)(\eta - E\eta) = \text{cov}(\xi, \eta)$ называется ковариацией случайных величин ξ и η . Для любых случайных величин имеем $D(\xi + \eta) = D\xi + D\eta + 2\text{cov}(\xi, \eta)$. Для независимых случайных величин, очевидно, $\text{cov}(\xi, \eta) = 0$, поэтому $D(\xi + \eta) = D\xi + D\eta$. Другие характеристики распределений мы рассмотрим после определения случайных величин на пространствах элементарных событий общего вида.

II.3. АКСИОМАТИКА КОЛМОГОРОВА. ИНТЕГРАЛ ЛЕБЕГА

Рассматриваются вероятностные пространства общего вида, а также случайные величины и их основные численные характеристики, большинство которых связано с *интегралом Лебега* от измеримых функций по вероятностной мере.

II.3.1. Вероятностное пространство и случайные величины

Во многих практических задачах, где естественно применять теорию вероятностей, приходится иметь дело с несчетным пространством элементарных событий. Примерами могут служить выбор направления на плоскости (одномерная задача) или в пространстве (двумерная задача), выбор точки в некотором объеме (трехмерная задача), выбор непрерывной функции, заданной на отрезке (бесконечномерная задача), и т. д. Можно было бы попытаться взять в качестве основного вероятностного пространства произвольное множество Ω и задать на классе всех его подмножеств \mathcal{A} функцию множеств P , обладающую приведенными в II.2 свойствами вероятностного распределения. К сожалению, даже в случае простейшего несчетного множества — отрезка единичной длины — невозможно без противоречий задать вероятности всех его подмножеств так, чтобы распределение вероятностей обладало естественными свойствами. Этот факт хорошо известен в теории меры, где приводятся примеры неизмеримых множеств (Колмогоров, Фомин, 1972, с. 248). Значит, в случае произвольного множества класс всех его подмножеств нельзя взять в качестве области задания вероятности. С другой стороны, слишком бедный класс подмножеств, выбранный для этой цели, использовать также непрактично. Во-первых, нельзя выбросить из задачи все многообразие возможностей, которое предполагает несчетное множество элементарных событий. Во-вторых, класс подмножеств можно замкнуть относительно некоторых теоретико-множественных операций, используя свойства вероятности, и, следовательно, можно предположить, что он уже замкнут.

Пусть Ω — множество элементарных событий. *Класс \mathcal{F} подмножеств множества Ω называется полем событий или σ -алгеброй событий*, если он обладает следующими свойствами:

- 1) $\Omega \in \mathcal{F}$;
- 2) если $A \in \mathcal{F}$, то $\bar{A} = \Omega \setminus A \in \mathcal{F}$;

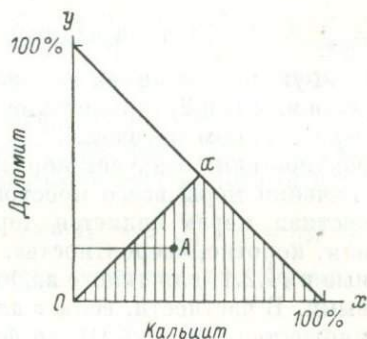
3) если $A_1, A_2, \dots \in \mathcal{F}$, то $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Таким образом, класс \mathcal{F} , на котором разумно задавать вероятность, замкнут относительно дополнения и счетного объединения множеств. Богатство σ -алгебры \mathcal{F} диктуется содержанием задачи. Так, например, при задании вероятности на $\Omega = [0, 1]$ и случайном выборе точки из этого интервала обычно берется класс \mathcal{F} , содержащий все открытые интервалы внутри интервала $[0, 1]$. Этот класс, называемый борелевским, содержит уже такое разнообразие подмножеств множества Ω , которого с большим запасом хватает для всех практических задач.

Пример II.2. Испытанием является извлечение из некоторого статистического ансамбля образца обычных карбонатных пород, в котором определяется химический состав. Исходом будет то или иное процентное содержание кальцита (CaCO_3) и доломита

Рис. II.1. Вероятностное пространство, совпадающее с точками треугольника составов карбонатных пород (кальцит—доломит).

Заштрихованное множество отвечает событию: кальцита больше, чем доломита.



$(\text{Ca}, \text{Mg})(\text{CO}_3)_2$. Пространство Ω элементарных событий здесь может быть взято в виде множества точек равнобедренного прямоугольного треугольника с катетами длиной в 100%, направленными по осям координат (рис. II.1). Декартовы координаты точки, взятой внутри треугольника или на его границе, представляют собой соответственно процентное содержание двух компонент, например кальцита и доломита. Событием в данном случае естественно назвать любое борелевское подмножество множества Ω . Напомним, что σ -алгеброй борелевских подмножеств пространства R^n ($n=1, 2, \dots$) называется наименьшая σ -алгебра, содержащая все n -мерные параллелепипеды (для $n=1$ — это отрезки, для $n=2$ — прямоугольники). Событием будет, например, подмножество A , выделенное условием: «кальцита не менее, чем доломита». Этому событию соответствует множество точек (x, y) треугольника, для которых $x \geq y$. Для решения задач, когда используется только это условие или дополнительное условие «кальцита меньше, чем доломита», достаточно использовать «бедную» σ -алгебру, состоящую из четырех событий $\{A, \bar{A}, \Omega, \emptyset\}$.

Если раньше мы называли событием любое подмножество $A \subset \Omega$, то теперь событиями мы назовем лишь элементы σ -алгебры \mathcal{F} , которая в большинстве случаев не совпадает с σ -алгеброй всех

подмножеств множества Ω . Каждому событию $A \in \mathcal{F}$ должна соответствовать некоторая вероятность (значение вероятности). В II.2.1 мы определяли это значение путем суммирования вероятностей элементарных событий, входящих в событие A . Теперь такой путь неприемлем, так как событие A может содержать несчетное множество элементарных событий, а каждое элементарное событие может иметь вероятность нуль. В современной аксиоматической теории вероятностей задание вероятностей на элементах σ -алгебры \mathcal{F} рассматривается как исходное данное, а именно: задается функция множеств P на \mathcal{F} такая, что

$$1) P(A) \geq 0 \text{ для любого } A \in \mathcal{F}, \text{ причем } P(\Omega) = 1;$$

$$2) \text{ если } A_1, A_2, \dots \in \mathcal{F} \text{ и } A_i \cap A_j = \emptyset (i \neq j), \text{ то } P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ (счетная аддитивность вероятности).}$$

Функция, заданная на σ -алгебре подмножеств и подчиняющаяся условиям 1) и 2) (кроме условия нормировки $P(\Omega) = 1$), называется мерой. Таким образом, P — это вероятностная мера. Слово «вероятностная» означает нормированность, т. е. равенство единице значения меры всего пространства. Синонимом термина «вероятностная мера» является термин «вероятностное распределение» или, коротко, «вероятность». Все свойства вероятности, приведенные в II.2.1, для такого задания вероятности остаются справедливыми. В частности, если σ -алгебра \mathcal{F} содержит одноточечные подмножества $\{\omega\}$ ($\omega \in \Omega$), то функция P задана и на элементарных событиях: $P(\{\omega\}) \geq 0$.

Совокупность пространства элементарных событий Ω , σ -алгебры событий \mathcal{F} и вероятности P , заданной на этой σ -алгебре, т. е. тройка (Ω, \mathcal{F}, P) , называется вероятностным пространством. Совокупность множества и σ -алгебры его подмножеств (в данном случае пара (Ω, \mathcal{F})) называется измеримым пространством (в этом случае не предполагается, что на классе подмножеств данного множества задана вероятность или какая-либо другая мера). В каждой конкретной задаче можно указать более или менее определенный вид этого пространства. Однако при усложнении условия задачи, например при переходе от единичного испытания к последовательности испытаний, приходится строить все новые и новые вероятностные пространства. Иногда предполагают, что существует универсальное вероятностное пространство, относительно которого измеримы все встречающиеся в задаче или в ее возможных обобщениях случайные величины. В теоретических построениях вообще не рассматривают природу вероятностного пространства, так как основной интерес представляют распределения случайных величин, заданных на этом пространстве. Система исходных понятий теории вероятностей, при которой событиям поставлены в соответствие подмножества некоторого множества Ω , а вероятностям событий — вероятностная (нормированная) мера на классе подмножеств,

подчиняющаяся определенным правилам, называется *аксиоматической Колмогорова* (Колмогоров, 1974). В соответствии с вынужденным ограничением области задания вероятности σ -алгеброй \mathcal{F} ограничивается и класс функций $\xi(\omega)$, которые называются случайными величинами. *Случайной величиной называют функцию ξ , заданную на Ω , принимающую вещественные значения и такую, что для любого $a \in R^1$ множество $\{\xi < a\}$ является событием, т. е. $\{\xi < a\} \in \mathcal{F}$ (на языке теории меры это означает, что ξ — функция, измеримая относительно σ -алгебры \mathcal{F} (Колмогоров, Фомин, 1972, с. 264)).*

Множество $\{\xi < a\}$ можно иначе записать в виде $\xi^{-1}(-\infty, a)$ — полный прообраз интервала $(-\infty, a)$ при отображении ξ . Нетрудно доказать, что для случайной величины ξ событием является также множество $\xi^{-1}B$, где B — любое борелевское множество на прямой.

Пусть $F_\xi(x) = P(\xi < x)$ — функция распределения случайной величины ξ . Очевидно, F_ξ — монотонная функция, у которой $\lim_{x \rightarrow -\infty} F_\xi(x) = 0$ и $\lim_{x \rightarrow +\infty} F_\xi(x) = 1$. Практический интерес представляют два частных вида функции распределения.

а). **Дискретный случай.** В этом случае функция F_ξ растет скачками или, другими словами, существует не более чем счетный набор

вещественных чисел $\{a_1, a_2, \dots\}$ такой, что $\sum_{i=1}^{\infty} P(\xi = a_i) = 1$.

В этом случае, как было отмечено выше, распределение случайной величины задается таблицей

$$\begin{pmatrix} a_1, a_2, \dots \\ p_1, p_2, \dots \end{pmatrix},$$

где $p_i = P(\xi = a_i)$.

б). **Абсолютно непрерывный случай.** В этом случае существует плотность распределения случайной величины ξ . *Плотностью называется функция $p_\xi(x)$ такая, что для любого $x \in R^1$ $F_\xi(x) =$*

$= \int_{-\infty}^x p_\xi(t) dt$, при этом $F_\xi(x)$ непрерывна на всей прямой. В частности, если в точке x непрерывна сама плотность p_ξ , то

$$p_\xi(x) = \frac{dF_\xi(x)}{dx}.$$

В геологической практике используются также распределения смешанного типа, при котором $F_\xi(x) = \alpha F_\xi^{(1)}(x) + (1 - \alpha) F_\xi^{(2)}(x)$ ($0 < \alpha < 1$) и $F_\xi^{(1)}$ — дискретная, а $F_\xi^{(2)}$ — абсолютно непрерывная функция распределения.

Пример II.3. Пусть в данном статистическом ансамбле, состоящем из образцов карбонатной породы, содержатся образцы чистого доломита, чистого известняка (кальцита) и образцы всевозможных смесей доломита с известняком. Типичное распреде-

ление вероятностей, характеризующее этот статистический ансамбль, сосредоточенное на отрезке $[0, 100]$ (по числу процентов доломита в образце), имеет две точки 0 и 100 с положительными значениями вероятности $P(C)$ и $P(D)$, а на интервале $(0, 100)$ — абсолютно непрерывное распределение вероятности со значением на этом интервале $1 - P(C) - P(D)$. Функцию распределения в настоящем примере можно представить в виде

$$F_{\xi}(x) = (P(C) + P(D)) F_{\xi}^{(1)}(x) + (1 - P(C) - P(D)) F_{\xi}^{(2)}(x),$$

где $F_{\xi}^{(1)}(x)$ — функция распределения дискретного распределения с таблицей

$$\left(\begin{array}{cc} 0, & 100 \\ \frac{P(C)}{P(C) + P(D)}, & \frac{P(D)}{P(C) + P(D)} \end{array} \right),$$

а $F_{\xi}^{(2)}(x)$ — функция распределения абсолютно непрерыв-

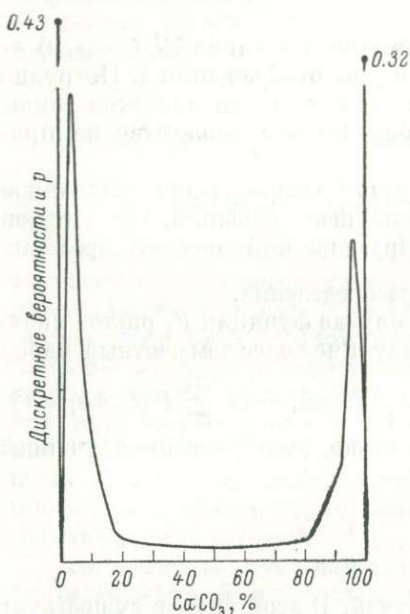


Рис. П.2. Распределение смешанного типа. Функция распределения процентного содержания CaCO_3 в чистой карбонатной породе в нижнекаменноугольных отложениях востока Русской платформы.

Жирные прямые — ординаты вероятностей дискретного распределения, их концы указаны золотыми кружками, цифры около них — вероятности; тонкая кривая — плотность непрерывного распределения.

ного распределения с некоторой плотностью $p_{\xi}(x)$ (рис. П.2).

Прежде чем перейти к примерам распределений, покажем, как строится интеграл, называемый *интегралом Лебега*, от измеримой функции, определенной на пространстве с мерой. В теории вероятностей интеграл Лебега служит для выражения математического ожидания случайной величины в форме, общей для разных видов распределений, и главное — для доказательства различных соотношений, использующих усреднение по вероятностной мере.

П.3.2. Интеграл Лебега

Функция $\xi(\omega)$, определенная на пространстве Ω , называется простой, если пространство Ω можно представить в виде счетной суммы

$$\Omega = \bigcup_{i=1}^{\infty} A_i, \quad A_i \cap A_j = \emptyset \quad (i \neq j)$$

попарно не пересекающихся множеств $A_i \in \mathcal{F}$, причем на каждом A_i функция $\xi(\omega)$ принимает постоянное значение

$$\xi(\omega) = c_i \quad (\text{если } \omega \in A_i).$$

Очевидно, простая функция $\xi(\omega)$ может быть представлена в виде

$$\xi(\omega) = \sum_{i=1}^{\infty} c_i I_{A_i}(\omega). \quad (\text{II. 3. 1})$$

Интегралом $\int_{\Omega} \xi(\omega) P(d\omega)$ от простой функции $\xi(\omega)$ по мере P называется выражение

$$\int_{\Omega} \xi(\omega) P(d\omega) = \sum_{i=1}^{\infty} c_i P(A_i),$$

если ряд в правой части абсолютно сходится.

Построение интеграла в общем случае разобьем на несколько этапов.

1) Если простая функция $\xi(\omega)$ допускает кроме представления (II. 3. 1) другое представление

$$\xi(\omega) = \sum_{j=1}^{\infty} d_j I_{B_j}(\omega),$$

где $B_j \cap B_k \neq \emptyset$ ($j \neq k$) и $\Omega = \bigcup_{j=1}^{\infty} B_j$, то

$$\sum_{i=1}^{\infty} c_i P(A_i) = \sum_{j=1}^{\infty} d_j P(B_j),$$

т. е. значение интеграла Лебега зависит только от самой функции $\xi(\omega)$, а не от выбора ее представления в виде ряда.

Доказательство. Положим $D_{i,j} = A_i B_j$ и заметим, что на множестве $D_{i,j}$ значение $f_{i,j}$ функции $\xi(\omega)$ равно сразу c_i и d_j . Поэтому имеем

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{i,j} P(D_{i,j}) = \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} f_{i,j} P(D_{i,j}) \right) = \sum_{i=1}^{\infty} \left(c_i \sum_{j=1}^{\infty} P(A_i B_j) \right) = \sum_{i=1}^{\infty} c_i P(A_i),$$

так как

$$A_i = \bigcup_{j=1}^{\infty} (A_i B_j),$$

где $A_i B_j \cap A_i B_k = \emptyset$ ($j \neq k$).

Аналогично доказывается, что

$$\sum_{j=1}^{\infty} d_j P(B_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{i,j} P(D_{i,j}).$$

Утверждение 1) доказано.

2). Если $\xi(\omega)$ и $\eta(\omega)$ — простые функции, причем существуют интегралы

$$\int_{\Omega} \xi(\omega) P(d\omega) \text{ и } \int_{\Omega} \eta(\omega) P(d\omega),$$

то $\xi(\omega) + \eta(\omega)$ — простая функция и

$$\int_{\Omega} (\xi(\omega) + \eta(\omega)) P(d\omega) = \int_{\Omega} \xi(\omega) P(d\omega) + \int_{\Omega} \eta(\omega) P(d\omega).$$

Доказательство. Утверждение очевидно, если при любом $\omega \in \Omega$

$$\xi(\omega) = \sum_{i=1}^{\infty} c_i I_{A_i}(\omega) \text{ и } \eta(\omega) = \sum_{j=1}^{\infty} d_j I_{B_j}(\omega).$$

Если, например, $\eta(\omega) = \sum_{j=1}^{\infty} d_j I_{B_j}(\omega)$, где $(B_1, B_2, \dots) \neq (A_1, A_2, \dots)$, то для доказательства утверждения представим обе функции в виде линейной комбинации индикаторов множеств $D_{ij} = A_i B_j$, как это делалось в 1).

3). Если $\xi(\omega)$ — простая функция, для которой существует интеграл $\int_{\Omega} \xi(\omega) P(d\omega)$, то

$$\left| \int_{\Omega} \xi(\omega) P(d\omega) \right| \leq \sup_{\omega \in \Omega} |\xi(\omega)|.$$

Доказательство очевидно.

4). Если последовательность простых функций (ξ_1, ξ_2, \dots) равномерно сходится к случайной величине ξ , то последовательность интегралов

$$\left(\int_{\Omega} \xi_1(\omega) P(d\omega), \int_{\Omega} \xi_2(\omega) P(d\omega), \dots \right)$$

фундаментальна в смысле Коши (Колмогоров, Фомин, 1972, с. 62), т. е.

$$\left| \int_{\Omega} \xi_n(\omega) P(d\omega) - \int_{\Omega} \xi_m(\omega) P(d\omega) \right| \rightarrow 0 \text{ (при } n, m \rightarrow \infty \text{)}.$$

Доказательство следует из оценки

$$\begin{aligned} & \left| \int_{\Omega} \xi_n(\omega) P(d\omega) - \int_{\Omega} \xi_m(\omega) P(d\omega) \right| = \\ & = \left| \int_{\Omega} (\xi_n(\omega) - \xi_m(\omega)) P(d\omega) \right| \leq \sup_{\omega \in \Omega} |\xi_n(\omega) - \xi_m(\omega)|, \end{aligned}$$

вытекающей из 2) и 3).

5). Для любой измеримой функции $\xi(\omega)$ существует последовательность $\xi_n(\omega)$ простых функций, сходящаяся к $\xi(\omega)$ равномерно.

Доказательство. Положим для каждого натурального n и целого k

$$A_{n, k} = \left\{ \omega : \frac{k}{n} \leq \xi(\omega) < \frac{k+1}{n} \right\} = \xi^{-1} \left[\frac{k}{n}, \frac{k+1}{n} \right) \in \mathcal{A},$$

а также

$$\xi_n(\omega) = \sum_{k=-\infty}^{\infty} \frac{k}{n} I_{A_{n, k}}(\omega).$$

Тогда для любого $\omega \in \Omega$

$$|\xi_n(\omega) - \xi(\omega)| \leq \frac{1}{n}.$$

Утверждение доказано.

6). Пусть $(\xi_n)^\infty$ — любая последовательность простых функций, сходящаяся к случайной величине ξ равномерно. Положим

$$\int_{\Omega} \xi(\omega) P(d\omega) = \lim_{n \rightarrow \infty} \int_{\Omega} \xi_n(\omega) P(d\omega).$$

Если существует предел справа, то он не зависит от выбора последовательности (ξ_n) . Назовем эту величину интегралом Лебега от функции ξ по мере P . Очевидно, что все интегралы $\int_{\Omega} \xi_n(\omega) P(d\omega)$, начиная с достаточно большого n , одновременно существуют или не существуют. В последнем случае говорят, что $\int_{\Omega} \xi(\omega) P(d\omega)$ не существует.

7). Пусть $\xi(\omega)$ и $\eta(\omega)$ — случайные величины, $a, b \in R^1$ и интегралы $\int_{\Omega} \xi(\omega) P(d\omega)$ и $\int_{\Omega} \eta(\omega) P(d\omega)$ существуют. Тогда существует интеграл

$$\int_{\Omega} (a\xi(\omega) + b\eta(\omega)) P(d\omega) = a \int_{\Omega} \xi(\omega) P(d\omega) + b \int_{\Omega} \eta(\omega) P(d\omega).$$

Доказательство вытекает из 2) и определения 6).

Мы определили интеграл Лебега от измеримой функции $\xi(\omega)$ по мере P . В теории вероятностей измеримая функция носит название случайной величины, а интеграл от нее по вероятностной мере — математического ожидания случайной величины.

Таким образом, $E\xi = \int_{\Omega} \xi(\omega) P(d\omega)$ — математическое ожидание ξ .

Если этот интеграл не существует, т. е. расходится интеграл $\int_{\Omega} |\xi(\omega)| P(d\omega)$, то говорят, что случайная величина не имеет математического ожидания.

Пусть f — некоторая измеримая вещественная функция, заданная на R^1 .

а). Если случайная величина ξ имеет дискретное распределение, т. е., согласно определению, является простой функцией, то мы имеем

$$Ef(\xi) = \int_{\Omega} f(\xi(\omega)) P(d\omega) = \sum_{i=1}^{\infty} f(a_i) P(\xi = a_i),$$

где (a_1, a_2, \dots) — значения, принимаемые случайной величиной ξ .

б). Если случайная величина ξ имеет плотность распределения $p_{\xi}(x)$, которую естественно считать неотрицательной функцией, а f_n — простая функция, принимающая значения c_1, c_2, \dots , то

$$\begin{aligned} Ef_n(\xi) &= \sum_{i=1}^{\infty} c_i P(f_n(\xi) = c_i) = \sum_{i=1}^{\infty} c_i \int_{f_n^{-1}(c_i)} p_{\xi}(x) dx = \\ &= \sum_{i=1}^{\infty} \int_{f_n^{-1}(c_i)} f_n(x) p_{\xi}(x) dx = \int_{-\infty}^{\infty} f_n(x) p_{\xi}(x) dx. \end{aligned}$$

Путем предельного перехода доказываем, что в этом случае для произвольной измеримой функции f

$$Ef(\xi) = \int_{-\infty}^{\infty} f(x) p_{\xi}(x) dx.$$

В частности, $E\xi = \int_{-\infty}^{\infty} x p_{\xi}(x) dx$, если $\int_{-\infty}^{\infty} |x| p_{\xi}(x) dx < \infty$.

В одномерном случае варианты математического ожидания а) и б) единообразно записывают также в виде интеграла Стильтьеса по неубывающей функции $F_{\xi}(x)$,

$$Ef(\xi) = \int_{-\infty}^{\infty} f(x) dF_{\xi}(x),$$

который имеет смысл для функции распределения $F_{\xi}(x)$ общего вида. Не вдаваясь в теорию интегралов Стильтьеса, заметим лишь, что в случае, когда мы имеем смешанный тип распределения, этот интеграл распадается на два, из которых первый сводится к ряду, а второй — к интегрированию в обычном смысле (в смысле Римана). Этот случай охватывает практически все возможные варианты.

П.3.3. Числовые характеристики случайных величин

Полная информация о случайной величине содержится в ее функции распределения. Однако в том случае, когда эта функция не имеет достаточно простого аналитического вида, эта информа-

ция трудно обозрима. В некоторых случаях хорошее представление о распределении случайной величины дают различные числовые характеристики этих распределений. Важнейшая из этих характеристик — математическое ожидание — упоминалась выше:

$$E\xi = \int_{\Omega} \xi(\omega) P(d\omega) = \int_{-\infty}^{\infty} x dF_{\xi}(x).$$

Кроме того, как и для дискретного вероятностного пространства, в общем случае характеристикой рассеяния случайной величины относительно естественного центра может служить дисперсия $D\xi = E(\xi - E\xi)^2$. Величина $\sigma = \sqrt{D\xi}$ называется *среднеквадратичным, или стандартным, отклонением*. Основные свойства математического ожидания и дисперсии перечислены в II.2. Перейдем к другим характеристикам случайных величин.

а). *Моментом* k -го порядка случайной величины ξ (или ее распределения) называется величина $E\xi^k$ ($k=1, 2, \dots$). Эти моменты называются также начальными.

б). *Центральным моментом* k -го порядка случайной величины ξ (или ее распределения) называется величина $E(\xi - E\xi)^k$ ($k=2, 3, \dots$). Второй центральный момент — это дисперсия.

Линейные функции от третьего и четвертого центральных моментов носят специальные названия:

$\frac{E(\xi - E\xi)^3}{\sigma^3}$ — *асимметрия* (смысл характеристики отражен в названии); $\frac{E(\xi - E\xi)^4}{\sigma^4} - 3$ — *эксцесс* (характеризует превышение «крутизны» функции распределения по отношению к «нормальной»; о нормальном распределении и его свойствах будет сказано далее).

в). *Абсолютные моменты* случайной величины:

$E|\xi|$ — *начальный*, используется для проверки условия существования математического ожидания;

$E|\xi - E\xi|$ — *центральный*, иногда используется для характеристики рассеяния; в аналитическом отношении менее удобен, чем дисперсия.

г). Дискретное распределение может быть охарактеризовано наиболее вероятным значением случайной величины, называемым *модой*

$$M_0(\xi) = a,$$

где случайная величина ξ принимает значения из множества $\{a_1, a_2, \dots\}$ и $P(\xi = a) = \max_{1 \leq i < \infty} P(\xi = a_i)$.

Модой абсолютно непрерывного распределения называется точка x , при которой плотность распределения $p_{\xi}(x)$ достигает максимума. Могут быть одномодальные распределения, двухмодальные, трехмодальные и т. д.

д). *Квантиль* x_p уровня p , где $0 < p < 1$, — это такое значение $x \in R^1$, при котором функция распределения принимает значение p , т. е. $F_\xi(x_p) = p$.

Квантиль уровня $1/2$ называется *медианой* $Me(\xi) = x$, если $F_\xi(x) = 1/2$.

Квантили уровней $1/4$ и $3/4$ называются *нижней* и *верхней квартилями*. Они широко использовались в седиментологии после работ Траска.

е). *Энтропией дискретного распределения* называется величина $H_1\xi = -\sum_{i=1}^{\infty} p_i \log p_i$, где ξ имеет распределение $\begin{pmatrix} a_1 & a_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$.

Энтропия абсолютно непрерывного распределения — это величина $H_2\xi = -\int_{-\infty}^{\infty} p_\xi(x) \log p_\xi(x) dx$, где $p \log p = 0$, если $p = 0$.

Обе эти величины могут служить мерой рассеяния распределений случайных величин. Для упрощения операций со случайными величинами часто вводятся специальные технические средства. К ним относятся производящая, характеристическая функции и преобразование Лапласа.

ж). Пусть случайная величина ξ принимает неотрицательные целые значения с вероятностями (p_0, p_1, p_2, \dots) , где $\sum_{i=0}^{\infty} p_i = 1$.

При любом s ($0 \leq s \leq 1$) определена величина $Es^\xi = \sum_{k=0}^{\infty} s^k p_k$. Функция $f_\xi(s) = Es^\xi$, заданная на интервале $[0, 1]$, называется *производящей функцией случайной величины* ξ . Она содержит полную информацию о распределении случайной величины ξ и в принципе, зная ее, можно восстановить распределение. В терминах производящих функций некоторые преобразования случайных величин выглядят особенно просто.

з). Для неотрицательных, но не обязательно целочисленных случайных величин роль производящей функции играет *преобразование Лапласа*, которое для любого $\lambda \geq 0$ определяется как величина

$$Ee^{-\lambda\xi} = \int_0^{\infty} e^{-\lambda x} dF_\xi(x).$$

Ее свойства аналогичны свойствам производящей функции.

и). Для случайных величин с произвольным распределением на оси R^1 , когда преобразование Лапласа может быть не определено, полную информацию о распределении содержит комплексная функция

$$Ee^{it\xi} = \int_{-\infty}^{\infty} e^{itx} dF_\xi(x) \quad (t \in R^1),$$

называемая *характеристической функцией распределения* (или случайной величины). Это есть не что иное, как преобразование Фурье распределения случайной величины ξ (Колмогоров, Фомин, 1972, с. 397). Так же как для производящей функции и преобразования Лапласа, соотношение между распределениями случайных величин и их характеристическими функциями взаимно однозначно и непрерывно. Свойства преобразования Фурье аналогичны свойствам преобразования Лапласа. Характеристические функции — это основной инструмент для доказательства предельных теорем теории вероятностей и, в частности, знаменитой центральной предельной теоремы.

П.4. ПРИМЕРЫ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ВЕЛИЧИН

Приведем несколько примеров часто встречающихся распределений случайных величин. Подробные сведения об одномерных распределениях содержатся в книге Кендалла и Стьюарта (1966).

П.4.1. Дискретные распределения

а. Биномиальное распределение. Так называется дискретное распределение, зависящее от двух параметров p ($0 < p < 1$) и N ($N=2, 3, \dots$) и сосредоточенное на целых числах $\{0, 1, 2, \dots, N\}$, причём

$$P(\xi = k) = C_N^k p^k (1-p)^{N-k} \quad (0 \leq k \leq N).$$

Биномиальное распределение появляется в схеме независимых испытаний с одной и той же вероятностью «успеха» p при i -том испытании ($i=1, 2, \dots$; так называемая схема Бернулли) и представляет собой распределение числа «успехов» при N испытаниях. Выводится биномиальное распределение с помощью формул сложения и умножения вероятностей.

Для этого случайную величину ξ представим в виде

$$\xi = \sum_{i=1}^N I_{A_i},$$

где A_i — событие, состоящее в том, что в i -том испытании был «успех»; I_{A_i} — индикаторная функция множества A_i .

Тогда $\{\xi = k\} = U \{I_{A_i} = 1 \text{ при } i \in C, I_{A_i} = 0 \text{ при } i \notin C\}$, где объединение ведётся по всем сочетаниям C из N элементов по k . Так как события, входящие в объединение, несовместимы, то по теореме сложения вероятность $P(\xi = k)$ равна сумме вероятностей этих событий, каждая из которых по теореме умножения равна $p^k (1-p)^{N-k}$. Например,

$$\begin{aligned} P(I_{A_1} = 1, \dots, I_{A_k} = 1, I_{A_{k+1}} = 0, \dots, I_{A_N} = 0) &= \\ = P(I_{A_1} = 1) \dots P(I_{A_k} = 1) P(I_{A_{k+1}} = 0) \dots P(I_{A_N} = 0) &= p^k (1-p)^{N-k}, \end{aligned}$$

так как случайные величины (I_{A_i}) ($i=1, \dots, N$) независимы в совокупности, так же как и для других сочетаний из N по k .

Отсюда получаем

$$P(\xi = k) = C_N^k p^k (1-p)^{N-k},$$

где C_N^k — число сочетаний из N по k .

Примеры биномиальных распределений часто встречаются в геологической практике. При анализе смеси зерен в шлихе отбирается некоторое число зерен шлиха, среди которых регистрируются зерна данного минерала. При соблюдении условий независимости и однородности отбора распределение числа зерен данного минерала хорошо описывается биномиальным законом. Биномиальное распределение используется также при оценке распространенности энантиоморфных типов кварца (Вистелиус, 1950).

Представляя случайную величину ξ в виде суммы независимых случайных величин, подсчитаем ее математическое ожидание и дисперсию:

$$E\xi = E \sum_{i=1}^N I_{A_i} = \sum_{i=1}^N E I_{A_i} = Np,$$

так как

$$P(I_{A_i} = 1) = p \text{ и } E I_{A_i} = p \cdot 1 + (1-p) \cdot 0 = p,$$

$$D\xi = \sum_{i=1}^N D I_{A_i} = Np(1-p),$$

так как

$$D I_{A_i} = E(I_{A_i} - p)^2 = (1-p)^2 \cdot p + (-p)^2 \cdot (1-p) = p(1-p).$$

В первом случае мы использовали свойство линейности математического ожидания, а во втором — свойство дисперсии суммы независимых случайных величин.

Вычислим значение моды биномиального распределения. Для этого сравним значения вероятностей для двух соседних значений — k и $k+1$:

$$\frac{p(\xi = k)}{p(\xi = k+1)} = \frac{C_N^k p^k (1-p)^{N-k}}{C_N^{k+1} p^{k+1} (1-p)^{N-k-1}} = \frac{1-p}{p} \cdot \frac{k+1}{N-k}.$$

Это отношение < 1 при $k < p(N+1) - 1$ и > 1 при $k > p(N+1) - 1$. Если не существует такого целого k , которое было бы в точности равно $p(N+1)$, то распределение обладает одной модой:

$$Mo \xi = k_1,$$

где k_1 — наименьшее среди тех k , для которых выполняется неравенство $k > p(N+1) - 1$ (или $k_1 = k_0 + 1$, где k_0 — наибольшее среди тех k , для которых выполняется неравенство $k < p(N+1) - 1$).

Если существует целое k_1 , для которого выполняется равенство $k_1 = p(N+1) - 1$, то распределение обладает двумя модами:

$$Mo_1 \xi = k_1 \text{ и } Mo_2 \xi = k_1 + 1.$$

Легко подсчитать производящую функцию биномиального распределения:

$$\text{для } 0 \leq s \leq 1 \quad \text{Es}^\xi = (\text{Es})^{\sum_{i=1}^N I_{A_i}} = \prod_{i=1}^N \text{Es}^{I_{A_i}} = (1 - p + ps)^N,$$

так как

$$\text{Es}^{I_{A_i}} = s^1 \cdot p + s^0 \cdot (1 - p) = 1 - p + ps.$$

б. Распределение Пуассона. Так называется распределение, сосредоточенное на неотрицательных целых числах, причем

$$P(\xi = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (k = 0, 1, 2, \dots),$$

где λ ($\lambda > 0$) — единственный параметр распределения. Распределение Пуассона можно рассматривать как предельное для последовательности биномиальных распределений с параметрами (p_n, N_n) .

Пусть $N_n \rightarrow \infty$ и $p_n \cdot N_n \rightarrow \lambda$ при $n \rightarrow \infty$.

Тогда для любого целого k

$$C_{N_n}^k p_n^k (1 - p_n)^{N_n - k} \rightarrow \frac{e^{-\lambda} \lambda^k}{k!}.$$

Этот факт можно доказать непосредственно (Гнеденко, 1961, с. 100). Но проще его установить, используя непрерывность соответствия между распределениями и их производящими функциями (Боровков, 1972, с. 121). Для распределения Пуассона имеем

$$\text{Es}^\xi = \sum_{k=0}^{\infty} s^k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda(1-s)}.$$

Пусть ξ_n — случайная величина, имеющая биномиальное распределение с параметрами p_n, N_n .

Пусть $p_n N_n = \lambda_n$, где $\lambda_n \rightarrow \lambda$ при $n \rightarrow \infty$.

Тогда

$$\text{Es}^{\xi_n} = (1 - p_n + p_n s)^{N_n} = \left(1 - \frac{(1-s)\lambda_n}{N_n}\right)^{N_n} \rightarrow e^{-\lambda(1-s)} \quad \text{при } n \rightarrow \infty.$$

Из сходимости производящих функций следует сходимость функций распределения. Распределение Пуассона имеет моменты всех порядков. Первые два момента:

$$\text{E}\xi = \sum_{k=1}^{\infty} k \frac{e^{-\lambda} \lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} = \lambda,$$

$$\text{E}\xi^2 = \sum_{k=1}^{\infty} k^2 \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k=1}^{\infty} k(k-1) \frac{e^{-\lambda} \lambda^k}{k!} + \lambda = \lambda^2 \sum_{k=2}^{\infty} \frac{e^{-\lambda} \lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2 + \lambda,$$

откуда

$$D\xi = \text{E}\xi^2 - (\text{E}\xi)^2 = \lambda.$$

Мода распределения Пуассона находится из условия $Mo(\xi) = k_1$, где k_1 — наименьшее из тех k , для которых $\frac{P(\xi = k)}{P(\xi = k + 1)} \leq 1$, т. е. $k \leq \lambda - 1$. При $k_1 = \lambda - 1$ распределение имеет две моды:

$$Mo \xi = \lambda - 1 \text{ и } Mo \xi = \lambda.$$

Наиболее естественно распределение Пуассона возникает в так называемых точечных процессах Пуассона. Рассмотрим, например, однородный точечный процесс Пуассона на прямой. При этом по всей прямой случайным образом размещены отмеченные точки так, что при любом n для любой системы n попарно непересекающихся интервалов A_1, A_2, \dots, A_n (где $A_i \cap A_j = \emptyset, A_i \subset R_1$) количества отмеченных точек $(\xi_{A_1}, \dots, \xi_{A_n})$ в этих интервалах — это взаимно независимые случайные величины, причем распределение случайной величины ξ_A зависит только от длины интервала A , а не от его положения на прямой.

Если к тому же распределение случайной величины ξ_A подчинено условию *ординарности* (смысл которого состоит в том, что для малой длины интервала вероятность попадания в этот интервал более одной точки есть величина более высокого порядка малости), то ξ_A имеет распределение Пуассона с параметром $a \cdot |A|$, где $a > 0$ и $|A|$ — длина интервала A (Гнеденко, 1964, с. 295).

Аналогично определяется многомерный точечный процесс Пуассона. В геологической практике точечные процессы Пуассона могут служить моделью многих явлений. Показано, например, что распределение количества зерен акцессорного минерала в шлихах из песчаных отложений с большой степенью точности соответствует распределению Пуассона (Вистелиус, 1968). Вместе со свойствами независимости количеств зерен в непересекающихся пробах это приводит к модели трехмерного точечного процесса Пуассона, где отмеченными точками являются зерна акцессорного минерала.

в. Геометрическое распределение. Это дискретное распределение, сосредоточенное на множестве неотрицательных целых чисел, причем

$$P(\xi = k) = (1 - \alpha) \alpha^k \quad (k = 0, 1, 2, \dots),$$

где α ($0 < \alpha < 1$) — параметр распределения. Этот параметр равен условной вероятности $P(\xi \geq k + 1 | \xi \geq k)$, которая не зависит от k .

Действительно, по определению условной вероятности (см. II.2.4),

$$P(\xi \geq k + 1 | \xi \geq k) = \frac{P(\{\xi \geq k + 1\} \cap \{\xi \geq k\})}{P(\xi \geq k)} = \frac{P(\xi \geq k + 1)}{P(\xi \geq k)} = \frac{\sum_{i=k+1}^{\infty} (1 - \alpha) \alpha^i}{\sum_{i=k}^{\infty} (1 - \alpha) \alpha^i} = \alpha.$$

Геометрическое распределение возникает в рассматриваемых далее марковских цепях в качестве распределения длины серии

повторяющихся элементов марковских цепей. Частным видом марковских цепей является упоминавшаяся выше последовательность независимых испытаний. Длина серии «успешных» испытаний до первого «неуспешного» имеет геометрическое распределение:

$$P(\xi = k) = P(I_{A_1} = 1, \dots, I_{A_k} = 1, I_{A_{k+1}} = 0) = \alpha^k (1 - \alpha),$$

где $\alpha = P(I_{A_i} = 1)$ — вероятность «успеха».

Мода геометрического распределения всегда равна нулю. Все его моменты конечны. Для подсчета его первых двух моментов воспользуемся свойством производящей функции.

Если $f_\xi(s)$ — производящая функция, то

$$E\xi = \left. \frac{df_\xi(s)}{ds} \right|_{s=1} \quad \text{и} \quad E\xi^2 - E\xi = \left. \frac{d^2f_\xi(s)}{ds^2} \right|_{s=1}.$$

Действительно, если $f_\xi(s) = \sum_{k=0}^{\infty} s^k p_k$, то

$$\frac{df_\xi(s)}{ds} = \sum_{k=1}^{\infty} k s^{k-1} p_k \quad \text{и} \quad \frac{d^2f_\xi(s)}{ds^2} = \sum_{k=1}^{\infty} k(k-1) s^{k-2} p_k,$$

откуда следуют формулы для $E\xi$ и $E\xi^2$.

Для геометрического распределения

$$f_\xi(s) = \sum_{k=0}^{\infty} s^k (1 - \alpha) \alpha^k = \frac{1 - \alpha}{1 - s\alpha}$$

и, следовательно,

$$E\xi = \left. \frac{d}{ds} \left(\frac{1 - \alpha}{1 - s\alpha} \right) \right|_{s=1} = \frac{\alpha}{1 - \alpha},$$

$$E\xi^2 - E\xi = \left. \frac{d^2}{ds^2} \left(\frac{1 - \alpha}{1 - s\alpha} \right) \right|_{s=1} = \left. \frac{d}{ds} \left(\frac{\alpha(1 - \alpha)}{(1 - s\alpha)^2} \right) \right|_{s=1} = \frac{2\alpha^2}{(1 - \alpha)^2},$$

откуда

$$E\xi^2 = \frac{\alpha(1 + \alpha)}{(1 - \alpha)^2} \quad \text{и} \quad D\xi = E\xi^2 - (E\xi)^2 = \frac{\alpha}{(1 - \alpha)^2}.$$

II.4.2. Абсолютно непрерывные распределения

а. Экспоненциальное распределение.

Это распределение имеет положительная случайная величина ξ , имеющая плотность распределения

$$p_\xi(x) = ae^{-ax} \quad (x > 0),$$

где a ($a > 0$) — параметр распределения. В случае геометрического распределения оно может рассматриваться в качестве предельного. Появляется оно в марковских процессах с непрерывным временем в той же ситуации, что и геометрическое для цепей Маркова. В процессах Пуассона на прямой по экспоненциальному

закону распределены расстояния между отмеченными точками. Важное свойство экспоненциального распределения — отсутствие последдействия.

Отождествим случайную величину ξ , имеющую экспоненциальное распределение, со временем жизни некоторого объекта. Тогда условное распределение остаточного времени жизни при условии, что объект уже прожил время t ($t \geq 0$), не зависит от t и имеет тот же самый вид, что и безусловное (это свойство и называется отсутствием последдействия).

Действительно,

$$P(\xi - t \geq t_1 | \xi \geq t) = \frac{P(\xi \geq t_1 + t)}{P(\xi \geq t)} = \frac{\int_{t_1+t}^{\infty} e^{-ax} dx}{\int_t^{\infty} e^{-ax} dx} = \frac{e^{-a(t_1+t)}}{e^{-at}} = e^{-at_1} = P(\xi \geq t_1).$$

Плотность экспоненциального распределения не имеет максимума на интервале $(0, \infty)$ — она монотонно убывает от a до 0. Легко подсчитать квантили этого распределения. Так как $F_{\xi}(x) = \int_0^x ae^{-at} dt = 1 - e^{-ax}$, то уравнение $F_{\xi}(x) = c$ ($0 < c < 1$) имеет единственный корень $x_c = -\frac{\log(1-c)}{a}$, который и является квантилью уровня c . В частности, $Me(\xi) = \frac{\log 2}{a}$.

Преобразование Лапласа экспоненциального распределения имеет вид:

$$L_{\xi}(\lambda) \equiv \int_0^{\infty} e^{-\lambda t} a e^{-at} dt = \frac{a}{\lambda + a} \quad (\lambda \geq 0).$$

Зная эту функцию, легко вычислить моменты экспоненциального распределения, так как, очевидно,

$$E\xi = -\frac{d}{d\lambda} L_{\xi}(\lambda) |_{\lambda=0} \quad \text{и} \quad E\xi^2 = \frac{d^2}{d\lambda^2} L_{\xi}(\lambda) |_{\lambda=0}.$$

Для экспоненциального распределения

$$E\xi = -\frac{d}{d\lambda} \left(\frac{a}{\lambda + a} \right) \Big|_{\lambda=0} = \frac{a}{(\lambda + a)^2} \Big|_{\lambda=0} = \frac{1}{a},$$

$$E\xi^2 = -\frac{d}{d\lambda} \frac{a}{(\lambda + a)^2} \Big|_{\lambda=0} = \frac{2a}{(\lambda + a)^3} \Big|_{\lambda=0} = \frac{2}{a^2},$$

откуда

$$D\xi = \frac{1}{a^2}.$$

б. Равномерное распределение. Так называют распределение случайной величины ξ , сосредоточенное на интервале (a, b) ($-\infty < a < b < \infty$) с плотностью

$$p_{\xi}(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases}.$$

Равномерное распределение является предельным для суммы по модулю M ($M \neq 0$) независимых и равно распределенных случайных величин при стремлении числа слагаемых к бесконечности.

Пусть, например, угол поворота единичного вектора на плоскости получен в результате n последовательных приращений угла:

$$\varphi_n = \sum_{i=1}^n \alpha_i,$$

где случайные величины α_i независимы и одинаково распределены.

Регистрируется угол $\Phi_n = \varphi_n - 2\pi k$, где $k = \left[\frac{\varphi_n}{2\pi} \right]$ — целая часть

величины $\frac{\varphi_n}{2\pi}$, т. е. $\Phi_n = \sum_{i=1}^n \alpha_i \pmod{2\pi}$ (сумма по модулю M , где

$M=2\pi$). При некоторых очень слабых предположениях о функции распределения слагаемых α_i последовательность распределений случайных величин Φ_n сходится (в некотором смысле) к равномерному распределению на интервале $(0, 2\pi)$. Идея сложения по модулю M большого числа малых случайных величин заложена в основу практической рекомендации — считать ошибку округления, например, с точностью до второго десятичного знака, случайной величиной с равномерным распределением на интервале между двумя допустимыми значениями округления.

Равномерное распределение имеет все моменты и квантили, вычислить которые предлагается читателю.

в. Н о р м а л ь н о е р а с п р е д е л е н и е. Характерируется плотностью распределения случайной величины

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (-\infty < x < \infty),$$

где a ($-\infty < a < \infty$) и σ ($\sigma > 0$) — параметры распределения. Это очень важное распределение в теории вероятностей, имеющее многочисленные применения в естествознании. Нормальное распределение возникает в качестве предельного при сложении большого числа независимых малых случайных величин. Соответствующий точный результат известен под названием *центральной предельной теоремы*.

П у с т ь

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Эта функция называется *интегралом Лапласа*. Известно, что

$$\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}, \text{ т. е. } \Phi(x) \text{ является функцией распределения.}$$

Один из вариантов центральной предельной теоремы состоит в следующем. Функция распределения случайной величины

$$X_n = \frac{\sum_{i=1}^n \xi_i - nE\xi_i}{\sqrt{nD\xi_i}},$$

где $(\xi_n)_1^\infty$ — последовательность взаимно независимых случайных величин с одной и той же функцией распределения и с конечной дисперсией, сходится к функции распределения $\Phi(x)$ при $n \rightarrow \infty$. Существует много других форм центральной предельной теоремы как для независимых, так и для зависимых случайных величин.

Нормальное распределение имеет все моменты. Для их подсчета мы используем свойства характеристических функций. Согласно определению характеристической функции случайной величины ξ с функцией распределения $F_\xi(x)$

$$\varphi_\xi(t) \equiv Ee^{it\xi} = \int_{-\infty}^{\infty} e^{itx} dF_\xi(x),$$

имеем

$$\frac{d^k}{dt^k} \varphi_\xi(t) = i^k \int_{-\infty}^{\infty} x^k e^{itx} dF_\xi(x)$$

(в предположении, что существует момент k -того порядка), откуда

$$E\xi^k = (-i)^k \frac{d^k}{dt^k} \varphi_\xi(t) \Big|_{t=0}.$$

Для нормального распределения

$$\begin{aligned} \varphi_\xi(t) &= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{itx - \frac{(x-a)^2}{2\sigma^2}} dx = \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-a-it\sigma^2)^2}{2\sigma^2} + ait - \frac{\sigma^2 t^2}{2}} dx = \exp\left(ait - \frac{\sigma^2 t^2}{2}\right), \end{aligned}$$

так как $\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-z)^2}{2\sigma^2}} dx = 1$ для любого комплексного z ; более

подробное вычисление см. в работе Боровкова (1972, с. 113).

Отсюда

$$E\xi = (-i) \frac{d\varphi_\xi(t)}{dt} \Big|_{t=0} = (a + i\sigma^2 t) \exp\left(ait - \frac{\sigma^2 t^2}{2}\right) \Big|_{t=0} = a,$$

$$E\xi^2 = \frac{d^2\varphi_\xi(t)}{dt^2} \Big|_{t=0} = (\sigma^2 + a^2 + 2\sigma^2ait - \sigma^4t^2) \exp\left(ait - \frac{\sigma^2t^2}{2}\right) \Big|_{t=0} = \sigma^2 + a^2,$$

$$E\xi^3 = i \frac{d^3\varphi_\xi(t)}{dt^3} \Big|_{t=0} =$$

$$= (a^3 + 3a\sigma^2 + 3\sigma^2(\sigma^2 + a^2)it - 3a\sigma^4t^2 + \sigma^6it^3) \exp\left(ait - \frac{\sigma^2t^2}{2}\right) \Big|_{t=0} = a^3 + 3a\sigma^2,$$

$$E\xi^4 = \frac{d^4\varphi_\xi(t)}{dt^4} \Big|_{t=0} = (a^4 + 6a^2\sigma^2 + 3\sigma^4 + 4a\sigma^2(a^2 + 3\sigma^2)it -$$

$$- 6a^2\sigma^4t^2 - 2a\sigma^6it^3 - \sigma^8t^4) \cdot \exp\left(ait - \frac{\sigma^2t^2}{2}\right) \Big|_{t=0} = a^4 + 6a^2\sigma^2 + 3\sigma^4.$$

Центральные моменты нормального распределения равны соответственно:

$$D\xi = E(\xi - E\xi)^2 = E\xi^2 - (E\xi)^2 = \sigma^2,$$

$$E(\xi - E\xi)^3 = E\xi^3 - 3E\xi^2E\xi + 2(E\xi)^3 = 0,$$

$$E(\xi - E\xi)^4 = E\xi^4 - 4E\xi^3E\xi + 6E\xi^2(E\xi)^2 - 3(E\xi)^4 = 3\sigma^4.$$

Отсюда асимметрия нормального распределения равна 0. Эцкессе также равен нулю:

$$\frac{E(\xi - E\xi)^4}{\sigma^4} - 3 = 0.$$

Как отмечалось выше, нормальное распределение служит образцом, с которым сравниваются все другие одномодальные распределения по степени сглаженности.

Г а м м а - р а с п р е д е л е н и е. Оно определяет распределение положительной случайной величины ξ с плотностью распределения (см. рис. II.3)

$$p_\xi(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (x > 0),$$

где α и β — положительные параметры, а $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$ — гамма-функция. Особенно часто используется частный вид гамма-распределения — распределение χ^2 . Так называется распределение случайной величины $\chi^2 = \xi_1^2 + \dots + \xi_n^2$, где $n=1, 2, \dots$, а (ξ_1, \dots, ξ_n) — независимые случайные величины, имеющие одно и то же нормальное распределение с параметрами $a=0$ и $\sigma=1$. При этом распределение χ^2 — это гамма-распределение с параметрами $\alpha = \frac{n}{2}$ и $\beta = \frac{1}{2}$. Число n — параметр распределения χ^2 — называется *числом степеней свободы*.

Преобразование Лапласа гамма-распределения вычисляется просто:

$$L_{\xi}(\lambda) \equiv Ee^{-\lambda\xi} = \int_0^{\infty} e^{-\lambda x} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx =$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-(\lambda+\beta)x} dx = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{(\lambda+\beta)^{\alpha}} = \left(\frac{\beta}{\lambda+\beta}\right)^{\alpha},$$

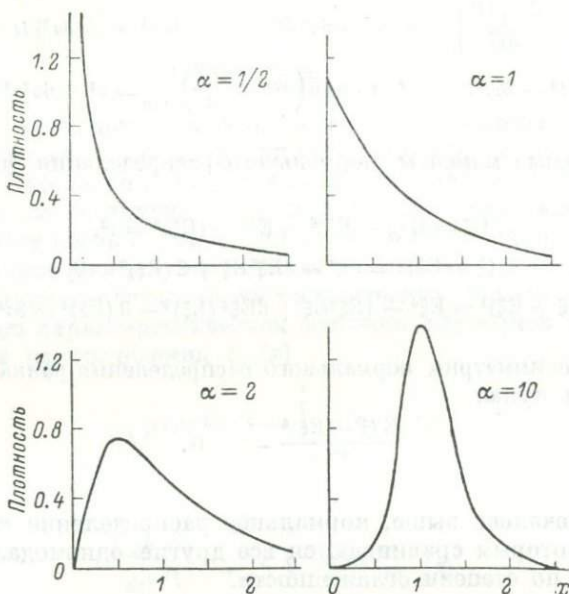


Рис. II.3. Влияние параметра α на форму плотности γ -распределения с математическим ожиданием, равным единице.

откуда

$$E\xi = -\frac{d}{d\lambda} L_{\xi}(\lambda) \Big|_{\lambda=0} = \frac{\alpha\beta^{\alpha}}{(\lambda+\beta)^{\alpha+1}} \Big|_{\lambda=0} = \frac{\alpha}{\beta},$$

$$E\xi^2 = \frac{d^2}{d\lambda^2} L_{\xi}(\lambda) \Big|_{\lambda=0} = \frac{\alpha(\alpha+1)\beta^{\alpha}}{(\lambda+\beta)^{\alpha+2}} \Big|_{\lambda=0} = \frac{\alpha(\alpha+1)}{\beta^2},$$

тогда

$$D\xi = E\xi^2 - (E\xi)^2 = \frac{\alpha}{\beta^2}.$$

Вычислим асимметрию гамма-распределения; имеем

$$E(\xi - E\xi)^3 = E\xi^3 - 3E\xi^2 E\xi + 2(E\xi)^3 = \frac{2\alpha}{\beta^3},$$

откуда

$$\frac{E(\xi - E\xi)^3}{\sigma^3} = 2\alpha^{-1/2}.$$

Асимметрия положительна и тем меньше, чем больше параметр α . Для распределения χ^2 , где $\alpha = n/2$, последнее свойство вытекает также из центральной предельной теоремы, из которой следует, что χ^2 распределено асимптотически нормально при $n \rightarrow \infty$.

д. Распределение Коши. Это распределение случайной величины ξ с плотностью

$$p_{\xi}(x) = \frac{1}{\pi(1+x^2)} \quad (-\infty < x < \infty).$$

Оно совпадает с распределением отношения ξ_1/ξ_2 независимых случайных величин ξ_1 и ξ_2 , распределенных нормально с параметрами $a=0$ и $\sigma=1$. Такое же распределение имеет $\operatorname{tg} \alpha$, где α — случайная величина, равномерно распределенная на отрезке $(-\pi/2, \pi/2)$. Плотность $p_{\xi}(x)$ этого распределения представляет собой симметричную относительно оси ординат функцию, при этом $\operatorname{Mo}(\xi)=0$. Однако математического ожидания не существует, так как

$$\int_{-\infty}^{\infty} \frac{|x| dx}{\pi(1+x^2)} = \frac{2}{\pi} \int_0^{\infty} \frac{x dx}{1+x^2} = \infty.$$

Можно показать, что характеристическая функция этого распределения равна $e^{-|\theta|}$ (Крамер, 1975, с. 272).

Распределением Коши называют также распределение случайной величины $\xi_1 = \lambda \xi + \mu$, где $\lambda > 0$ и $-\infty < \mu < \infty$, плотность которой $p_{\xi_1}(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x - \mu)^2}$.

Таким образом, распределение Коши — это двухпараметрическое распределение, при этом $\operatorname{Mo}_{\xi_1} = \operatorname{Me}_{\xi_1} = \mu$ — естественный центр распределения. Параметр λ характеризует степень рассеяния случайной величины ξ_1 относительно центра и численно равен полуразности верхней и нижней квантилей. Действительно,

$$F_{\xi_1}(x) = \int_{-\infty}^x \frac{1}{\pi} \frac{\lambda dx}{\lambda^2 + (x - \mu)^2} = \frac{1}{\pi} \operatorname{arctg} \frac{x - \mu}{\lambda} + \frac{1}{2}.$$

Если $F_{\xi_1}(x) = \frac{1}{4}$, то $\operatorname{arctg} \frac{x - \mu}{\lambda} = -\frac{\pi}{4}$, т. е. $\frac{x - \mu}{\lambda} = -1$ и $x = \mu - \lambda$ — нижняя квантиль.

Если $F_{\xi_1}(x) = \frac{3}{4}$, то $\operatorname{arctg} \frac{x - \mu}{\lambda} = \frac{\pi}{4}$, т. е. $\frac{x - \mu}{\lambda} = 1$ и $x = \mu + \lambda$ — верхняя квантиль.

Разность между верхней и нижней квантилями равна 2λ .

П.5. ВЕКТОРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Рассматриваются конечномерные векторные случайные величины и связанные с ними понятия: произведение вероятностных пространств, индуцированные вероятностные распределения, мар-

гинальные распределения, характеристики и примеры распределений векторных случайных величин, условные распределения относительно случайных величин.

II.5.1. Произведение вероятностных пространств

Как уже отмечалось, одна из задач теории вероятностей — нахождение вероятностей одних событий по заданным вероятностям других событий. Иногда требуется расширить исходное множество элементарных событий и класс измеримых множеств. Один из методов конструирования нового вероятностного пространства из заданных — *произведение вероятностных пространств*.

Пусть $\{(\Omega_1, \mathcal{F}_1, P_1), \dots, (\Omega_n, \mathcal{F}_n, P_n)\}$ ($n=1, 2, \dots$) — заданная система вероятностных пространств. Их произведением называется вероятностное пространство (Ω, \mathcal{F}, P) , где $\Omega = \Omega_1 \times \dots \times \Omega_n$ — множество всех конечных последовательностей $\omega = (\omega_1, \dots, \omega_n)$ ($\omega_i \in \Omega_i$), называемое декартовым произведением множеств $\Omega_1, \dots, \Omega_n$, \mathcal{F} — наименьшая σ -алгебра, содержащая все n -мерные «параллелепипеды» $A_1 \times A_2 \times \dots \times A_n$ ($A_i \in \mathcal{F}_i$), т. е. декартовы произведения множеств A_1, \dots, A_n . Эту σ -алгебру обозначают $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n$. P — вероятностная мера, заданная на \mathcal{F} , такая, что для любых $A_i \in \mathcal{F}_i$ ($i=1, \dots, n$)

$$P(A_1 \times \dots \times A_n) = \prod_{i=1}^n P_i(A_i).$$

Существование и единственность такой вероятностной меры, которую обозначают $P_1 \times \dots \times P_n$, доказаны, например, в книге Невё (1969, с. 113). Часто все $(\Omega_i, \mathcal{F}_i, P_i)$ ($i=1, \dots, n$) — это экземпляры одного и того же вероятностного пространства

$$(\Omega_i, \mathcal{F}_i, P_i) = (\Omega_1, \mathcal{F}_1, P_1).$$

Так поступают при конструировании вероятностного пространства в схеме Бернулли, когда ищут вероятности событий в последовательности независимых испытаний. В обычных обозначениях множество $A_1 \times \dots \times A_n$ означает событие, при котором происходит и событие A_1 при первом испытании, и событие A_2 при втором испытании, и событие A_n при n -м испытании, т. е. это пересечение событий $\{\pi_i \in A_i\}$ ($i=1, \dots, n$), где $\pi_i(\omega) = \omega_i$ — функция от $\omega = (\omega_1, \dots, \omega_n)$, сопоставляющая каждому ω его i -тую координату. Событие $\{\pi_i \in A_i\}$ состоит в том, что i -тая координата вектора ω принадлежит множеству A_i .

Тогда

$$P(A_1 \times \dots \times A_n) = P\left(\bigcap_{i=1}^n \{\pi_i \in A_i\}\right) = \prod_{i=1}^n P_i(A_i),$$

но, согласно тому же определению,

$$P(\{\pi_i \in A_i\}) = P\left(\bigcap_{j \neq i} \{\pi_j \in \Omega_j\} \cap \{\pi_i \in A_i\}\right) = \prod_{j \neq i} P_j(\Omega_j) \cdot P_i(A_i) = P_i(A_i),$$

так как $P_j(\Omega_j) = 1$.

Следовательно, мера P выбрана так, что события $\{\pi_i \in A_i\}$ независимы при любых A_1, \dots, A_n ($A_i \in \mathcal{F}_i$).

Пусть теперь на пространстве Ω заданы случайные величины ξ_1, \dots, ξ_n так, что ξ_i фактически зависит только от ω_i (так будет, если $\xi_i(\omega)$ представлена в виде $\xi_i(\pi_i(\omega))$, где ξ_i — некоторая случайная величина, заданная на пространстве Ω_i). Тогда случайные величины ξ_i взаимно независимы.

Действительно, для любых a_i

$$\{\xi_i < a_i\} = \{\omega : \xi_i(\pi_i(\omega)) < a_i\} = \pi_i^{-1}\{\xi_i < a_i\} = \{\pi_i \in A_i\},$$

где $A_i = \{\xi_i < a_i\} \in \mathcal{F}_i$. Но события $\{\pi_i \in A_i\}$ взаимно независимы. Следовательно, и ξ_i взаимно независимы.

В теории вероятностей часто приходится иметь дело с последовательностями случайных величин. Предыдущее построение показывает, что всегда можно построить вероятностное пространство, на котором заданы случайные величины, имеющие те же самые распределения, и на котором они будут взаимно независимы (мы показали это для конечных последовательностей, но то же справедливо и для бесконечных). Поэтому часто не заботятся о конкретном виде вероятностного пространства и считают, что на нем можно построить сколь угодно большую систему взаимно независимых случайных величин.

II.5.2. Распределение векторных случайных величин

Мы определили случайные величины как измеримые функции, отображающие измеримое пространство (Ω, \mathcal{F}) в измеримое пространство (R^1, \mathcal{B}^1) , где \mathcal{B}^1 — σ -алгебра борелевских подмножеств на прямой. Это означает, что для случайной величины ξ (ее значения $\xi(\omega)$ называются реализациями случайной величины ξ) полный образ любого борелевского множества $A \in \mathcal{B}^1$, т. е. множество $\xi^{-1}A = \{\omega : \xi(\omega) \in A\}$, принадлежит σ -алгебре \mathcal{F} .

Теперь мы определим векторную случайную величину ξ как измеримое отображение пространства (Ω, \mathcal{F}) в измеримое пространство (R^n, \mathcal{B}^n) , где R^n — обычное n -мерное векторное пространство, а \mathcal{B}^n — σ -алгебра борелевских подмножеств пространства R^n . Это означает, что для любого $A \in \mathcal{B}^n$ множество $\xi^{-1}A$ принадлежит σ -алгебре \mathcal{F} .

Ранее мы говорили о распределениях одномерных случайных величин. Теперь мы уточним это понятие, рассмотрим распределение векторных случайных величин.

Пусть (Ω, \mathcal{F}, P) — вероятностное пространство, на котором задана векторная случайная величина ξ .

Распределением случайной величины ξ называется вероятностная мера μ_ξ , определенная на борелевских подмножествах $A \in \mathcal{B}^n$ равенством

$$\mu_\xi(A) = P(\xi \in A) = P(\xi^{-1}A).$$

Легко проверить, что μ_ξ — это именно вероятность, определенная на пространстве (R^n, \mathcal{B}^n) , так как

1) $\mu_\xi(A) \geq 0$ и $\mu_\xi(R^n) = P(\xi \in R^n) = 1$,

2) если $A_n \in \mathcal{B}^n$ для $n = 1, 2, \dots$ и $A_i \cap A_j = \emptyset$ ($i \neq j$), то

$$\mu_\xi\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\xi \in \bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} \{\xi \in A_n\}\right) = \sum_{n=1}^{\infty} P(\xi \in A_n) = \sum_{n=1}^{\infty} \mu_\xi(A_n),$$

так как

$$\{\xi \in A_i\} \cap \{\xi \in A_j\} = \emptyset \quad (i \neq j).$$

Итак, $(R^n, \mathcal{B}^n, \mu_\xi)$ — это новое вероятностное пространство. Говорят, что случайная величина ξ индуцирует вероятность (или вероятностную меру) μ_ξ на множестве своих значений.

В частности, если $n=1$, т. е. ξ — одномерная случайная величина, то ее распределение μ_ξ — это вероятность, заданная на (R^1, \mathcal{B}^1) , т. е. на множестве вещественных чисел с заданной на нем σ -алгеброй борелевских подмножеств. Мы характеризовали вероятность μ_ξ

а) или таблицей вида

$$\begin{pmatrix} a_1, a_2, \dots \\ p_1, p_2, \dots \end{pmatrix},$$

где $a_i \in R^1$, $p_i \geq 0$ и $\sum_{i=1}^{\infty} p_i = 1$ (в дискретном случае),

б) или плотностью распределения $p_\xi(x)$, где $P(\xi < a) = \int_{-\infty}^a p_\xi(x) dx$, $a \in R^1$ и p_ξ — неотрицательная измеримая функция (в абсолютно непрерывном случае),

в) или функцией распределения (в общем случае)

$$F_\xi(a) = P(\xi < a).$$

В терминах вероятностной меры μ_ξ эти способы задания можно записать в следующем виде:

а) $\mu_\xi\{a_i\} = p_i \quad (i = 1, 2, \dots)$,

б) $\mu_\xi(-\infty, a) = \int_{-\infty}^a p_\xi(x) dx$,

в) $\mu_\xi(-\infty, a) = F_\xi(a)$.

Для случая $n > 1$ основными встречающимися в практике распределениями векторных случайных величин будут распределения трех типов: дискретное, абсолютно непрерывное и смешанное (образованное из двух предыдущих типов).

Дискретное распределение — это случай, когда существует не более чем счетное множество точек $\{a_1, a_2, \dots\}$ ($a_i \in R^n$), на котором сосредоточена мера μ_ξ , т. е. $\mu_\xi\{a_1, a_2, \dots\} = 1$, и, следовательно, распределение может быть задано таблицей

$$\begin{pmatrix} a_1, a_2, \dots \\ \mu_\xi(a_1), \mu_\xi(a_2), \dots \end{pmatrix}.$$

Абсолютно непрерывным распределением называется распределение μ_ξ , для которого существует такая неотрицательная измеримая функция p_ξ , называемая плотностью распределения μ_ξ , что для любого борелевского множества $A \subset R^n$

$$\mu_\xi(A) = \int_A p_\xi(x) dx.$$

В этом выражении справа стоит интеграл Лебега от функции p_ξ по так называемой мере Лебега на множестве A . Мы не будем уточнять понятие «мера Лебега». Отметим только, что n -мерная мера Лебега — это n -мерный объем данного множества. Все практически используемые плотности распределения — это непрерывные или кусочнонепрерывные функции, а для таких функций интеграл Лебега по мере Лебега совпадает с обычным интегралом Римана. Следовательно, если мы будем обозначать точку $x \in R^n$ в обычной координатной форме $x = (x_1, \dots, x_n)$, то предыдущий интеграл, записанный в виде

$$\int_A p_\xi(x_1, \dots, x_n) dx_1 \dots dx_n,$$

можно вычислять, как обычный n -кратный.

Распределение смешанного типа не нуждается в специальных комментариях. Заметим, что полное описание распределения μ_ξ произвольного типа можно дать с помощью n -мерной функции распределения F_ξ . Применяя, как и выше, символ $\pi_i(\xi)$ для обозначения i -той координаты случайного вектора ξ , определим

$$F_\xi(x_1, \dots, x_n) = P\left(\bigcap_{i=1}^n \{\pi_i(\xi) < x_i\}\right) = \mu_\xi((-\infty, x_1) \times (-\infty, x_2) \times \dots \times (-\infty, x_n)).$$

Нетрудно выразить значение меры μ_ξ через F_ξ на n -мерном параллелепипеде. В отличие от одномерного случая выражение получается громоздкое и неудобное для практического использования. Кроме того при преобразовании системы координат функция F_ξ преобразуется чрезвычайно сложно (Тугубалин, 1972, с. 72).

В общем случае n -мерная функция распределения применяется мало. Исключение составляет случай, когда координаты случайного вектора являются взаимно независимыми случайными величинами. Тогда

$$F_{\xi}(x_1, \dots, x_n) = \prod_{i=1}^n F_{\xi_i}(x_i),$$

где $\xi_i = \pi_i(\xi)$ — одномерная случайная величина.

Часто требуется охарактеризовать векторную случайную величину некоторым числом. В общем виде это делается по следующей схеме. С каждой точкой пространства R^n сопоставляется вещественное число. Иными словами, задается функция $f(x)$ ($x \in R^n$). Предположим, что f — это измеримая функция, т. е. для любого $A \in \mathcal{B}^1$

$$f^{-1}(A) \in \mathcal{B}^n,$$

где \mathcal{B}^1 — это борелевская σ -алгебра подмножеств из области значений функции f , а \mathcal{B}^n — борелевская σ -алгебра подмножеств из области определения функции f . Тогда функция от ω $f(\xi)$ — это одномерная случайная величина (с реализациями $f(\xi(\omega))$).

Действительно, для любого $A \in \mathcal{B}^1$

$$\{f(\xi) \in A\} = \xi^{-1}\{f \in A\} \in \mathcal{F},$$

так как

$$\{f \in A\} = f^{-1}(A) \in \mathcal{B}^n$$

и ξ — случайная векторная величина.

Следовательно, $f(\xi)$ — измеримая функция, т. е. случайная величина. Математическое ожидание этой случайной величины

$$E f(\xi) = \int_{\Omega} f(\xi(\omega)) P(d\omega).$$

Так как f — это случайная величина на вероятностном пространстве $(R^n, \mathcal{B}^n, \mu_{\xi})$, то

$$E f(\xi) = \int_{R^n} f(x) \mu_{\xi}(dx).$$

Что это действительно так, следует из вычислений:

$$\begin{aligned} \int_{\Omega} f(\xi(\omega)) P(d\omega) &= \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} P\left(\omega : \frac{k}{n} \leq f(\xi(\omega)) < \frac{k+1}{n}\right) = \\ &= \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} P\left(\xi^{-1}\left(x : \frac{k}{n} \leq f(x) < \frac{k+1}{n}\right)\right) = \\ &= \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} \mu_{\xi}\left(x : \frac{k}{n} \leq f(x) < \frac{k+1}{n}\right) = \int_{R^n} f(x) \mu_{\xi}(dx). \end{aligned}$$

Кроме того случайная величина $f(\xi)$ индуцирует вероятностную меру $\mu_{f(\xi)}$ на измеримом пространстве (R^1, \mathcal{B}^1) , превращая его в вероятностное пространство $(R^1, \mathcal{B}^1, \mu_{f(\xi)})$. При этом

$$E f(\xi) = \int_{R^1} x \mu_{f(\xi)}(dx) = \int_{R^1} x dF_{f(\xi)}(x).$$

Эта формула обосновалась уже в разделе II.3.2. Вычисление интеграла $\int_{R^n} f(x) \mu_\xi(dx)$ для дискретного и абсолютно непрерывного типа распределения аналогично одномерному случаю.

а) Дискретное распределение μ_ξ :

$$\int_{R^n} f(x) \mu_\xi(dx) = \sum_{i=1}^{\infty} f(a_i) \mu_\xi\{a_i\},$$

где $\sum_{i=1}^{\infty} \mu_\xi\{a_i\} = 1$, $a_i \in R^n$.

б) Абсолютно непрерывное распределение μ_ξ с плотностью $p_\xi(x)$:

$$\int_{R^n} f(x) \mu_\xi(dx) = \int_{R^n} f(x) p_\xi(x) dx;$$

последний интеграл для случая, когда функции f и p_ξ непрерывны, всюду, за исключением, может быть, некоторых точек и линий, сводится к n -кратному интегралу

$$\int_{R^n} f(x_1, \dots, x_n) p_\xi(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Рассмотрим в качестве вещественной функции от векторного аргумента проекцию $\pi_i(x)$ ($x \in R^n$, $\pi_i(x) \in R^1$), где при координатном представлении $x = (x_1, \dots, x_n)$, $\pi_i(x) = x_i$ ($i = 1, 2, \dots, n$).

Распределение случайной величины $\pi_i(\xi)$, т. е. вероятностная мера $\mu_{\pi_i(\xi)}$, называется *маргинальным распределением по отношению к μ_ξ* . При любом борелевском $A \subset R^1$

$$\begin{aligned} \mu_{\pi_i(\xi)}(A) &= \mu_\xi(\pi_i \in A) = P(\pi_i(\xi) \in A) = P\left(\bigcap_{j \neq i} \{\pi_j(\xi) \in R^1\} \cap \{\pi_i(\xi) \in A\}\right) = \\ &= \mu_\xi\left(\bigcap_{j \neq i} \{\pi_j \in R^1\} \cap \{\pi_i \in A\}\right) = \mu_\xi\left(\underbrace{R^1 \times \dots \times R^1}_{i-1} \times A \times \underbrace{R^1 \times \dots \times R^1}_{n-i}\right), \end{aligned}$$

т. е. это мера «слоя», где множеством A ограничена только одна i -тая координата, а остальные координаты могут принимать любые вещественные значения.

Маргинальное распределение $\mu_{\pi_i(\xi)}$ называют также *проекцией* распределения μ_ξ на i -тую координатную ось. Можно рассмотреть проекции более общего вида, например, на некоторое k -мерное подпространство, задаваемые осями координат (i_1, \dots, i_k) . Более общей концепцией является преобразование исходного распреде-

ления μ_ξ , порождаемого линейным отображением пространства R^n в пространство R^k . Ниже это преобразование будет рассмотрено специально.

Найдем математическое ожидание случайной величины $\pi_i(\xi)$:

$$E\pi_i(\xi) = \int_{R^n} \pi_i(x) \mu_\xi(dx) = \int_{R^1} x \mu_{\pi_i(\xi)}(dx).$$

Вектор

$$E\xi = (E\pi_1(\xi), E\pi_2(\xi), \dots, E\pi_n(\xi))$$

называется математическим ожиданием векторной случайной величины ξ .

Таким образом,

$$E\pi_i(\xi) = \pi_i(E\xi).$$

Это частный случай свойства линейности математического ожидания. Пользуясь обозначением интеграла Лебега от векторнозначной случайной величины по мере P , который определяется по координатно так, как это делается в П. 3. 2, можно записать:

$$E\xi = \int_{\Omega} \xi(\omega) P(d\omega) = \int_{R^n} \mathbf{x} \mu_\xi(dx).$$

П.5.3. Характеристики векторных случайных величин

Рассмотрим векторную случайную величину $\xi(\omega)$ на вероятностном пространстве (Ω, \mathcal{F}, P) . Такие величины, естественно, возникают, когда исследуются совместные распределения n случайных величин (ξ_1, \dots, ξ_n) , которые можно считать координатами вектора $\xi = (\xi_1, \dots, \xi_n)$ или — в других обозначениях — $\pi_i(\xi) = \xi_i$. Выше мы определили математическое ожидание векторной случайной величины

$$E\xi = (a_1, \dots, a_n),$$

где $a_i = E\xi_i = E\pi_i(\xi)$.

Естественно ожидать, что характеристики распределения векторной случайной величины ξ не сводятся к характеристикам распределений его координат ξ_i . В общем случае это действительно так, за исключением немногих характеристик, таких как $E\xi$, и случая взаимной независимости координат случайного вектора. Мы перечислим некоторые из этих характеристик.

а. Моменты векторной случайной величины. Моментами векторной случайной величины ξ (или ее распределения) называются интегралы (если они существуют)

$$E(\xi_1^{k_1} \dots \xi_n^{k_n}) = \int_{R^n} f(\mathbf{x}) \mu_\xi(dx),$$

где $f(\mathbf{x}) = (\pi_1(\mathbf{x}))^{k_1} \dots (\pi_n(\mathbf{x}))^{k_n}$.

При координатном выражении $\mathbf{x} = (x_1, \dots, x_n)$ эта функция представляется в виде $f(x_1, \dots, x_n) = x_1^{k_1} \dots x_n^{k_n}$. В частности,

$$E(\xi_1^{k_1} \dots \xi_n^{k_n}) = \sum_{i=1}^{\infty} a_{i1}^{k_1} \dots a_{in}^{k_n} \mu_{\xi}(\mathbf{a}_i)$$

для дискретного распределения, где

$$\mathbf{a}_i = (a_{i1}, \dots, a_{in}) \text{ и } \sum_{i=1}^{\infty} \mu_{\xi}(\mathbf{a}_i) = 1,$$

$$E(\xi_1^{k_1} \dots \xi_n^{k_n}) = \int_{R^n} x_1^{k_1} \dots x_n^{k_n} p_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n$$

для абсолютно непрерывного распределения с плотностью p_{ξ} . Число $k = k_1 + \dots + k_n$ ($k_i \geq 0$) называется порядком момента. Моменты первого порядка — это математические ожидания координат случайного вектора (или его маргинальных распределений)

$$E\xi_i = \int_{R^n} \pi_i(\mathbf{x}) \mu_{\xi}(d\mathbf{x}).$$

б. Центральные моменты. Центральными моментами называются интегралы

$$E(\xi_1 - E\xi_1)^{k_1} \dots (\xi_n - E\xi_n)^{k_n} = \int_{R^n} \prod_{i=1}^n (\pi_i(\mathbf{x}) - E\xi_i)^{k_i} \mu_{\xi}(d\mathbf{x}),$$

где $k_i \geq 0$ и $k = k_1 + \dots + k_n$ — порядок центрального момента. Наибольшее употребление находят центральные моменты второго порядка. Обозначая

$$\lambda_{ij} = E(\xi_i - E\xi_i)(\xi_j - E\xi_j),$$

можно записать все центральные моменты второго порядка в виде матрицы

$$\Lambda = \begin{pmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \dots & \dots & \dots \\ \lambda_{n1} & \dots & \lambda_{nn} \end{pmatrix}.$$

Величины $\lambda_{ii} = E(\xi_i - E\xi_i)^2$ — не что иное, как дисперсии координат случайного вектора ξ ($\lambda_{ii} = D\xi_i$), а $\lambda_{ij} = \lambda_{ji}$ — это ковариация i -той и j -той координат (см. II. 2. 2), $\lambda_{ij} = \text{cov}(\xi_i, \xi_j)$. Матрицу Λ называют *ковариационной матрицей* случайного вектора ξ (или его распределения).

Число $\rho_{ij} = \frac{\text{cov}(\xi_i, \xi_j)}{\sqrt{D\xi_i D\xi_j}}$ называется *коэффициентом корреляции* случайных величин ξ_i и ξ_j . По абсолютной величине ρ_{ij} не больше единицы (следствие неравенства $(E(\xi_i - E\xi_i)(\xi_j - E\xi_j))^2 \leq E(\xi_i - E\xi_i)^2 \cdot E(\xi_j - E\xi_j)^2$ — интегральной формы неравенства Коши — Буняковского (Колмогоров, Фомин, 1972, с. 47)). Когда

$\rho_{ij} = 0$, говорят, что случайные величины ξ_i и ξ_j некоррелированы. Если ξ_i и ξ_j — независимые случайные величины, то они некоррелированы, так как в этом случае $\text{cov}(\xi_i, \xi_j) = E\xi_i\xi_j - E\xi_i E\xi_j = 0$. Обратное, вообще говоря, неверно.

Пусть a_1, \dots, a_n — произвольные вещественные числа. Тогда

$$E \left(\sum_{i=1}^n (\xi_i - E\xi_i) a_i \right)^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} a_i a_j \geq 0.$$

Матрица Λ является матрицей коэффициентов этой квадратичной формы. Неотрицательность этой квадратичной формы означает по определению неотрицательную определенность матрицы Λ .

в. Матрица информационных коэффициентов корреляции. Пусть $\pi_{ij}(x) = (\pi_i(x), \pi_j(x))$ — проекция вектора $x \in R^n$ на плоскость, соответствующую i -той и j -той координатам. Таким образом, если $x = (x_1, \dots, x_n)$, то $\pi_{ij}(x) = (x_i, x_j)$ — двумерный вектор. Очевидно, π_{ij} — измеримое отображение пространства R^n в пространство R^2 . Если теперь ξ — n -мерная векторная случайная величина, то $\pi_{ij}(\xi)$ — двумерная векторная случайная величина. Распределение $\mu_{\pi_{ij}(\xi)}$, где $\mu_{\pi_{ij}(\xi)}(A) = \mu_{\xi}(\pi_{ij} \in A) = P(\pi_{ij}(\xi) \in A)$ ($A \in \mathcal{B}^2$), является распределением случайного вектора $\pi_{ij}(\xi)$. Оно также называется маргинальным по отношению к распределению μ_{ξ} . Пусть теперь маргинальные распределения $\mu_{\pi_i(\xi)}$, $\mu_{\pi_j(\xi)}$ и $\mu_{\pi_{ij}(\xi)}$ абсолютно непрерывны, т. е. существуют плотности этих распределений $p_{\pi_i(\xi)}$, $p_{\pi_j(\xi)}$ и $p_{\pi_{ij}(\xi)}$. Тогда определена величина, называемая количеством информации одной компоненты $\pi_i(\xi)$ случайного вектора относительно другой компоненты $\pi_j(\xi)$:

$$J_{ij}(\xi) = \int_{R^2} p_{\pi_{ij}(\xi)}(x, y) \log \frac{p_{\pi_{ij}(\xi)}(x, y)}{p_{\pi_i(\xi)}(x) p_{\pi_j(\xi)}(y)} dx dy.$$

Очевидно, эта величина симметрична относительно перестановки i и j . Она отражает меру стохастической зависимости случайных величин $\pi_i(\xi)$ и $\pi_j(\xi)$. Если эти случайные величины независимы, то $J_{ij}(\xi) = 0$.

Действительно, пусть $A = A_1 \times A_2$ ($A_1, A_2 \in \mathcal{B}^1$) — «прямоугольник», лежащий в плоскости R^2 , со «сторонами» A_1 и A_2 . Тогда

$$\begin{aligned} \int_A p_{\pi_{ij}(\xi)}(x, y) dx dy &= \mu_{\pi_{ij}(\xi)}(A) = \\ &= P(\pi_{ij}(\xi) \in A) = P((\pi_i(\xi) \in A_1) \cap (\pi_j(\xi) \in A_2)) = \\ &= P(\pi_i(\xi) \in A_1) P(\pi_j(\xi) \in A_2) = \mu_{\pi_i(\xi)}(A_1) \mu_{\pi_j(\xi)}(A_2) = \\ &= \int_{A_1 \times A_2} p_{\pi_i(\xi)}(x) p_{\pi_j(\xi)}(y) dx dy. \end{aligned}$$

Так как это равенство справедливо для всех A_1 и A_2 и, согласно предположению, $P_{\pi_i(\xi)}$ и $P_{\pi_j(\xi)}$ кусочно-непрерывны, то

$$P_{\pi_{ij}(\xi)}(x, y) = P_{\pi_i(\xi)}(x) P_{\pi_j(\xi)}(y)$$

всюду на плоскости R^2 и поэтому $J_{ij}(\xi) = 0$.

Аналогично доказывается, что совместная плотность n независимых случайных величин равна произведению плотностей этих случайных величин, если только совместная плотность существует.

Кроме матрицы $(J_{ij}(\xi))_{n \times n}$ в работах по математической геологии для характеристики многомерных распределений используется также матрица, составленная из элементов

$$\rho_{ij}(\xi) = \sqrt{1 - e^{-2|J_{ij}(\xi)|}},$$

называемых *информационными коэффициентами корреляции* (Кульбак, 1967; Vistelius, 1967), при этом $0 \leq \rho_{ij} < 1$. Значение $\rho_{ij} = 1$ приписывают такому совместному распределению компонент $\pi_i(\xi)$ и $\pi_j(\xi)$, когда существуют плотности распределений $\mu_{\pi_i(\xi)}$ и $\mu_{\pi_j(\xi)}$, но не существует плотность самого распределения $\mu_{\pi_{ij}(\xi)}$. Так будет, например, если с вероятностью единица $F(\pi_i(\xi), \pi_j(\xi)) = 0$ для какой-нибудь функции F , определенной на плоскости R^2 .

В то же время любое взаимно однозначное и непрерывное преобразование R^n в R^n сохраняет матрицу информационных коэффициентов корреляции неизменной.

г. Характеристическая функция многомерного распределения. Пусть $\mathbf{t} = (t_1, \dots, t_n)$ и $\mathbf{x} = (x_1, \dots, x_n)$ — два n -мерных вектора, $(\mathbf{t}, \mathbf{x}) = t_1 x_1 + \dots + t_n x_n$ — скалярное произведение этих векторов. При этом определена случайная величина (\mathbf{t}, ξ) , где

$$(\mathbf{t}, \xi(\omega)) = t_1 \pi_1(\xi(\omega)) + \dots + t_n \pi_n(\xi(\omega)).$$

Характеристической функцией случайного вектора ξ (или его распределения) называется комплексная функция, определенная для всех $\mathbf{t} \in R^n$,

$$E e^{i(\mathbf{t}, \xi)} = \int_{R^n} e^{i(\mathbf{t}, \mathbf{x})} \mu_{\xi}(d\mathbf{x}) = \int_{R^n} \exp i \sum_{k=1}^n t_k \pi_k(\mathbf{x}) \mu_{\xi}(d\mathbf{x}).$$

Соответствие между распределениями векторных случайных величин и их характеристическими функциями взаимно однозначно и непрерывно. Если моменты второго порядка существуют, то в окрестности точки $\mathbf{t} = (0, \dots, 0) \in R^n$ имеет место разложение

$$E e^{i(\mathbf{t}, \xi)} = e^{i(\mathbf{t}, E\xi)} \left(1 - \frac{1}{2} \sum_{k=1}^n \sum_{j=1}^n \lambda_{kj} t_k t_j + o \left(\sum_{j=1}^n t_j^2 \right) \right),$$

где $\frac{o(a)}{a} \rightarrow 0$, если $a \rightarrow 0$ ($o(a)$ называется величиной более высокого порядка малости по сравнению с a).

Если координаты случайного вектора независимы, то, очевидно,

$$E e^{i(t, \xi)} = \prod_{k=1}^n E e^{i t_k \pi_k(\xi)}.$$

И обратно, если это представление справедливо для всех $t = (t_1, \dots, t_n) \in R^n$, то координаты $(\pi_1(\xi), \dots, \pi_n(\xi))$ взаимно независимы.

Для характеристики многомерных распределений используются также многомерные производящие функции и многомерные преобразования Лапласа, свойства которых обобщают свойства соответствующих одномерных преобразований.

$E \prod_{k=1}^n s_k^{\pi_k(\xi)}$ — многомерная производящая функция,

где $(s_1, \dots, s_n) \in R^n$, причем $0 \leq s_k \leq 1$ и все $\pi_k(\xi)$ — неотрицательные целочисленные случайные величины.

$E \prod_{k=1}^n e^{-\lambda_k \pi_k(\xi)}$ — многомерное преобразование Лапласа,

где $(\lambda_1, \dots, \lambda_n) \in R^n$, причем $\lambda_k \geq 0$ и все $\pi_k(\xi)$ — неотрицательные случайные величины.

П.5.4. Примеры распределений векторных случайных величин

а. Полиномиальное распределение. Пусть $\xi = (\xi_1, \dots, \xi_N) (N = 1, 2, \dots)$ — конечная последовательность независимых и одинаково распределенных случайных величин, заданных на некотором вероятностном пространстве. Каждая случайная величина может принимать n различных значений a_1, a_2, \dots, a_n с вероятностями p_1, p_2, \dots, p_n , где $p_i > 0$ и $\sum_{i=1}^n p_i = 1$.

Рассмотрим случайный вектор $k(\xi) = (k_1(\xi), \dots, k_n(\xi))$, где для данной реализации $\xi(\omega)$ $k_i(\xi(\omega))$ — число тех координат вектора $\xi(\omega)$, которые равны a_i . Распределение случайного вектора $k(\xi)$ называется полиномиальным. Это дискретное распределение, сосредоточенное на точках $x(x_1, \dots, x_n) \in R^n$ с неотрицательными целочисленными координатами, причем $\sum_{i=1}^n x_i = N$. Нетрудно подсчитать, что

$$P(k(\xi) = x) = \frac{N!}{x_1! \dots x_n!} p_1^{x_1} \dots p_n^{x_n}.$$

Натуральные числа N и n , а также вероятности p_1, \dots, p_{n-1} являются параметрами полиномиального распределения ($p_n = 1 - p_1 - \dots - p_{n-1}$ не является независимым параметром). Вероятность $P(k(\xi) = x)$ является коэффициентом при $\alpha_1^{x_1} \dots \alpha_n^{x_n}$ в разложении полинома $(p_1 \alpha_1 + \dots + p_n \alpha_n)^N$. Биномиальное распределение является частным случаем полиномиального при $n = 2$. Оно же является маргинальным для полиномиального.

Действительно,

$$P(k_1(\xi) = x_1) = \sum P(\mathbf{k}(\xi) = \mathbf{x}) = \sum \frac{N!}{x_1! x_2! \dots x_n! p_1^{x_1} p_2^{x_2} \dots p_n^{x_n}} =$$

$$= \frac{N!}{(N-x_1)! p_1^{x_1}} \sum \frac{(N-x_1)!}{x_2! \dots x_n! p_2^{x_2} \dots p_n^{x_n}}$$

(сумма берется по всем $x_2, \dots, x_n \geq 0$, таким что $\sum_{i=2}^n x_i = N - x_1$).

Полагая в последней сумме $N' = N - x_1$ и $p_i = p'_i \alpha_i$, где $\alpha_i = 1 - p_1$, мы будем иметь $\sum_{i=2}^n p'_i = 1$, и поэтому последняя сумма равна

$$\sum \frac{N'!}{x_2! \dots x_n!} (p'_2)^{x_2} \dots (p'_n)^{x_n} \alpha_2^{x_2} \dots \alpha_n^{x_n} = (p'_2 \alpha_2 + \dots + p'_n \alpha_n)^{N'} =$$

$$= (1 - p_1)^{N-x_1},$$

т. е. случайная величина $k_1(\xi)$ имеет биномиальное распределение с параметрами p_1 и N .

То же относится к любой другой компоненте $k_i(\xi)$ вектора $\mathbf{k}(\xi)$ (ее параметры p_i и N).

Тем же методом легко подсчитать, что маргинальное двумерное распределение, например вектора $(k_i(\xi), k_j(\xi))$, — это полиномиальное распределение третьей степени с параметрами N, p_i, p_j .

Подсчитать ковариационную матрицу случайного вектора $\mathbf{k}(\xi)$ можно непосредственно, воспользовавшись аналитическим выражением для совместных распределений пар компонент. Но лучше это сделать с помощью следующего приема. Заметим, что

$$k_i(\xi) = \sum_{k=1}^N I_{\{a_i\}}(\xi_k),$$

где

$$I_{\{a_i\}}(\xi_k) = \begin{cases} 0, & \text{если } \xi_k \neq a_i, \\ 1, & \text{если } \xi_k = a_i, \end{cases}$$

откуда

$$EI_{\{a_i\}}(\xi_k) = p_{ik} = p_i,$$

и если $i \neq j$, то

$$I_{\{a_i\}}(\xi_k) I_{\{a_j\}}(\xi_k) = 0.$$

Тогда при $i = j$

$$\lambda_{ii}(k(\xi)) = Dk_i(\xi) = Np_i(1 - p_i) \quad (\text{см. II. 4. 1 (a)})$$

и, если $i \neq j$, то

$$\lambda_{ij}(k(\xi)) = \text{cov}(k_i(\xi), k_j(\xi)) = Ek_i(\xi) k_j(\xi) - Ek_i(\xi) Ek_j(\xi) =$$

$$= \sum_{k=1}^N \sum_{l=1}^N EI_{\{a_i\}}(\xi_k) I_{\{a_j\}}(\xi_l) - \sum_{k=1}^N \sum_{l=1}^N EI_{\{a_i\}}(\xi_k) \cdot EI_{\{a_j\}}(\xi_l) =$$

$$= \sum_{\substack{k \neq l \\ 1 \leq k, l \leq N}} p_{ik} p_{jl} - \sum_{k=1}^N \sum_{l=1}^N p_{ik} p_{jl} = -Np_i p_j.$$

Полиномиальное распределение имеет многочисленные применения в геологической практике. Вектор, представляющий состав в пробе с фиксированным числом зерен, отобранной из некоторого статистического ансамбля так, что обеспечивается независимость и однородность условий отбора каждого зерна, имеет полиномиальное распределение. Полиномиальное распределение играет важную роль в статистике марковских цепей, которые часто применяются для моделирования геологических явлений.

Производящая функция полиномиального распределения подсчитывается просто. Для вектора $\mathbf{k}(\xi) = (k_1(\xi), \dots, k_n(\xi))$ при $0 \leq s_i \leq 1$

$$\begin{aligned} E \prod_{i=1}^n s_i^{k_i}(\xi) &= \sum_{\substack{x_1, \dots, x_n \geq 0 \\ \sum_{i=1}^n x_i = N}} \frac{N!}{x_1! x_2! \dots x_n!} (p_1 s_1)^{x_1} \dots (p_n s_n)^{x_n} = \\ &= (p_1 s_1 + \dots + p_n s_n)^N. \end{aligned}$$

Так как компонента $k_n(\xi)$ может быть представлена в виде

$$k_n(\xi) = N - \sum_{i=1}^{n-1} k_i(\xi),$$

то распределение характеризуется $n-1$ -мерной производящей функцией

$$\begin{aligned} E \prod_{i=1}^{n-1} s_i^{k_i}(\xi) &= (p_1 s_1 + \dots + p_{n-1} s_{n-1} + p_n)^N = \\ &= \left(1 - \sum_{i=1}^{n-1} p_i (1 - s_i)\right)^N \quad (\text{см. II. 4.1 (а) для } n=2). \end{aligned}$$

б. Нормальное распределение. Это абсолютно непрерывное распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ с плотностью распределения

$$p_{\xi}(x_1, \dots, x_n) = C e^{-\frac{1}{2} Q(x_1, \dots, x_n)}, \quad (\text{II. 5.1})$$

где $Q(x_1, \dots, x_n) = \sum_{i,j=1}^n b_{ij}(x_i - a_i)(x_j - a_j)$ — положительно определенная квадратичная форма, C , a_i и b_{ij} — вещественные постоянные. При этом $C = (\sqrt{2\pi})^{-n} \sqrt{D}$, где D — определитель матрицы $(b_{ij})_{n \times n}$.

Таким образом, параметрами нормального распределения являются n , a_i , b_{ij} ($1 \leq i, j \leq n$).

Нормальное распределение часто используется для описания статистических ансамблей. Обоснованием хорошего соответствия этого описания наблюдаемым опытным данным принято считать уже упоминавшуюся нами центральную предельную теорему,

доказательство основных вариантов которой в многомерном случае принципиально не отличается от одномерного. В теоретическом плане нормальное распределение относится к числу наиболее исследованных и часто естественно возникающих при тех или иных расчетах благодаря многим экстремальным свойствам этого распределения. Доказано, например, что среди абсолютно непрерывных распределений одномерных случайных величин с плотностью, положительной при всех $x \in R^1$, и конечной дисперсией, значение которой фиксировано, нормальные распределения имеют наибольшую энтропию (см. II.3.3 (е)). Аналогичным свойством максимума энтропии обладают многомерные нормальные распределения (Рао, 1968, с. 477). Другое замечательное свойство нормального распределения доказано В. П. Скитовичем (1954), оно состоит в следующем.

Пусть ξ_1, \dots, ξ_n — взаимно независимые случайные величины и существуют такие наборы ненулевых коэффициентов (a_1, \dots, a_n) и (b_1, \dots, b_n) , что линейные формы $\sum_{i=1}^n a_i \xi_i$ и $\sum_{i=1}^n b_i \xi_i$ являются независимыми случайными величинами. Тогда величины ξ_i распределены нормально. Свойства нормального распределения и его параметров наиболее просто вытекают из вида его характеристической функции.

Нетрудно показать (Гнеденко, 1961, с. 251), что

$$E e^{i(t, \xi)} = \exp \left(i \sum_{j=1}^n a_j t_j - \frac{1}{2} \sum_{k=1}^n \sum_{j=1}^n c_{kj} t_j t_k \right) \quad (t = (t_1, \dots, t_n) \in R^n), \quad (\text{II. 5. 2})$$

где матрица $(C_{kj})_{n \times n}$ является обратной для матрицы (b_{ij}) . Отсюда следует, что

$$a_i = E \xi_i, \quad c_{kj} = \lambda_{kj} = \text{cov}(\xi_k, \xi_j) \quad (k \neq j) \quad \text{и} \quad c_{jj} = \lambda_{jj} = D \xi_j \quad (\text{см. II. 5. 3 (r)}).$$

Поэтому матрицы (c_{kj}) и (b_{ij}) симметричны.

Для любого $k (1 \leq k < n)$ k -мерное маргинальное распределение также является нормальным. Для доказательства этого найдем характеристическую функцию вектора из k компонент, например $\xi' = (\xi_1, \dots, \xi_k)$. Для этого достаточно положить в выражении для $E e^{i(t, \xi)}$ $t_{k+1} = \dots = t_n = 0$. Получим

$$E e^{i(t', \xi')} = \exp \left(i \sum_{j=1}^k a_j t_j - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k c_{ij} t_j t_i \right) \quad (t' = (t_1, \dots, t_k) \in R^k),$$

что является характеристической функцией k -мерного нормального распределения. Если компоненты случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ независимы, то $\text{cov}(\xi_i, \xi_j) = 0 \quad (i \neq j)$ (т. е. ξ_i и ξ_j некоррелированы). Для нормально распределенного вектора справедливо также обратное утверждение. Действительно, если матрица (c_{ij}) является диагональной, то и обратная ей матрица

(b_{ij}) также диагональная, и плотность (II.5.1) представляется в виде произведения плотностей одномерных нормальных распределений, т. е. компоненты ξ_i — независимые нормально распределенные величины.

До сих пор мы говорили о невырожденном нормальном распределении, при котором определители матриц (b_{ij}) и (c_{ij}) не равны нулю. Нормальным распределением называют также распределение с характеристической функцией вида (II.5.2), у которого ковариационная матрица имеет определитель, равный нулю. Это так называемое вырожденное, или несобственное, нормальное распределение.

Пусть ранг ковариационной матрицы равен r ($1 \leq r < n$). Известно (Крамер, 1975, с. 327), что в этом случае существует линейное множество (линия, плоскость, гиперплоскость) L_r размерности r такое, что все распределение сосредоточено на L_r (т. е. $\mu_{\xi}(L_r) = 1$) и не сосредоточено ни на каком другом линейном множестве меньшей размерности. В этом случае существует ровно $n-r$ линейных соотношений между компонентами ξ_i , которые выполняются с вероятностью единица.

Вырожденное нормальное распределение не является абсолютно непрерывным. Однако каждая из n компонент ξ_i может быть представлена в виде линейной функции от r независимых и нормально распределенных величин, совместное распределение которых абсолютно непрерывно в пространстве R^r .

Рассмотрим линейное преобразование пространства R^n в пространство R^m

$$y_i = \sum_{k=1}^n c_{ik} x_k \quad (i=1, 2, \dots, m)$$

с матрицей $C = (c_{ik})_{m \times n}$, где m не обязательно равно n . В матричном обозначении это преобразование имеет вид $y = xC^T$ (для удобства в качестве основного обозначения для вектора мы выбрали вектор-строку); таким образом, $y = (y_1, \dots, y_m)$, $x = (x_1, \dots, x_n)$; индекс «Т» сверху обозначает транспонирование; таким образом, y^T и x^T — это соответствующие векторы-столбцы, и предыдущее равенство можно записать в виде

$$y^T = Cx^T.$$

Это преобразование определяет новую векторную случайную величину с m -мерным распределением.

Пусть

$$\varphi_{\xi}(t) = E e^{i(t, \xi)} \quad (t = (t_1, \dots, t_m) \in R^m)$$

является характеристической функцией случайного вектора $\xi = (\xi_1, \dots, \xi_n)$. Определим характеристическую функцию преобразованного случайного вектора $\eta = (\eta_1, \dots, \eta_m) = \xi C^T$. Так как (t, ξ) можно записать в матричном виде как $t \xi^T$, то, определив вектор t как $t = uC$, где $u = (u_1, \dots, u_m) \in R^m$, получим

$$(t, \xi) = (uC, \xi) = uC \cdot \xi^T = u \cdot (C \cdot \xi^T) = u \cdot (\xi C^T)^T = (u, \eta)$$

(здесь мы использовали сочетательный закон умножения матриц и правило транспонирования).

Следовательно, характеристическая функция φ_η случайного вектора η вычисляется по правилу

$$\varphi_\eta(u) = \varphi_\xi(uC).$$

Пусть теперь ξ имеет нормальное распределение с характеристической функцией

$$\varphi_\xi(t) = \exp \left(i \sum_{j=1}^n a_j t_j - \frac{1}{2} \sum_{j,k=1}^n \lambda_{jk} t_j t_k \right),$$

которую можно записать в матричном виде как

$$\exp \left(it \cdot a^T - \frac{1}{2} t \cdot \Lambda \cdot t^T \right),$$

где $a = (a_1, \dots, a_n)$ — математическое ожидание, а Λ — ковариационная матрица случайного вектора ξ .

Тогда

$$\varphi_\eta(u) = \exp \left(iuCa^T - \frac{1}{2} u\Lambda C^T u^T \right).$$

Очевидно, $u\Lambda C^T u^T$ — неотрицательно определенная квадратичная форма, и, следовательно, случайный вектор η имеет нормальное (может быть, вырожденное) распределение. В частности, линейная функция от n компонент случайного вектора, распределенного нормально, также распределена нормально. Оказывается верен гораздо более глубокий обратный результат: если сумма двух независимых случайных величин распределена нормально, то каждая из этих случайных величин также распределена нормально (Феллер, 1967, с. 600).

Рассмотрим плотность невырожденного нормального распределения (II.5.4). Она представляет собой непрерывную положительную функцию с максимумом в точке $E\xi = a = (a_1, \dots, a_n)$, убывающую до нуля на бесконечности. При любом $c \neq 0$ уравнение

$$\sum_{i,j=1}^n b_{ij} (x_i - a_i)(x_j - a_j) = c^2$$

определяет поверхность уровня одинаковой плотности вероятности $p_\xi(x)$. Эта поверхность представляет собой эллипсоид с центром в точке a . С помощью поворота координатной системы можно сделать главные оси эллипсоида параллельными координатным осям. Этому повороту соответствует ортогональное преобразование пространства R^n , а новому положению эллипсоидов одинакового уровня — независимость компонент преобразованного случайного вектора. Эллипсоид, определенный для $c^2 = n + 2$, назы-

вается эллипсоидом рассеяния. Он может определяться для любого n -мерного распределения с конечными вторыми моментами и служит для сравнения распределений по степени сосредоточения вероятности около центра (Крамер, 1975, с. 331). Квадрат его объема пропорционален $D^{-1} = |\Lambda|$ (определитель ковариационной матрицы). Эта величина называется *обобщенной дисперсией*.

Как было отмечено, маргинальные распределения нормально распределенного вектора также нормальны. Однако нормальность одномерных маргинальных распределений не является достаточной для того, чтобы распределение вектора было нормальным. Приведем пример.

Пусть ξ и η — двумерные случайные векторы, распределенные нормально с математическими ожиданиями $E\xi = E\eta = (0, 0)$ и ковариационными матрицами $\Lambda_\xi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ и $\Lambda_\eta = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$ соответственно.

Рассмотрим плотность

$$p(x, y) = \frac{1}{2} p_\xi(x, y) + \frac{1}{2} p_\eta(x, y).$$

Тогда

$$p_1(x) = \int_{-\infty}^{\infty} p(x, y) dy = \frac{1}{2} \int_{-\infty}^{\infty} (p_\xi(x, y) + p_\eta(x, y)) dy$$

и

$$p_2(x) = \int_{-\infty}^{\infty} p(y, x) dy = \frac{1}{2} \int_{-\infty}^{\infty} (p_\xi(y, x) + p_\eta(y, x)) dy$$

являются плотностями маргинальных распределений. Но

$$p_\xi(x, y) = p_\xi(y, x), \quad p_\eta(x, y) = p_\eta(y, x)$$

и

$$\int_{-\infty}^{\infty} p_\xi(x, y) dy = \int_{-\infty}^{\infty} p_\eta(x, y) dy,$$

так как это плотности одного и того же нормального распределения с параметрами 0 и 1. Отсюда $p_1(x)$ и $p_2(x)$ — плотности нормального распределения с параметрами 0 и 1. В то же время $p(x, y)$ не является плотностью нормального распределения, так как соответствующая характеристическая функция равна $\frac{1}{2} e^{-\frac{1}{2}(t_1^2 + t_2^2)}$ +

$+ \frac{1}{2} e^{-\frac{1}{2}(t_1^2 + t_2^2 + t_1 t_2)}$ и ее логарифм не является полиномом второй степени от t_1, t_2 .

в. Распределение Коши. Это распределение представляет собой абсолютно непрерывное распределение случайной векторной величины ξ с плотностью распределения

$$p_{\xi}(x_1, \dots, x_n) = \frac{c_n}{\left(1 + \sum_{i=1}^n x_i^2\right)^{\frac{n+1}{2}}},$$

где

$$c_n = \frac{\pi^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)}.$$

В этой форме распределение Коши имеет только один параметр — размерность n . Иногда распределением Коши называют также распределение случайного вектора $\eta = a + \xi \cdot B$ при любом векторе $a \in R^n$ и матрице B порядка $n \times n$, которые в этом случае являются параметрами распределения.

Распределение Коши представляет собой пример распределения, которое не имеет математического ожидания и тем более моментов высших порядков. Плотность p_{ξ} симметрична относительно начала координат, где она имеет максимум. Поверхности одинакового уровня плотности представляют собой сферы. Все маргинальные распределения Коши являются также распределениями Коши меньшей размерности. Показать это проще всего с помощью характеристической функции. Докажем, что

$$E e^{i(t, x)} = e^{-|t|},$$

где $|t| = \sqrt{t_1^2 + \dots + t_n^2}$, если $t = (t_1, \dots, t_n)$.

Для этого в интеграле

$$\int_{R^n} \frac{e^{i \sum_{j=1}^n t_j x_j} dx_1 \dots dx_n}{\left(1 + \sum_{i=1}^n x_i^2\right)^{\frac{n+1}{2}}}$$

сделаем замену переменных, перейдя к сферической системе координат. Замечая, что $(t \cdot x) = \sum_{j=1}^n t_j x_j = |t| \cdot |x| \cos(\widehat{t, x})$, направим одну ось отсчета углов по вектору t .

Якобиан преобразования имеет вид (Фихтенгольд, 1960, с. 402)

$$r^{n-1} \sin^{n-2} \varphi_1 \cdot F(\varphi_2, \dots, \varphi_{n-1}),$$

где $\varphi_1 = (\widehat{t, x})$ и $F(\varphi_2, \dots, \varphi_{n-1})$ — множитель, зависящий только от всех остальных углов сферической системы координат.

Обозначив $|x| = r$, где $|x| = \sqrt{x_1^2 + \dots + x_n^2}$, мы получим

$$C \int_0^{\infty} dr \int_0^{\pi} d\varphi \frac{e^{i|t|r \cos \varphi}}{(1+r^2)^{\frac{n+1}{2}}} r^{n-2} \sin^{n-2} \varphi,$$

где $C = \int_{(\varphi_2, \dots, \varphi_{n-1})} F(\varphi_2, \dots, \varphi_{n-1}) d\varphi_2, \dots, d\varphi_{n-1}$ (интеграл берется

по области всех возможных углов $\varphi_2, \dots, \varphi_{n-1}$ в сферической системе координат, которая одна и та же для любых r и φ_1).

Теперь перейдем обратно к декартовой системе координат, положив $r \cos \varphi_1 = u$ и $r \sin \varphi_1 = v$. Якобиан этого преобразования равен r^{-1} , и мы имеем

$$C \int_{-\infty}^{\infty} du \int_0^{\infty} dv \frac{e^{i|t|u}}{(1+u^2+v^2)^{\frac{n+1}{2}}} v^{n-2}.$$

Вычисляем интеграл по v :

$$\begin{aligned} \int_0^{\infty} \frac{v^{n-2} dv}{(1+u^2+v^2)^{\frac{n+1}{2}}} &= \frac{1}{(n-1)(1+u^2)} \cdot \frac{v^{n-1}}{(1+u^2+v^2)^{\frac{n-1}{2}}} \Big|_0^{\infty} = \\ &= \frac{1}{(n-1)(1+u^2)}. \end{aligned}$$

Получаем

$$C_1 \int_{-\infty}^{\infty} \frac{e^{i|t|u} du}{1+u^2},$$

где $C_1 = \frac{C}{n-1}$, т. е. вычисление характеристической функции n -мерного распределения Коши свелось к одномерному случаю. Так как характеристическая функция равна единице при $t=0 = (0, \dots, 0)$, то можно не уточнять значение постоянной C . Окончательно получаем

$$Ee^{i(t, \xi)} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{i|t|u} du}{1+u^2} = e^{-|t|} \quad (\text{см. П. 4.2 (д)}).$$

В отличие от одномерного случая $|t| = \sqrt{t_1^2 + \dots + t_n^2}$ и, следовательно, любое маргинальное распределение — это распределение Коши. Для доказательства достаточно приравнять нулю соответствующие координаты вектора t .

Распределение Коши можно получить с помощью проекции поверхности сферы на касательную плоскость, осуществляемой лучом, проходящим из центра сферы. Пробегая все точки поверхности сферы, луч осуществляет взаимно однозначное соответствие полусферы и плоскости. Пусть теперь на полусфере задано равномерное распределение вероятностей. Это такое распределение, что значение вероятности на любой области полусферы пропорционально площади этой области. Заданная проекция индуцирует вероятностное распределение на плоскости. Можно показать, что это индуцированное распределение и есть распределение Коши с размерностью, на единицу меньшей, чем размерность сферы.

Очевидно, плотность $p_{\xi}(x_1, \dots, x_n)$ не равна произведению n плотностей одномерных распределений Коши. Поэтому компоненты случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ — зависимые случайные величины. Ковариация компонент ξ_i, ξ_j ($i \neq j$) не определена. Для численной оценки зависимости можно применить информационный коэффициент корреляции (см. П.5.3 (в)).

Подставляя в $\int_{R^2} p_{\xi_1, \xi_2}(x, y) \log \frac{p_{\xi_1, \xi_2}(x, y)}{p_{\xi_1}(x) p_{\xi_2}(y)} dx dy$ значения двумерных и одномерных плотностей Коши, получим, что

$$J(\xi_1, \xi_2) = \log \frac{\pi}{2} + 4 \log 2 - 3 = 0.2242 \dots,$$

откуда

$$\rho(\xi_1, \xi_2) = \sqrt{1 - e^{-2J(\xi_1, \xi_2)}} = 0.6 \dots$$

П.5.5. Условные распределения относительно случайных величин

В п. П.2.1 мы определили условную вероятность $P(A|B)$ события A относительно события B как отношение $\frac{P(AB)}{P(B)}$, предполагая, что $P(B) > 0$. Рассмотрим условные вероятности относительно события B для всех $A \in \mathcal{F}$. Очевидно, функция множеств $P(\cdot | B)$ подчиняется условиям 1) и 2) из П.3.1. Таким образом, мы получаем новое вероятностное пространство $(\Omega, \mathcal{F}, P(\cdot | B))$. Особенностью вероятностной меры $P(\cdot | B)$ является лишь то, что она сосредоточена на множестве B (которое предполагается измеримым, т. е. $B \in \mathcal{F}$), так как

$$P(B|B) = \frac{P(B)}{P(B)} = 1.$$

Рассмотрим случайную величину ξ , т. е. измеримую функцию, заданную на измеримом пространстве (Ω, \mathcal{F}) . При новой вероятностной мере $P(\cdot | B)$ у нее будет новое распределение $\mu_{\xi}(\cdot | B)$, где для любого $a \in R'$

$$\mu_{\xi}((-\infty, a) | B) = P(\xi < a | B),$$

которое может характеризоваться функцией распределения $F_{\xi}(\cdot | B)$, плотностью $p_{\xi}(\cdot | B)$ и т. д. Интеграл Лебега от ξ по этой мере (если он существует) называется условным математическим ожиданием относительно B и обозначается $E(\xi | B)$.

Таким образом, все понятия, связанные с безусловной вероятностью P , остаются без изменений при замене P на $P(\cdot | B)$. Принципиально новое начинается тогда, когда мы рассматриваем условную вероятность как функцию от условия, т. е. от множества B . Заметим, что условная вероятность $P(A | B)$ как функция от B не является вероятностной мерой. Например, если $A = \Omega$, то $P(A | B) = P(A | \bar{B}) = 1$, в то время, как $B\bar{B} = \emptyset$. В то же время условные вероятности $P(A | B)$ при различных B не могут быть заданы произвольно.

Пусть Ω разбито на конечное число непересекающихся частей: $\Omega = \bigcup_{i=1}^n B_i$, $B_i B_j = \emptyset$ ($i \neq j$) и B сложено из этих же частей:

$$B = \bigcup_{k=1}^m B_{i_k},$$

где $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$.

Тогда

$$P(A | B) P(B) = \sum_{k=1}^m P(A | B_{i_k}) P(B_{i_k}).$$

Правая часть напоминает определение математического ожидания дискретной случайной величины φ , имеющей распределение

$$\begin{pmatrix} a_1, \dots, a_m, & 0 \\ P(B_{i_1}), \dots, P(B_{i_m}), & 1 - P(B) \end{pmatrix}$$

где $a_k = P(A | B_{i_k})$ и $P(\varphi = a_k) = P(B_{i_k})$.

Эта аналогия, а также потребность в определении условных вероятностей относительно случайных величин приводит к новому подходу к условной вероятности. В современной теории вероятностей и в ее приложениях значительно более важную роль, чем условные вероятности относительно отдельных событий, играют условные вероятности относительно систем событий. Примером раздела теории вероятностей, который основан на условных вероятностях относительно системы событий, является теория марковских процессов.

В предыдущем примере в качестве этой системы событий удобно взять разбиение $\mathfrak{B} = (B_1, \dots, B_n)$ и определить условную вероятность $P(A | \mathfrak{B})$ относительно этого разбиения. При этом $P(A | \mathfrak{B})$ считается такой случайной величиной φ , которая принимает постоянные значения на каждом B_i , равные $P(A | B_i)$.

Тогда

$$P(AB) = P(A|B)P(B) = \sum_{B_i \in \mathfrak{B}} P(A|B_i)P(\varphi = P(A|B_i)) = \\ = \int_B \varphi P(d\omega) = \int_B P(A|\mathfrak{B})P(d\omega) = E(P(A|\mathfrak{B}) \cdot I_B).$$

Это определение теряет свою ясность при переходе от конечных разбиений к бесконечным, и в частности к несчетным. Нас будут интересовать в первую очередь разбиения, производимые на множестве элементарных событий Ω случайными величинами, принимающими несчетное множество значений.

Пусть ξ — такая случайная величина, которая принимает все значения из интервала (a, b) . Несчетным разбиением будет разбиение $\mathfrak{B}_\xi = \{\{\xi = x\} : x \in (a, b)\}$.

Условной вероятностью события A относительно разбиения \mathfrak{B}_ξ , производимого случайной величиной ξ (или относительно случайной величины ξ), называется такая случайная величина γ , измеримая относительно σ -алгебры \mathcal{F}_ξ , что для любого $B \in \mathcal{F}_\xi$

$$P(AB) = \int_B \gamma P(d\omega).$$

В этом определении \mathcal{F}_ξ означает « σ -алгебра, порожденная случайной величиной ξ ». Так называется под- σ -алгебра σ -алгебры \mathcal{F} , содержащая все события вида $\{\xi < a\}$ ($a \in R^1$), причем это наименьшая из всех таких σ -алгебр. Случайной величиной, измеримой относительно некоторой σ -алгебры $\mathcal{F}_1 \subset \mathcal{F}$, называется функция φ , определенная на Ω , такая, что для любого $a \in R^1$ $\{\varphi < a\} \in \mathcal{F}_1$.

В тех случаях, которые представляют интерес, \mathcal{F}_ξ не совпадает с \mathcal{F} . Условную вероятность события A относительно случайной величины ξ обозначают как $P(A|\xi)$. Значения этой новой случайной величины в точке $\omega \in \Omega$ обозначаются как $P(A|\xi)(\omega)$.

Основное равенство, определяющее эту условную вероятность,

$$P(AB) = E(P(A|\xi) \cdot I_B),$$

где $B \in \mathcal{F}_\xi$. Возникает вопрос о существовании таких условных вероятностей для данных $A \in \mathcal{F}$ и случайных величин ξ . Положительный ответ на этот вопрос дается известной теоремой Радона—Никодима (Колмогоров, 1974, с. 72).

Следующим шагом является определение *условных распределений вероятностей относительно случайной величины ξ* . Так называется функция $P(\cdot|\xi)$ (отсутствующий аргумент ее обозначен точкой) такая, что при любом $A \in \mathcal{F}$ $P(A|\xi)$ — это определенная выше условная вероятность события A относительно случайной величины ξ , а для любого $\omega \in \Omega$ (точнее, почти для всех $\omega \in \Omega$, т. е. для всех ω , за исключением ω , принадлежащих множеству, вероятность которого равна 0) функция $P'(\cdot) = P(\cdot|\xi)(\omega)$ —

это вероятностное распределение на (Ω, \mathcal{F}) . Не слишком ограничительные достаточные условия существования таких условных распределений вероятностей относительно ξ приведены, например, в книге Гихмана и Скорохода (1971, с. 51).

Если определено условное распределение вероятностей, то естественно определить *условное математическое ожидание случайной величины η относительно случайной величины ξ* , которое обозначается $E(\eta|\xi)$ и равно $\int_{\Omega} \eta P(d\omega|\xi)$, а также условное распределение случайной величины η относительно случайной величины ξ . Мы будем обозначать это распределение как $\mu_{\eta}(\cdot|\xi)$.

Таким образом,

$$\mu_{\eta}(A|\xi) = P(\eta \in A|\xi),$$

а

$$E(\eta|\xi) = \int_{R^1} x \mu_{\eta}(dx|\xi).$$

Заметим, что $P(A|\xi)$ можно представить в виде $E(I_A|\xi)$. Поэтому иногда в качестве исходного понятия определяют условное математическое ожидание, а потом уже условные вероятности и условные распределения вероятностей.

Рассмотренные выше определения условных вероятностей, распределений и математических ожиданий относительно случайной величины ξ без существенных изменений переносятся на случай векторной случайной величины $\xi = (\xi_1, \dots, \xi_n)$. Таким образом, например, $P(A|\xi_1, \dots, \xi_n)$ — это случайная величина, измеримая относительно σ -алгебры \mathcal{F}_{ξ} (т. е. наименьшей под- σ -алгебры σ -алгебры \mathcal{F} , содержащей все множества вида $\{\xi_1 < a_1\}, \dots, \{\xi_n < a_n\}$ ($a_1, \dots, a_n \in R^n$), и такая, что для любого $B \in \mathcal{F}_{\xi}$

$$P(AB) = E(P(A|\xi_1, \dots, \xi_n) \cdot I_B).$$

Рассмотрим $E(\eta|\xi)$ для двух частных случаев, когда ξ имеет дискретное и абсолютно непрерывное распределение.

Если ξ — «дискретная» случайная величина,

$$\mu_{\xi}(a_i) > 0 \text{ и } \sum_{i=1}^{\infty} \mu_{\xi}(a_i) = 1,$$

где $\{a_1, a_2, \dots\} \subset R^1$, то $E(\eta|\xi)$ также принимает счетное множество значений:

$$E(\eta|\xi = a_1), E(\eta|\xi = a_2), \dots,$$

где $E(\eta|\xi = a_i)$ — условное математическое ожидание относительно события $\{\xi = a_i\}$, имеющего положительную вероятность, причем

$$E(\eta|\xi = a_i) = \int_{\Omega} \eta P(d\omega|\xi = a_i).$$

То же относится к условному распределению вероятностей $P(\cdot | \xi)$ или к условному распределению $\mu_\eta(\cdot | \xi)$ случайной величины η .

Например, если условное распределение случайной величины η абсолютно непрерывно для тех $\omega \in \Omega$, когда $\xi = a_i$, то определена условная плотность распределения $p_\eta(x | \xi = a_i)$ ($x \in R^1$) относительно события $\{\xi = a_i\}$, причем она может не существовать для других a_j , для которых условное распределение $\mu_\eta(\cdot | \xi = a_j)$ может быть дискретно. Если оно дискретно, то его можно задать таблицей

$$\begin{pmatrix} b_1 & b_2 & \dots \\ (\mu_\eta(b_1 | \xi = a_j), \mu_\eta(b_2 | \xi = a_j), \dots) \end{pmatrix},$$

где

$$\mu_\eta(b_i | \xi = a_j) > 0 \text{ и } \sum_{i=1}^{\infty} \mu_\eta(b_i | \xi = a_j) = 1.$$

Если ξ имеет абсолютно непрерывное распределение с плотностью $p_\xi(x)$ ($x \in R^1$), то задать $E(\eta | \xi)$ набором условных математических ожиданий относительно событий вида $\{\xi = x\}$ невозможно, так как их вероятности все равны нулю. Известно лишь, что $E(\eta | \xi)$ — это \mathcal{F}_ξ -измеримая функция (случайная величина). Можно показать (Ито, 2, 1963, с. 7), что такая случайная величина представима в виде $\varphi(\xi(\omega))$ (функция φ от функции ξ от ω) и, значит, если $\xi(\omega) = x$, то $E(\eta | \xi)(\omega) = \varphi(x)$. Чтобы не усложнять обозначений, и в этом случае пишут

$$E(\eta | \xi)(\omega) = E(\eta | \xi = x).$$

Почти при всех $x \in R^1$ функция $E(\eta | \xi = x)$ обладает всеми свойствами математического ожидания. Например, если $\eta = a_1\eta_1 + a_2\eta_2$, где η_1 и η_2 некоторые случайные величины, то

$$E(\eta | \xi = x) = a_1 E(\eta_1 | \xi = x) + a_2 E(\eta_2 | \xi = x) \text{ (линейность).}$$

Далее,

$$\inf_{\omega \in \Omega} \eta(\omega) \leq E(\eta | \xi = x) \leq \sup_{\omega \in \Omega} \eta(\omega)$$

и, если $\eta(\omega) \equiv 1$, то $E(\eta | \xi = x) = 1$ (подробнее о свойствах условных математических ожиданий см.: Ито, 2, 1963, с. 8).

$E(\eta | \xi = x)$ можно представить в виде интеграла по условному распределению вероятностей относительно ξ при значении $\xi = x$, т. е. по мере $P(\cdot | \xi = x)$. Смысл этого обозначения при $P(\xi = x) = 0$, как и для математического ожидания, отличается от случая $P(\xi = x) > 0$.

Пусть условное распределение $\mu_\eta(\cdot | \xi = x)$ (т. е. условное распределение случайной величины η относительно ξ при тех ω , для которых $\xi(\omega) = x$) является абсолютно непрерывным и $p_\eta(\cdot | \xi = x)$ — плотность этого распределения. Предположим

также, что существует плотность $p_{\eta, \xi}$ распределения случайного вектора (η, ξ) . Тогда

$$p_{\eta}(y | \xi = x) = \frac{p_{\eta, \xi}(y, x)}{p_{\xi}(x)}.$$

Действительно, вычислим условное математическое ожидание $E(f(\eta) | \xi)$. Имеем

$$\begin{aligned} E(f(\eta) \cdot I_{\{\xi \in A\}}) &= \int_{R^2} f(y) I_A(x) p_{\eta, \xi}(y, x) dy dx = \\ &= \int_A \left(\int_{R^1} f(y) \frac{p_{\eta, \xi}(y, x)}{p_{\xi}(x)} dy \right) p_{\xi}(x) dx. \end{aligned}$$

С другой стороны,

$$E(f(\eta) \cdot I_{\{\xi \in A\}}) = \int_{\{\xi \in A\}} E(f(\eta) | \xi) P(d\omega) = \int_A E(f(\eta) | \xi = x) p_{\xi}(x) dx.$$

Так как в качестве A можно взять любое множество $A \in \mathcal{F}$, то

$$E(f(\eta) | \xi = x) = \int_{R^1} f(y) \frac{p_{\eta, \xi}(y, x)}{p_{\xi}(x)} dy.$$

В частности,

$$P(\eta < a | \xi = x) = E(I_{\{\eta < a\}} | \xi = x) = \int_{-\infty}^a \frac{p_{\eta, \xi}(y, x)}{p_{\xi}(x)} dy.$$

Отсюда, согласно определению плотности условного распределения,

$$p_{\eta}(y | \xi = x) = \frac{p_{\eta, \xi}(y, x)}{p_{\xi}(x)}.$$

Эта формула для плотности условного распределения (или условной плотности, как ее иногда называют) аналогична формуле условной вероятности относительно события положительной вероятности.

В качестве примера вычислим плотности условных распределений одной координаты двумерного случайного вектора относительно другой координаты, когда этот вектор имеет распределение: а) нормальное, б) Коши.

а). Пусть

$$p_{\xi, \eta}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \cdot \frac{1}{1-\rho^2} \left(\frac{(x-a_1)^2}{\sigma_1^2} - \frac{2\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2} \right)\right),$$

где $\sigma_1, \sigma_2 > 0$; $-1 < \rho < 1$; $a_1, a_2 \in R^1$.

Тогда

$$p_{\xi}(x) = \int_{-\infty}^{\infty} p_{\xi, \eta}(x, y) dy = \frac{1}{\sqrt{2\pi} \sigma_1} e^{-\frac{(x-a_1)^2}{2\sigma_1^2}},$$

$$p_{\eta}(y | \xi = x) = \frac{1}{\sqrt{2\pi} \sigma_2 \sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)} \left((y-a_2) - (x-a_1) \frac{\rho\sigma_2}{\sigma_1} \right)^2}.$$

Таким образом, условное распределение η относительно ξ — это нормальное распределение с параметрами $\left((x-a_1) \frac{\rho\sigma_2}{\sigma_1}, \sigma_2 \sqrt{1-\rho^2} \right)$. Так как в определении плотности нормального распределения координаты равноправны, то условное распределение ξ относительно η также нормальное с соответствующими параметрами.

б). Пусть

$$p_{\xi, \eta}(x, y) = \frac{1}{2\pi(1+x^2+y^2)^{3/2}}.$$

Тогда, как было показано ранее (см. П. 5. 4 (в)),

$$p_{\xi}(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

и, следовательно,

$$p_{\eta}(y | \xi = x) = \frac{1+x^2}{2(1+x^2+y^2)^{3/2}}.$$

Отсюда видно, что условное распределение η относительно ξ уже не является распределением Коши. В частности, существует условное математическое ожидание $E(\eta | \xi = x)$, равное нулю при любом $x \in R^1$, так как условная плотность симметрична относительно нуля и

$$\int_{-\infty}^{\infty} \frac{|y| dy}{(1+x^2+y^2)^{3/2}} = 2 \int_0^{\infty} \frac{y dy}{(1+x^2+y^2)^{3/2}} = \frac{-2}{(1+x^2+y^2)^{1/2}} \Big|_0^{\infty} = \frac{2}{1+x^2} < \infty.$$

Аналогично получаем, что $E(\xi | \eta = y) = 0$.

Пусть случайная величина ξ может принимать любые значения из интервала (a, b) и кроме того задано совместное распределение случайных величин η и ξ такое, что для всех $x \in (a, b)$ существует $E(\eta | \xi = x)$. Функция $\varphi(x) = E(\eta | \xi = x)$ называется кривой регрессии η на ξ . Аналогично кривая $\psi(y) = E(\xi | \eta = y)$ при $y \in (c, d)$, где η может принимать любые значения из интервала (c, d) , называется кривой регрессии ξ на η .

Если ξ и η — независимые случайные величины, то обе кривые регрессии представляются собой прямые линии, проходящие через точку $(E\xi, E\eta)$ перпендикулярно друг другу: $E(\eta | \xi = x) = E\eta$ — прямая, параллельная оси абсцисс, $E(\xi | \eta = y) = E\xi$ — прямая, параллельная оси ординат. Однако, как показывает пример дву-

мерного распределения Коши, такая форма и расположение кривых регрессии не достаточны для того, чтобы ξ и η были независимыми.

Кривые регрессии обладают важным свойством минимальности.

Пусть требуется найти такую функцию $f(x)$, чтобы средний квадрат расстояния от точки (ξ, η) до точки $(\xi, f(\xi))$ был минимальным. Заметим, что обе эти точки случайны (случайные векторы), но если в первом случае распределение случайной точки «размазано», вообще говоря, по плоскости, то во втором оно сосредоточено на кривой $y=f(x)$. Такой функцией будут $f(x)=E(\eta | \xi=x)$.

Действительно, пусть f_1 — измеримая функция. Тогда

$$\begin{aligned} E(\eta - f_1(\xi))^2 &= \int_{R^2} (y - f_1(x))^2 \mu_{\xi, \eta}(dx, dy) = \\ &= \int_{R^1} \mu_{\xi}(dx) \int_{R^1} (y - f_1(x))^2 \mu_{\eta}(dy | \xi = x) = \\ &= \int_{R^1} \mu_{\xi}(dx) \int_{R^1} (y - E(\eta | \xi = x) + E(\eta | \xi = x) - f_1(x))^2 \cdot \mu_{\eta}(dy | \xi = x) = \\ &= \int_{R^1} (D(\eta | \xi = x) + (E(\eta | \xi = x) - f_1(x))^2) \mu_{\xi}(dx) \geq \int_{R^1} D(\eta | \xi = x) \mu_{\xi}(dx), \end{aligned}$$

где $D(\eta | \xi = x) = E((\eta - E(\eta | \xi = x))^2 | \xi = x)$ — дисперсия распределения $\mu_{\eta}(\cdot | \xi = x)$.

В связи с обсуждением условных вероятностей относительно случайных величин вернемся к вопросу стохастической независимости. Как было определено в П. 2. 2, случайные величины ξ и η называются стохастически независимыми, если для любых $a_1, a_2 \in R_1$

$$P(\{\xi < a_1\} \cap \{\eta < a_2\}) = P(\xi < a_1) P(\eta < a_2).$$

Отсюда следует, что для любого измеримого $A \subset R^1$

$$\begin{aligned} P(\{\xi \in A\} \cap \{\eta < a_2\}) &= E(P(\eta < a_2 | \xi) \cdot I_{\{\xi \in A\}}) = \\ &= E(P(\eta < a_2) \cdot I_{\{\xi \in A\}}) = P(\eta < a_2) P(\xi \in A). \end{aligned}$$

Так как постоянная функция является \mathcal{F}_{ξ} -измеримой, то отсюда следует, что $P(\eta < a_2 | \xi) = P(\eta < a_2)$ — условная вероятность события $\{\eta < a_2\}$ относительно случайной величины ξ равна безусловной вероятности того же события. Аналогично получаем, что $P(\xi < a_1 | \eta) = P(\xi < a_1)$, и так как a_1, a_2 — произвольны, то условное распределение одной случайной величины относительно другой равно безусловному распределению.

Рассмотрим случайный вектор $\xi = (\xi_1, \dots, \xi_n)$. Компоненты этого вектора — случайные величины ξ_1, \dots, ξ_n — называются взаимно стохастически независимыми (или стохастически независимыми в совокупности), если для любых $a_1, \dots, a_n \in R^1$

$$P\left(\bigcap_{i=1}^n \{\xi_i < a_i\}\right) = \prod_{i=1}^n P(\xi_i < a_i).$$

Как показывает пример, принадлежащий Бернштейну, из попарной стохастической независимости компонент вектора ξ не следует их взаимная стохастическая независимость. Пусть $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$,

$$P(\omega_i) = \frac{1}{4} \quad (i = 1, 2, 3, 4), \quad A_1 = \{\omega_1, \omega_4\}, \quad A_2 = \{\omega_2, \omega_4\}, \quad A_3 = \{\omega_3, \omega_4\}.$$

Тогда

$$P(A_j) = \frac{1}{2} \quad (j = 1, 2, 3) \quad \text{и} \quad P(A_1 A_2) = P(A_1 A_3) = P(A_2 A_3) = P(\omega_4) = \frac{1}{4},$$

т. е. события из набора $\{A_1, A_2, A_3\}$ попарно стохастически независимы. Однако $P(A_1 A_2 A_3) = P(\omega_4) = \frac{1}{4}$, в то время как $P(A_1) \times P(A_2) P(A_3) = \frac{1}{8}$. Следовательно, эти события не являются взаимно стохастически независимыми, и, значит, то же можно сказать о случайных величинах $I_{A_1}, I_{A_2}, I_{A_3}$.

Нетрудно показать, что случайные величины ξ_1, \dots, ξ_n стохастически независимы в совокупности, если для любого $i \in \{1, \dots, n\}$

$$P(\xi_i < a_i | \xi_1, \dots, \xi_{i-1}) = P(\xi_i < a_i),$$

т. е. условное распределение каждой компоненты относительно случайного вектора, составленного из предыдущих компонент, равно безусловному распределению этой компоненты (т. е. соответствующему маргинальному распределению). Действительно, если указанное свойство имеет место, то

$$\begin{aligned} P\left(\bigcap_{i=1}^n \{\xi_i < a_i\}\right) &= E\left(P_i(\xi_n < a_n | \xi_1, \dots, \xi_{n-1}) \cdot I_{n-1}\left(\bigcap_{i=1}^{n-1} \{\xi_i < a_i\}\right)\right) = \\ &= P(\xi_n < a_n) P\left(\bigcap_{i=1}^{n-1} \{\xi_i < a_i\}\right) = \prod_{i=1}^n P(\xi_i < a_i). \end{aligned}$$

В общем случае для задания распределения вектора ξ достаточно знать распределение первой компоненты μ_{ξ_1} и условные распределения $\mu_{\xi_i}(\cdot | \xi_1, \dots, \xi_{i-1})$ ($i = 2, \dots, n$). При этом

$$\begin{aligned} P\left(\bigcap_{i=1}^n \{\xi_i < a_i\}\right) &= \mu_{\xi}((-\infty, a_1) \times \dots \times (-\infty, a_n)) = \\ &= \int_{(-\infty, a_1) \times \dots \times (-\infty, a_n)} \mu_{\xi_1}(dx_1) \mu_{\xi_2}(dx_2 | \xi_1 = x_1) \dots \\ &\quad \dots \mu_{\xi_n}(dx_n | (\xi_1, \dots, \xi_{n-1}) = (x_1, \dots, x_{n-1})), \end{aligned}$$

где $\mu_{\eta}(\cdot | \xi = x)$, когда ξ — векторная случайная величина имеет тот же смысл, что и для одномерной случайной величины (Гихман, Скороход, 1, 1971, с. 56).

Распределение вектора ξ называется *марковским*, если условные распределения $\mu_{\xi_i}(\cdot | \xi_1, \dots, \xi_{i-1})$ ($i = 2, \dots, n$) совпадают с условными распределениями $\mu_{\xi_i}(\cdot | \xi_{i-1})$.

Таким образом, для марковского распределения

$$P\left(\bigcap_{i=1}^n \{\xi_i < a_i\}\right) = \int_{(-\infty, a_1) \times \dots \times (-\infty, a_n)} \mu_{\xi_1}(dx_1) \mu_{\xi_2}(dx_2 | \xi_1 = x_1) \dots \mu_{\xi_n}(dx_n | \xi_{n-1} = x_{n-1}).$$

Марковские распределения конечномерных и бесконечномерных векторов (последние называются *цепями Маркова*) подробно обсуждаются в гл. IV и V.

II.6. ПРЕОБРАЗОВАНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Рассмотрим некоторые преобразования случайных величин. Если ξ — n -мерная векторная случайная величина и f — измеримая функция, заданная на R^n со значениями в R^m , то $f(\xi)$ — m -мерная векторная случайная величина с распределением $\mu_{f(\xi)}$, которое вычисляется по правилу

$$\mu_{f(\xi)}(A) = \mu_{\xi}(f^{-1}(A)) \quad (A \in \mathcal{B}^m),$$

где $f^{-1}(A) = \{(x_1, \dots, x_n) : f(x_1, \dots, x_n) \in A\}$. (см. II.5.2)

Как было отмечено выше, многомерные функции распределения находят ограниченное применение потому, что не существует удобных формул, связывающих функции распределения преобразованных и не преобразованных векторных случайных величин. Более удобны в этом отношении плотности распределений.

Пусть μ_{ξ} — абсолютно непрерывное распределение и p_{ξ} — плотность распределения случайного вектора ξ .

Тогда

$$\mu_{f(\xi)}(A) = \int_{f^{-1}(A)} p_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Пусть $m = n$ и $g(x_1, \dots, x_n) = (g_1(x_1, \dots, x_n), \dots, g_n(x_1, \dots, x_n))$, где g_i — гладкие функции. Как известно из курса дифференциального и интегрального исчисления (Фихтенгольц, 1960, с. 388), предыдущий интеграл в этом случае равен

$$\int_{g^{-1}(f^{-1}(A))} p_{\xi}(g(x_1, \dots, x_n)) |J_g(x_1, \dots, x_n)| dx_1 \dots dx_n,$$

где $J_g(x_1, \dots, x_n)$ — так называемый якобиан преобразования g , равный определителю матрицы Якоби $(a_{ij}(x_1, \dots, x_n))_{n \times n}$, где $a_{ij}(x_1, \dots, x_n) = \frac{\partial g_i}{\partial x_j}$ (частная производная вычислена в точке

(x_1, \dots, x_n)). Множество $g^{-1}(f^{-1}(A))$ совпадает с A при всех $A \in \mathcal{B}^n$ в том и только в том случае, если $g^{-1} = f$. При этом

$$\int_A p_{\xi}(f^{-1}(x_1, \dots, x_n)) |J_{f^{-1}}(x_1, \dots, x_n)| dx_1 \dots dx_n$$

и, следовательно,

$$p_{f(\xi)}(x_1, \dots, x_n) = p_{\xi}(f^{-1}(x_1, \dots, x_n)) |J_{f^{-1}}(x_1, \dots, x_n)|. \quad (\text{II. 6. 1})$$

Это общая формула преобразования плотности распределения, соответствующего преобразованию f векторной случайной величины в случае, если обратное преобразование f^{-1} имеет гладкие компоненты.

II.6.1. Линейное преобразование

Пусть $f = (f_1, \dots, f_n)$, где

$$f_i(x_1, \dots, x_n) = a_i + \sum_{j=1}^n b_{ji} x_j \quad (i = 1, \dots, n).$$

Это преобразование можно записать в матричном виде как

$$f(x) = a + xB,$$

где $a = (a_1, \dots, a_n) \in R^n$, $B = (b_{ij})_{n \times n}$ — матрица. Отображение f обратимо в том и только в том случае, когда определитель матрицы B не равен нулю. При этом

$$f^{-1}(x) = (x - a) B^{-1},$$

где B^{-1} — матрица, обратная B . Заметим, что

$$J_{f^{-1}}(x_1, \dots, x_n) = (\det B^{-1}) = (\det B)^{-1} \neq 0.$$

По формуле (II.6.1) плотность распределения случайной величины $a + \xi B$ связана с плотностью p_{ξ} соотношением

$$p_{f(\xi)}(x_1, \dots, x_n) = p_{\xi} \left(\sum_{j=1}^n (x_j - a_j) b'_{j1}, \dots, \sum_{j=1}^n (x_j - a_j) b'_{jn} \right) (\det B)^{-1},$$

где $B^{-1} = (b'_{ij})$.

Пусть $n = 2$, $f_1(x_1, x_2) = x_1 + x_2$ и $f_2(x_2) = x_2$.

Тогда

$$a(0, 0), \quad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \text{ и } \det B = 1.$$

Поэтому

$$B^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \text{ и } p_{f(\xi)}(x_1, x_2) = p_{\xi}(x_1 - x_2, x_2).$$

Отсюда мы получаем плотность распределения суммы двух случайных величин. Действительно,

$$p_{\xi_1+\xi_2}(x_1) = p_{f_1(\xi)}(x_1) = \int_{-\infty}^{\infty} p_{f(\xi)}(x_1, x_2) dx_2 = \int_{-\infty}^{\infty} p_{\xi}(x_1 - x_2, x_2) dx_2.$$

Пусть ξ_1 и ξ_2 независимы.

Тогда, очевидно,

$$p_{\xi}(x_1, x_2) = p_{\xi_1}(x_1) p_{\xi_2}(x_2)$$

и

$$p_{\xi_1+\xi_2}(x_1) = \int_{-\infty}^{\infty} p_{\xi_1}(x_1 - x_2) p_{\xi_2}(x_2) dx_2. \quad (\text{II. 6. 2})$$

Операция получения из двух функций третьей по правилу, приведенному в правой части формулы (II.6.2.), а также результат этой операции называется сверткой этих двух функций. Таким образом, плотность распределения суммы двух независимых случайных величин равна свертке плотностей распределений этих случайных величин. Для этой операции над функциями принято обозначение $p_{\xi_1+\xi_2} = p_{\xi_1} * p_{\xi_2}$.

Очевидно, $p_{\xi_1} * p_{\xi_2} = p_{\xi_2} * p_{\xi_1}$.

Кроме того, как нетрудно показать, операция свертки подчиняется сочетательному закону, поэтому последовательное применение свертки можно записать без скобок в виде

$$p_{\xi_1} * p_{\xi_2} * \dots * p_{\xi_n} = \underset{i=1}{*}^n p_{\xi_i},$$

и если все p_{ξ_i} равны друг другу, то $\underset{i=1}{*}^n p_{\xi_i} = p^{n*}$, где $p_{\xi_i} = p$.

Формула, аналогичная (II.6.2), справедлива для распределений общего вида. В терминах функций распределений она записывается в виде

$$F_{\xi_1+\xi_2}(x) = \int_{-\infty}^{\infty} F_{\xi_1}(x - x_1) dF_{\xi_2}(x_1),$$

где интеграл Стильтьеса, стоящий справа, равен интегралу Лебега от функции $F_{\xi_1}(x - x_1)$ по мере $\mu_{\xi_2}(dx_1)$ (замечание об интеграле Стильтьеса см. в II.3.2). Эта операция над функциями распределения, а также аналогичная операция над самими распределениями, соответствующая сложению независимых случайных величин, называется композицией функций распределения (а также соответствующих распределений), или, так же как и для плотностей, — сверткой. Пишут также, что

$$F_{\xi_1+\xi_2} = F_{\xi_1} * F_{\xi_2}, \quad \mu_{\xi_1+\xi_2} = \mu_{\xi_1} * \mu_{\xi_2} \text{ и т. д.}$$

Так как характеристическая функция суммы двух независимых случайных величин равна произведению характеристических функ-

ций этих случайных величин, то операции свертки над функциями распределения соответствует операция произведения над характеристическими функциями. Справедливо и обратное соответствие. Тем же свойством обладают преобразование Лапласа и производящая функция распределений случайных величин.

Рассмотрим еще одно линейное преобразование:

$$f(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Если ξ_1, \dots, ξ_n — независимые случайные величины, то независимыми будут и $\frac{1}{n} \xi_1, \dots, \frac{1}{n} \xi_n$ и потому

$$p_{f(\xi)}(x) = \left(\prod_{i=1}^n p_{\frac{1}{n} \xi_i} \right) (x).$$

Если все p_{ξ_i} одинаковы, то

$$p_{f(\xi)}(x) = p_{\frac{1}{n} \xi_1}^{*n}(x).$$

Пусть $\varphi_{\xi}(t) = Ee^{it\xi}$ — характеристическая функция случайной величины ξ .

Тогда

$$\varphi_{\frac{1}{n} \xi_1}(t) = Ee^{\frac{it\xi_1}{n}} = \varphi_{\xi_1}\left(\frac{t}{n}\right)$$

и

$$\varphi_{f(\xi)}(t) = \left(\varphi_{\frac{1}{n} \xi_1}(t) \right)^n = \left(\varphi_{\xi_1}\left(\frac{t}{n}\right) \right)^n.$$

Если ξ_1 имеет распределение Коши, то $\varphi_{\xi_1}(t) = e^{-|t|}$ (см. II. 4. 2 (д)) и, следовательно,

$$\varphi_{\frac{1}{n} \sum_{i=1}^n \xi_i}(t) = \left(e^{-\left| \frac{t}{n} \right|} \right)^n = e^{-|t|}.$$

Отсюда делаем вывод, что средние из n независимых случайных величин, распределенных по закону Коши, также имеют распределение Коши. Как мы увидим дальше, этот вывод ставит распределение Коши в особое положение по отношению к большинству практически используемых распределений, для которых дисперсия среднего арифметического стремится к нулю, так что само это среднее арифметическое стремится в некотором смысле к постоянной величине, равной математическому ожиданию исходных случайных величин.

Пусть, например, ξ_1 имеет нормальное распределение со средним 0 и дисперсией 1.

Тогда $\varphi_{\xi_1}(t) = e^{-t^2/2}$ и, следовательно, $\varphi\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) = e^{-t^2/2n} \rightarrow 1$ ($n \rightarrow \infty$), что соответствует сходимости к нулю случайной величины $\frac{1}{n} \sum_{i=1}^n \xi_i$, так как $e^{t \cdot 0} = 1$.

П.6.2. Некоторые нелинейные преобразования

а. Произведение случайных величин. Пусть

$$f(x_1, x_2) = (x_1 \cdot x_2, x_2).$$

Предположим для определенности, что $x_1 > 0$, т. е. полуплоскость отображается на всю плоскость.

Тогда

$$f^{-1}(x_1, x_2) = \left(\frac{x_1}{x_2}, x_2 \right) \quad (x_2 \neq 0)$$

и

$$J_{f^{-1}}(x_1, x_2) = \begin{vmatrix} \frac{1}{x_2} & 0 \\ -\frac{x_1}{x_2^2} & 1 \end{vmatrix} = \frac{1}{x_2}.$$

Следовательно,

$$p_{f(\xi_1, \xi_2)}(x_1, x_2) = p_{(\xi_1, \xi_2)}\left(\frac{x_1}{x_2}, x_2\right) \left| \frac{1}{x_2} \right|.$$

Отсюда

$$p_{\xi_1 \cdot \xi_2}(x_1) = \int_{-\infty}^{\infty} p_{(\xi_1, \xi_2)}\left(\frac{x_1}{x_2}, x_2\right) \left| \frac{1}{x_2} \right| dx_2,$$

и если ξ_1 и ξ_2 независимы, то

$$p_{\xi_1 \cdot \xi_2}(x_1) = \int_{-\infty}^{\infty} p_{\xi_1}\left(\frac{x_1}{x_2}\right) p_{\xi_2}(x_2) \left| \frac{1}{x_2} \right| dx_2.$$

Эта формула, так же как и формула для свертки двух плотностей, может быть обобщена на случай распределений общего вида. Очевидно, формула верна и для $x_1 < 0$.

б. Отношение двух случайных величин. Пусть

$$f(x_1, x_2) = \left(\frac{x_1}{x_2}, x_2 \right).$$

Это отображение задано всюду на плоскости R^2 , кроме оси абсцисс, когда $x_2 = 0$. Предположим сначала, что $x_1 > 0$.

Тогда

$$f^{-1}(x_1, x_2) = (x_1 x_2, x_2)$$

и

$$J_{f^{-1}}(x_1, x_2) = \begin{vmatrix} x_2 & 0 \\ x_1 & 1 \end{vmatrix} = x_2.$$

Следовательно,

$$P_{f(\xi_1, \xi_2)}(x_1, x_2) = P_{(\xi_1, \xi_2)}(x_1 x_2, x_2) |x_2|.$$

Отсюда

$$P_{\xi_1/\xi_2}(x_1) = \int_{-\infty}^{\infty} P_{(\xi_1, \xi_2)}(x_1 x_2, x_2) |x_2| dx_2.$$

Та же формула справедлива для $x_1 < 0$. Если ξ_1 и ξ_2 независимы, то

$$P_{\xi_1/\xi_2}(x_1) = \int_{-\infty}^{\infty} p_{\xi_1}(x_1 x_2) p_{\xi_2}(x_2) |x_2| dx_2.$$

Так как, согласно предположению, распределение случайного вектора (ξ_1, ξ_2) абсолютно непрерывно, то можно не выбрасывать точку $x_2 = 0$ из области интегрирования вследствие того, что интеграл по множеству $\{0\}$ равен нулю.

Пусть случайный вектор $\xi = (\xi_1, \xi_2)$ имеет невырожденное нормальное распределение с вектором математических ожиданий (a_1, a_2) и ковариационной матрицей

$$\begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix},$$

где

$$\sigma_i > 0, \quad -1 < \rho < 1.$$

Тогда

$$P_{\xi_1/\xi_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{1}{1-\rho^2} \left(\left(\frac{x_1-a_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-a_1}{\sigma_1}\right)\left(\frac{x_2-a_2}{\sigma_2}\right) + \left(\frac{x_2-a_2}{\sigma_2}\right)^2\right)\right) \quad (\text{см. II. 5. 4б}).$$

Плотность распределения случайной величины $\eta = \frac{\xi_1}{\xi_2}$ равна

$$P_{\eta}(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} |z| \exp\left(-\frac{1}{2} \frac{1}{1-\rho^2} \left(\left(\frac{xz-a_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{xz-a_1}{\sigma_1}\right)\left(\frac{z-a_2}{\sigma_2}\right) + \left(\frac{z-a_2}{\sigma_2}\right)^2\right)\right) dz.$$

Групируя члены в показателе и выделяя полный квадрат, приводим это выражение к виду

$$C \int_{-\infty}^{\infty} |z| e^{-A(x-B)^2} dz,$$

где A, B, C зависят от параметров распределения и от x , причем $A > 0$. Заметим, что последний интеграл не изменится, если заменить B на $-B$, и, следовательно, он равен

$$\begin{aligned} C \int_{-\infty}^{\infty} |z| e^{-A(x-|B|)^2} dz &= \\ &= -C \int_{-\infty}^0 z e^{-A(x-|B|)^2} dz + C \int_0^{\infty} z e^{-A(x-|B|)^2} dz = \\ &= -2C \int_{-\infty}^0 z e^{-A(x-|B|)^2} dz + C \int_{-\infty}^{\infty} z e^{-A(x-|B|)^2} dz = \\ &= -2C \int_{-\infty}^0 (z - |B|) e^{-A(x-|B|)^2} dz - 2C|B| \int_{-\infty}^0 e^{-A(x-|B|)^2} dz + \\ &+ C \int_{-\infty}^{\infty} (z - |B|) e^{-A(x-|B|)^2} dz + C|B| \int_{-\infty}^{\infty} e^{-A(x-|B|)^2} dz. \end{aligned}$$

Третий интеграл, очевидно, равен нулю. Осуществляя замену переменных — в первом интеграле $(z - |B|)^2 = u$, во втором $A(z - |B|)^2 = \frac{t^2}{2}$ — и замечая, что четвертый интеграл равен $\sqrt{\frac{\pi}{A}}$, получаем

$$\begin{aligned} C \int_{\frac{B^2}{2}}^{\infty} e^{-Au} du - \frac{2|B|C}{\sqrt{2A}} \int_{-\infty}^{-|B|\sqrt{2A}} e^{-\frac{t^2}{2}} dt + |B|C \sqrt{\frac{\pi}{A}} = \\ = \frac{C}{A} e^{-AB^2} - 2|B|C \sqrt{\frac{\pi}{A}} \Phi(-|B|\sqrt{2A}) + |B|C \sqrt{\frac{\pi}{A}}, \end{aligned}$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ называется функцией Лапласа, которая

не выражается через элементарные функции; для ее значений — обычно в виде $\Phi(x) - \frac{1}{2}$ — составлены таблицы, которые можно найти в любом справочнике по теории вероятностей и статистике. В итоге мы имеем

$$p_{\eta}(x) = \frac{C}{A} e^{-AB^2} + |B|C \sqrt{\frac{\pi}{A}} (1 - 2\Phi(-|B|\sqrt{2A})).$$

В результате простых, но громоздких вычислений получаем следующие значения коэффициентов:

$$A = \frac{\sigma_1^2 - 2x\sigma_1\sigma_2\rho + x^2\sigma_2^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)} > 0,$$

$$B = a_2 - \sigma_2 \frac{(a_2x - a_1)(x\sigma_2 - \rho\sigma_1)}{\sigma_1^2 - 2x\sigma_1\sigma_2\rho + x^2\sigma_2^2},$$

$$C = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{(a_2x - a_1)^2}{2(\sigma_1^2 - 2x\sigma_1\sigma_2\rho + x^2\sigma_2^2)}\right).$$

В общем виде это выражение трудно обозримо. Однако при $a_1 = a_2 = 0$, $B = 0$ и $C = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}$ оно значительно упрощается. В этом случае

$$p_\eta(x) = \frac{1}{\pi} \frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{x^2\sigma_2^2 - 2\rho x\sigma_1\sigma_2 + \sigma_1^2}.$$

Нетрудно показать, что это выражение является плотностью случайной величины

$$\frac{\sigma_1\sqrt{1-\rho^2}}{\sigma_2} \xi_0 + \rho \frac{\sigma_1}{\sigma_2},$$

где ξ_0 имеет стандартное распределение Коши. Следовательно, p_η — плотность распределения Коши с параметрами сдвига $\rho \frac{\sigma_1}{\sigma_2}$ и масштабом $\frac{\sigma_1\sqrt{1-\rho^2}}{\sigma_2}$.

Интересно проанализировать эту плотность при $a_2 \rightarrow \infty$. При этом $B \rightarrow \infty$ и

$$p_\eta(x) \sim C |B| \sqrt{\frac{\pi}{A}} = \frac{1}{\sqrt{2\pi}} \frac{a_2\sigma_1^2 - x\sigma_1\sigma_2\rho + \sigma_2^2 a_1x - a_1\sigma_1\sigma_2\rho}{(\sigma_1^2 - 2x\sigma_1\sigma_2\rho + x^2\sigma_2^2)^{3/2}} \times \\ \times \exp\left(-\frac{(a_2x - a_1)^2}{2(\sigma_1^2 - 2x\sigma_1\sigma_2\rho + x^2\sigma_2^2)}\right).$$

Если обозначить

$$f(x) = \frac{a_2x - a_1}{\sqrt{\sigma_1^2 - 2x\sigma_1\sigma_2\rho + x^2\sigma_2^2}},$$

то

$$f'(x) = \frac{a_2\sigma_1^2 - x\sigma_1\sigma_2\rho + \sigma_2^2 a_1x - a_1\sigma_1\sigma_2\rho}{(\sigma_1^2 - 2x\sigma_1\sigma_2\rho + x^2\sigma_2^2)^{3/2}}$$

и, следовательно,

$$p_\eta(x) \sim \frac{1}{\sqrt{2\pi}} f'(x) e^{-\frac{1}{2}(f(x))^2} = \frac{d}{dx} \Phi(f(x)).$$

Отсюда

$$P(\eta < x) \sim \Phi\left(\frac{a_2x - a_1}{\sqrt{\sigma_1^2 - 2x\sigma_1\sigma_2\rho + x^2\sigma_2^2}}\right).$$

Последняя функция является функцией распределения случайной величины $g(\xi_0)$, где ξ_0 имеет стандартное нормальное распределение (с параметрами $(0; 1)$), а $g=f^{-1}$ — нелинейное преобразование.

Семейство функций распределения несколько более общего вида

$$\Phi\left(\frac{x-a}{\sqrt{\pi(x)}}\right),$$

где $\pi(x)$ — полином, называемый дисперсионной функцией, рассмотрено Бернштейном (Bernstein, 1925).

Полученная оценка функции распределения отношения ξ_1/ξ_2 двух случайных величин, имеющих нормальное распределение, при больших значениях $E \xi_2 = a_2$ может применяться на практике в том случае, когда a_2/σ_2 достаточно велико, так что можно с большой практической достоверностью предположить, что ξ_2 принимает только положительные значения.

Пусть известно, например, что ξ_2 не может принимать отрицательные значения и нуль. Вместе с тем распределение ξ_2 по крайней мере в окрестности математического ожидания близко к нормальному. Если отношение математического ожидания к стандартному отклонению этого нормального закона достаточно велико, то при расчетах принимают распределение случайной величины ξ_2 равным этому нормальному, пренебрегая тем фактом, что при этом допускаются с положительной вероятностью (очень малой) отрицательные значения величины ξ_2 . В такой ситуации хорошим приближением для распределения отношения ξ_1/ξ_2 , где ξ_1 также распределено нормально, является полученное выше распределение Бернштейна.

В геологической практике с такими случайными величинами приходится иметь дело при решении задач о распределении состава горной породы, где ξ_1 и ξ_2 или абсолютные величины массы двух компонент в данном объеме, или относительные величины, выраженные в процентах или в других долях. Так, при определении абсолютного возраста породы устанавливается содержание в ней двух изотопов. Возраст породы является функцией отношения этих количеств. Если содержание изотопов варьирует от одной пробы к другой, сохраняя некоторую статистическую устойчивость, то естественно считать его случайным вектором. Распределение этого вектора можно оценить. Хорошие оценки получаются при предположении, что абсолютное содержание изотопов имеет нормальное распределение. При этом можно полагать, что распределение отношения содержания изотопов очень близко к рассмотренному выше распределению Бернштейна.

в. Процентный пересчет. Так мы будем называть преобразование

$$f(x_1, \dots, x_n) = \left(\frac{x_1}{\sum_{i=1}^n x_i}, \dots, \frac{x_n}{\sum_{i=1}^n x_i} \right),$$

когда $\sum_{i=1}^n x_i > 0$. При таком преобразовании по вектору абсолютного содержания компонент, скажем, в данном образце горной породы определяется вектор относительного содержания компонент, который обычно выражается в процентах. В этом случае мы не будем давать выражение для преобразованной плотности, а ограничимся исследованием распределения величины $\eta = \frac{\xi_1}{\sum_{i=1}^n \xi_i}$, когда известно распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$.

Пусть случайный вектор ξ имеет нормальное распределение с вектором математических ожиданий $\mathbf{a} = (a_1, \dots, a_n)$ и ковариационной матрицей $\Lambda = (\lambda_{ij})_{n \times n}$. В важных для геохимии случаях, когда вектор ξ отражает абсолютное содержание компонент, значения величин ξ_i не могут быть отрицательными. Однако их распределения можно считать нормальными, так как они близки к нормальным с очень большими отношениями математических ожиданий к стандартным отклонениям. При выводе приближенного распределения случайной величины η мы будем учитывать это предположение. Из него следует, что вероятность события $\left\{ \sum_{i=1}^n \xi_i \leq 0 \right\}$ пренебрежимо мала. При этом

$$P(\eta < x) \sim P((1-x)\xi_1 - x\xi_2 - \dots - x\xi_n < 0) = P(\xi \cdot \mathbf{L}^T < 0),$$

где $\mathbf{L} = (1-x, -x, \dots, -x)$ — матрица $1 \times n$ линейного преобразования пространства R^n в пространство R^1 . Согласно (II. 5. 4 (б)), случайная величина $\xi \cdot \mathbf{L}^T$ имеет нормальное распределение с математическим ожиданием $b = \mathbf{a} \cdot \mathbf{L}^T$ и дисперсией $\sigma^2 = \mathbf{L} \cdot \Lambda \cdot \mathbf{L}^T$, которая положительна, если матрица Λ положительно определена. Несложный подсчет показывает, что

$$\mathbf{a} \mathbf{L}^T = a_1 - x \sum_{i=1}^n a_i$$

и

$$\mathbf{L} \cdot \Lambda \cdot \mathbf{L}^T = x^2 \sum_{i,j=1}^n \lambda_{ij} - 2x \sum_{i=1}^n \lambda_{i1} + \lambda_{11}.$$

Следовательно,

$$P(\xi \cdot L^r < 0) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^0 e^{-\frac{(t-b)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-b/\sigma} e^{-u^2/2} du =$$

$$= \Phi \left(\frac{x \sum_{i=1}^n a_i - a_1}{\sqrt{x^2 \sum_{i,j=1}^n \lambda_{ij} - 2x \sum_{i=1}^n \lambda_{i1} + \lambda_{11}}} \right).$$

Итак, мы получили, что распределение случайной величины η при указанных выше предположениях приблизительно равно распределению Бернштейна. Этим же способом можно вывести приблизительное распределение для отношения ξ_1/ξ_2 , которое мы получили выше в предположении, что $E\xi_2$ достаточно велико, из точного выражения для плотности распределения этого отношения.

Исследуем свойства симметрии распределения Бернштейна. Мы называем функцию распределения F симметричной, если при любом $x > 0$

$$F(a-x) = 1 - F(a+x),$$

где a — медиана распределения ($F(a) = 1/2$).

Говорят, что функция распределения имеет правую асимметрию, если для любых $x > 0$

$$F(a-x) < 1 - F(a+x),$$

и левую асимметрию, если

$$F(a-x) > 1 - F(a+x).$$

Рассмотрим распределение Бернштейна с функцией распределения

$$F(x) = \Phi \left(\frac{x-a}{\sqrt{\alpha x^2 - 2\beta x + \gamma}} \right),$$

где $\alpha x^2 - 2\beta x + \gamma > 0$ при всех x . Так как $\Phi(0) = 1/2$, то a является медианой данного распределения.

Очевидно, $F(a-x) < 1 - F(a+x)$ тогда и только тогда, когда

$$-\frac{(a-x)-a}{\sqrt{\alpha(a-x)^2 - 2\beta(a-x) + \gamma}} > \frac{(a+x)-a}{\sqrt{\alpha(a+x)^2 - 2\beta(a+x) + \gamma}},$$

так как функция нормального распределения Φ симметрична относительно своей медианы, равной нулю, и строго возрастает на всей прямой. Последнее неравенство эквивалентно неравенству

$$\alpha(a-x)^2 - 2\beta(a-x) + \gamma < \alpha(a+x)^2 - 2\beta(a+x) + \gamma,$$

т. е. $\beta x < \alpha x^2$.

Так как $x > 0$, то $\beta < \alpha$. Следовательно, при этом условии функция распределения имеет правую асимметрию. Аналогично при $\beta > \alpha$ имеет место левая асимметрия, а при $\beta = \alpha$ — симметрия функции F .

Для приближенной функции распределения отношения $\xi_1 / \sum_{i=1}^n \xi_i$, выведенной выше, когда все ξ_i распределены нормально и $\sum_{i=1}^n a_i > 0$, имеем

$$a = \frac{a_1}{\sum_{i=1}^n a_i}, \quad \alpha = \sum_{i,j=1}^n \lambda_{ij} \left/ \left(\sum_{i=1}^n a_i \right)^2 \right.,$$

$$\beta = \sum_{i=1}^n \lambda_{i1} \left/ \left(\sum_{i=1}^n a_i \right)^2 \right., \quad \gamma = \lambda_{11} \left/ \left(\sum_{i=1}^n a_i \right)^2 \right..$$

Условие правой асимметрии для этой функции распределения имеет вид

$$\frac{\sum_{i=1}^n \lambda_{i1}}{\sum_{i,j=1}^n \lambda_{ij}} < \frac{a_1}{\sum_{i=1}^n a_i}.$$

Заметим, что вывод распределения Бернштейна для нормально распределенного вектора с положительными математическими ожиданиями справедлив также для *вырожденного нормального распределения*. В этом случае квадратичная форма $x \cdot \Lambda \cdot x^T$ ($x \in R^n$, Λ — матрица $n \times n$) при некоторых x может обращаться в нуль. Поэтому при выводе распределения Бернштейна нужно убедиться в том, что дисперсионная функция отлична от нуля. В рассматриваемом случае для этого достаточно, чтобы дискриминант $\beta^2 - \alpha\gamma$ квадратичной формы $\alpha x^2 - 2\beta x + \gamma$ был меньше нуля,

что для отношения $\xi_1 / \sum_{i=1}^n \xi_i$ приводит к условию

$$\left(\sum_{i=1}^n \lambda_{i1} \right)^2 < \lambda_{11} \cdot \sum_{i,j=1}^n \lambda_{ij}.$$

Вырожденное нормальное распределение случайного вектора ξ будет, например, при условии $\sum_{i=1}^n \xi_i = C$, где C — некоторое число. Очевидно, в этом случае процентное преобразование сводится к изменению масштаба исходных случайных величин и нормальность распределения сохраняется.

Процентное преобразование случайных величин, имеющих совместное распределение, отличное от нормального (например, логарифмически нормальное или многомерное гамма-распределение), рассмотрено в работе Фааса и Сарманова (1968).

Пример II.4. Генетические соотношения между концентрациями элементов и процентный пересчет

Если взять минерал, основа кристаллической решетки которого образована кремнекислородными тетраэдрами, и зафиксировать объем пробы, то очевидно, что колебания в содержании, скажем, редких щелочей, входящих в изоморфные примеси, не связаны с количеством Si в пробе. Построение совместного распределения по большому числу однородных проб покажет, что количество Si постоянно, а содержание щелочи варьирует. Если произвести процентный пересчет или «замкнуть сумму анализа», скажем, считая содержание компонентов в граммах на тонну или в числе частиц на миллион, то график совместного распределения покажет линейную связь между содержаниями Si и изучаемой щелочи; коэффициент корреляции между ними будет -1 . Это хорошо известно минералагам и кристаллохимикам, которые при пересчете анализа дают цифры, отнесенные к числу атомов кислорода, входящих в элементарную ячейку.

Каково влияние процентного пересчета в случае минералов, имеющих некоторый постоянный элемент (скажем, число атомов кислорода), ясно. Однако как поступать при изучении горных пород, исследуя соотношения между составляющими их химических анализов, неизвестно. Все способы пересчета в петрохимии (Харкера, Нигли и Заварицкого) этот вопрос игнорируют. Исключительная важность его была понята Бартом, предложившим пересчитывать анализ на постоянное число атомов в некоторой кислородной ячейке. Это предложение, в основе совершенно верное, иногда отвергалось из-за споров об упаковке кислородных атомов.

Обобщая сказанное, следует отметить, что геохимиков и петрологов, особенно изучающих редкие и рассеянные элементы, совершенно не интересуют содержания элементов, представляющие их относительное количество в некотором постоянном весе. Эти отношения сами по себе несут искаженную информацию о геохимических явлениях. Интерес представляют отношения между количеством элементов, содержащихся в некотором элементарном объеме. Элементарным объемом мы называем объем такого минимального куба, для которого вероятность того, что свойства породы находятся в пределах, определяющих ее принадлежность к данному классу пород, больше заданного, близкого к единице уровня (например, больше 0.999). Очевидно, что в абсолютных единицах элементарные объемы будут различны —

в крупнозернистых породах они больше, в мелкозернистых — меньше. Однако для данного типа породы и ее фациальной разности, именно для элементарного объема, функция распределения вероятностей химических компонентов или минералов несет наименее искаженную информацию о геохимических явлениях.

Для практического изучения породы нужно разбивать ее на некоторые части (скажем, кубы), из которых каждая заведомо охватывает элементарный объем, и изучать результаты не пересчитанных, а непосредственно определенных (взвешенных)

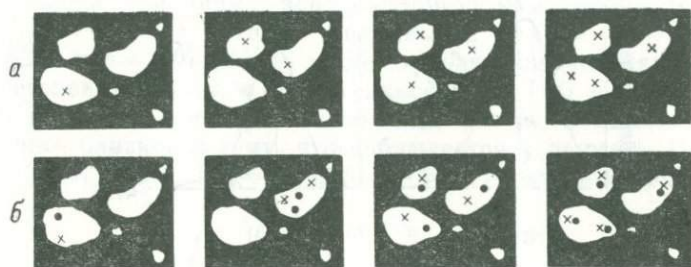


Рис. 11.4. Ложная в генетическом смысле корреляция, возникающая после процентного пересчета.

a — между весовыми содержаниями U и V нет стохастической зависимости; после процентного пересчета появляется линейная зависимость с коэффициентом корреляции $r_{U,V} = -1$.

b — между весовыми содержаниями U и V существует линейная зависимость с $r_{U,V} = +1$, после процентного пересчета появляется линейная зависимость с $r_{U,V} = -1$.

Черное — U_3O_8 , белое — поры; площадь, занятая U_3O_8 , на всех рисунках одна и та же (увел. 44 000 \times , материал Японской электронно-оптической лаборатории). \times — V , одна крестик — одна весовая единица; \bullet — вторичный U , внедрившийся в поры вместе с V , одна точка — одна весовая единица.

количество элементов в этих объемах. В этом случае результаты анализов допускают наиболее простую предметную трактовку. Подчеркнем еще раз, что это особенно важно при изучении поведения в породах и минералах редких и рассеянных элементов. Эти элементы могут занимать в породе весьма разнообразное положение, а пересчет на постоянную сумму смазывает это различие. Рисунок 11.4 поясняет сказанное. Разъяснения имеются также в статье Иванова и Подольского (1971).

Ниже приводятся численные примеры того, как влияет на маргинальные распределения, в частности на их симметрию, процентный пересчет величин, имеющих многомерное нормальное распределение в элементарном объеме.

a). Случайный вектор $(\xi_1, \xi_2, \xi_3, \xi_4)$, представляющий веса компонент в элементарном объеме, имеет математические ожидания $E\xi_1=70$, $E\xi_2=14$, $E\xi_3=4$ и $E\xi_4=25 \cdot 10^{-6}$ и ковариационную матрицу

$$\Lambda = \begin{pmatrix} 25 & 10 & 5 & -4 \cdot 10^{-5} \\ 10 & 4 & 2 & 0 \\ 5 & 2 & 1 & 0 \\ -4 \cdot 10^{-5} & 0 & 0 & 64 \cdot 10^{-12} \end{pmatrix}$$

(эти данные примерно отвечают составу элементарного объема грубозернистого гранита в граммах, где ξ_1 — SiO_2 , ξ_2 — Al_2O_3 , ξ_3 — K_2O и ξ_4 — рассеянный элемент).

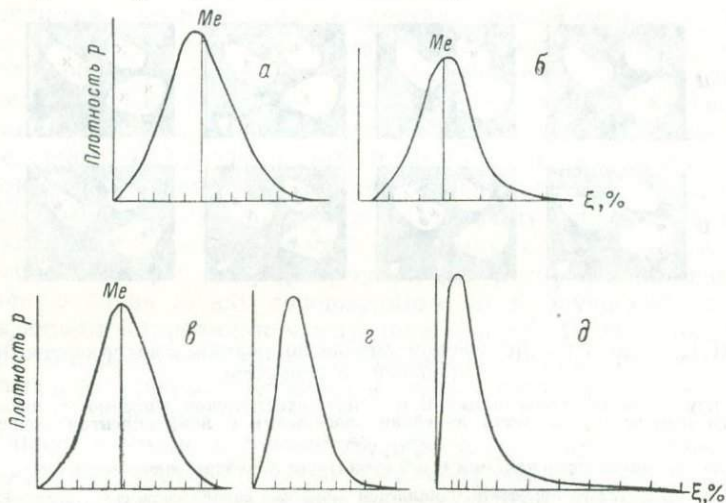


Рис. 11.5. Возникновение асимметрии под влиянием процентного пересчета. а — рассеянный элемент в аналоге гранита, умеренная правая асимметрия; б — то же для другой ковариационной матрицы; в — изменена ковариационная матрица, симметричная плотность; г — рассеянный элемент в аналоге диамта, заметная правая асимметрия; д — рассеянный элемент в аналоге руды, резко выраженная правая асимметрия. Me — медиана.

Тогда случайная величина $\eta_4 = \left(\xi_4 / \sum_{i=1}^4 \xi_i \right) \cdot 10^7$ приблизительно имеет распределение Бернштейна вида

$$P(\eta_4 < y) = \Phi \left(\frac{y - 2.84}{\sqrt{0.0083y^2 + 0.1033y + 0.8264}} \right).$$

В проверяемом неравенстве $\beta < 0$ и $a \cdot \alpha > 0$. Следовательно, $\beta < a \cdot \alpha$ и функция распределения имеет правую асимметрию (см. рис. II. 5, а).

б). Пусть случайный вектор $(\xi_1, \xi_2, \xi_3, \xi_4)$ имеет нормальное распределение с тем же вектором математических ожиданий, что и в II. 5, а, но с другой ковариационной матрицей

$$\Lambda = \begin{pmatrix} 25 & -5 & -2.5 & 2 \cdot 10^{-5} \\ -5 & 4 & 0 & 0 \\ -2.5 & 0 & 1 & 4 \cdot 10^{-6} \\ 2 \cdot 10^{-5} & 0 & 4 \cdot 10^{-6} & 64 \cdot 10^{-12} \end{pmatrix}.$$

В этом случае $a=2.84$, $\alpha=0.0019$, $\beta=0.0310$, $\gamma=0.8264$. Так как $a \cdot \alpha=0.0053$, то $\beta > a \cdot \alpha$ и, следовательно, имеет место левая асимметрия (рис. П.5, б).

в). Снова изменим ковариационную матрицу, положив $\lambda_{14}=\lambda_{41}=-2 \cdot 10^{-5}$. Тогда получим $\alpha=0.0019$, $\beta=-0.0206$, $\gamma=0.8264$ и $a=2.84$. Имеет место правая асимметрия; функция распределения близка к симметричной (рис. П.5, в).

г). Рассмотрим теперь соотношение тех же компонент, более или менее близкое к тому, что наблюдается у дацитов. При этом $E \xi_1=62$, $E \xi_2=20$, $E \xi_3=2$, $E \xi_4=25 \cdot 10^{-6}$ и матрица ковариаций

$$\Lambda = \begin{pmatrix} 100 & 20 & 6 & -8 \cdot 10^{-5} \\ 20 & 4 & 1.2 & 0 \\ 6 & 1.2 & 0.36 & 0 \\ -8 \cdot 10^{-5} & 0 & 0 & 64 \cdot 10^{-12} \end{pmatrix}.$$

Соответствующий закон распределения Бернштейна имеет параметры $a=2.84$, $\alpha=0.0192$, $\beta=-0.1033$, $\gamma=0.8264$. На рис. П.5, г видна еще более резкая, чем в предыдущих примерах, правая асимметрия.

д). Рассмотрим случай, когда в породе (или в руде) определяется только содержание рассеянного элемента и общее количество проанализированного вещества. Пусть ξ_1 соответствует общему количеству вещества, а ξ_2 — рассеянному элементу. Случайный вектор (ξ_1, ξ_2) распределен нормально с параметрами $E \xi_1=3$, $E \xi_2=3 \cdot 10^{-6}$ и

$$\Lambda = \begin{pmatrix} 1 & -0.95 \cdot 10^{-6} \\ -0.95 \cdot 10^{-6} & 1.21 \cdot 10^{-12} \end{pmatrix}.$$

Тогда случайная величина $\eta = \frac{\xi_2}{\xi_1} \cdot 10^6$ с большой точностью описывается распределением Бернштейна с параметрами $a=1$, $\alpha=0.111$, $\beta=-0.105$ и $\gamma=0.134$.

Соответствующая плотность приведена на рис. П.5, д. Она имеет очень резко выраженную правую асимметрию.

Нужно иметь в виду, конечно, что все сказанное о процентном пересчете относится к элементарному объему. Перенос полученных результатов на образец требует специального рассмотрения, что автор надеется сделать во второй части этой книги.

П.6.3. Рандомизация

Пусть ξ_1, \dots, ξ_n — случайные величины, определенные на вероятностном пространстве (Ω, \mathcal{F}, P) , а $\mathfrak{B} = \{A_1, \dots, A_n\}$ — разбиение пространства Ω на n частей, независимое от данных случайных величин. Это значит, что для любых i, j и $a \in R^1$ события A_i и $\{\xi_j < a\}$ независимы.

Рассматривается случайная величина $\xi = \sum_{i=1}^n \xi_i I_{A_i}$, которую мы называем рандомизированной случайной величиной, полученной из случайных величин ξ_i в результате «разыгрывания» номера в соответствии с независимым разбиением \mathfrak{B} . Из определения следует, что

$$\xi(\omega) = \xi_i(\omega),$$

если $\omega \in A_i$, так как $A_i \cap A_j = \emptyset$, если $i \neq j$.

Пусть F_1, \dots, F_n — функции распределения исходных случайных величин. Тогда по формуле полной вероятности

$$\begin{aligned} F(x) = P(\xi < x) &= P\left(\sum_{i=1}^n \xi_i I_{A_i} < x\right) = \sum_{i=1}^n P(A_i) P(\{\xi_i < x\} | A_i) = \\ &= \sum_{i=1}^n P(A_i) P(\xi_i < x) = \sum_{i=1}^n P(A_i) F_i(x), \end{aligned}$$

где $\sum_{i=1}^n P(A_i) = 1$. Такая операция получения взвешенного среднего из совокупности n функций распределения называется суперпозицией (или смесью) функций распределения. Таким образом, рандомизация случайных величин соответствует суперпозиция их функций распределения.

Термин «суперпозиция» мы применяем также к самому распределению рандомизированной случайной величины, к плотности распределения, если она существует, и к другим характеристикам распределения. Например, k -тый момент рандомизированной случайной величины равен суперпозиции k -тых моментов компонент:

$$E\left(\sum_{i=1}^n \xi_i I_{A_i}\right)^k = \sum_{i=1}^n P(A_i) E\xi_i^k.$$

Однако дисперсия рандомизированной случайной величины не равна суперпозиции дисперсий компонент:

$$\begin{aligned} D\left(\sum_{i=1}^n \xi_i I_{A_i}\right) &= \sum_{i,j=1}^n (E(\xi_i \xi_j) E(I_{A_i} I_{A_j}) - E\xi_i E\xi_j E I_{A_i} E I_{A_j}) = \\ &= \sum_{i,j=1}^n (E(\xi_i \xi_j) - E\xi_i E\xi_j) E(I_{A_i} I_{A_j}) + \end{aligned}$$

$$\begin{aligned}
& + \sum_{i,j=1}^n E\xi_i E\xi_j (E(I_{A_i} I_{A_j}) - E I_{A_i} E I_{A_j}) = \\
& = \sum_{i=1}^n P(A_i) D\xi_i + \sum_{i,j=1}^n E\xi_i E\xi_j \text{cov } I_{A_i} I_{A_j}.
\end{aligned}$$

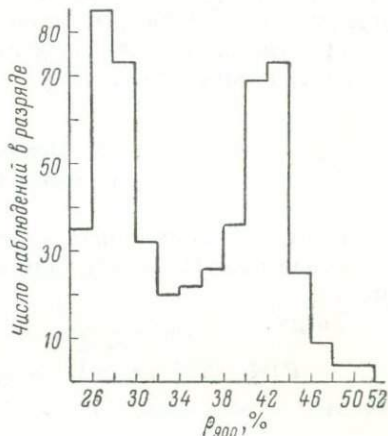
Так как ковариационная матрица случайного вектора $I = (I_{A_1}, \dots, I_{A_n})$ неотрицательно определена (см. II.5.3 (б)), то последнее слагаемое неотрицательно. Следовательно,

$$D\left(\sum_{i=1}^n \xi_i I_{A_i}\right) \geq \sum_{i=1}^n P(A_i) D\xi_i,$$

т.е. дисперсия рандомизированной случайной величины не меньше суперпозиции дисперсий компонент.

Рис. II.6. Распределение частот значений коэффициента спектральной яркости для длины волны 900 мкм (ρ_{900}) в современных песках Центральных Каракумов (Романова, 1971).

Суперпозиция одновыпуклых распределений, типичных для различных частей исследуемой территории.



В геологических задачах суперпозиция распределений встречается очень часто (Вистеллиус, 1948; Романова, 1971). В первой из этих работ при изучении формы обломочных частиц исследуется коэффициент окатанности Вадела. В рамках принятой модели этот коэффициент является случайной величиной. Распределение этой случайной величины, как следует из проведенной статистической обработки материала, хорошо описывается некоторым абсолютно непрерывным распределением с двугорбой кривой плотности. При более подробном анализе обнаруживается, что это распределение в основном является суперпозицией двух нормальных распределений:

$$F(x) = pF_1(x) + (1-p)F_2(x).$$

В данном случае удастся грубо оценить параметры компонент — функций распределения F_1 и F_2 , а также коэффициент усреднения p . Хорошим примером подобного исследования является работа Романовой (1971), где проводится разложение суперпозиции распределений спектральной яркости песков Кара-Кумов (рис. II.6).

П.7. ПОСЛЕДОВАТЕЛЬНОСТИ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН И ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

При составлении моделей геологических явлений иногда удобно представлять параметры модели в виде предела последовательности некоторых случайных величин. Ряд важных законов распределения находит обоснование с помощью предельного перехода. Почти вся техника математической статистики основана на законе больших чисел для последовательностей независимых случайных величин и, значит, на предельных теоремах. Из сказанного ясна роль раздела теории вероятностей, изучающего пределы последовательностей случайных величин. Мы затронем только две важнейшие предельные теоремы: закон больших чисел и центральную предельную теорему в ее простейшей форме.

П.7.1. Неравенство Чебышева и закон больших чисел

Пусть ξ — случайная величина, заданная на вероятностном пространстве (Ω, \mathcal{F}, P) , для которой существуют два первых момента.

Тогда

$$P(|\xi - E\xi| \geq \varepsilon) = \int_{\Omega} I_{\{|\xi - E\xi| \geq \varepsilon\}} P(d\omega) \leq \int_{\Omega} \frac{(\xi - E\xi)^2}{\varepsilon^2} P(d\omega),$$

так как при $\omega \in \{|\xi - E\xi| < \varepsilon\}$

$$I_{\{|\xi - E\xi| \geq \varepsilon\}}(\omega) = 0, \quad \frac{(\xi(\omega) - E\xi)^2}{\varepsilon^2} \geq 0,$$

а при $\omega \in \{|\xi - E\xi| \geq \varepsilon\}$

$$I_{\{|\xi - E\xi| \geq \varepsilon\}}(\omega) = 1, \quad \frac{(\xi(\omega) - E\xi)^2}{\varepsilon^2} \geq 1.$$

Полученное неравенство, которое можно переписать в виде

$$P(|\xi - E\xi| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} D\xi,$$

называется *неравенством Чебышева*.

Пусть (ξ_1, ξ_2, \dots) — последовательность независимых случайных величин. Это означает, что для любого n случайные величины ξ_1, \dots, ξ_n взаимно независимы. Пусть эти случайные величины имеют ограниченные вторые моменты. Рассмотрим среднее арифметическое из n первых членов этой последовательности

$$\frac{1}{n} \sum_{i=1}^n \xi_i.$$

В этом случае имеем

$$E\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right) = \frac{1}{n}\sum_{i=1}^n E\xi_i.$$

Очевидно,

$$D(c\xi) = c^2 D\xi,$$

где c — постоянная величина. Отсюда по независимости величин ξ_i

$$D\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right) = \frac{1}{n^2} D\left(\sum_{i=1}^n \xi_i\right) = \frac{1}{n^2} \sum_{i=1}^n D\xi_i.$$

Пусть все $D\xi_i \leq M < \infty$.

Тогда

$$D\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right) \leq \frac{1}{n} M$$

и по неравенству Чебышева

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n \xi_i - \frac{1}{n}\sum_{i=1}^n E\xi_i\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \frac{1}{n} M.$$

Отсюда ясно, что при достаточно больших n вероятность события

$$\left\{\left|\frac{1}{n}\sum_{i=1}^n \xi_i - \frac{1}{n}\sum_{i=1}^n E\xi_i\right| \geq \varepsilon\right\}$$

при любом наперед заданном $\varepsilon > 0$ очень мала и может быть сделана сколь угодно малой при n , стремящемся к бесконечности.

Пусть все $E\xi_i$ равны a .

Тогда

$$\frac{1}{n}\sum_{i=1}^n E\xi_i = a$$

и предыдущее неравенство переписется в виде

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n \xi_i - a\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \frac{1}{n} M.$$

Ясно, что среднее из n первых случайных величин при $n \rightarrow \infty$ сходится в некотором смысле к постоянной величине a . Говорят, что последовательность случайных величин $(\xi_n)_{n=1}^{\infty}$ сходится по вероятности к случайной величине ξ , если для любого $\varepsilon > 0$ при $n \rightarrow \infty$

$$P(|\xi_n - \xi| \geq \varepsilon) \rightarrow 0.$$

Таким образом, последовательность случайных величин $(\eta_n)_1^\infty$, где $\eta_n = \frac{1}{n} \sum_{i=1}^n \xi_i$, при указанных предположениях сходится по

вероятности к постоянной величине a , а в более общем случае можно утверждать, что последовательность разностей $\eta_n - E\eta_n$ сходится по вероятности к нулю.

Доказанные свойства последовательностей независимых случайных величин являются частными случаями так называемого закона больших чисел. Закон больших чисел, в частности, устанавливает связь между вероятностью события и частотой появления события в последовательности однородных испытаний.

Действительно, пусть в последовательности независимых испытаний вероятность события A при i -том испытании имеет одну и ту же величину p для всех i . Обозначив $A_i = \{A \text{ при } i\text{-том испытании}\}$, мы можем записать

$$p = P(A_i) = EI_{A_i}.$$

Тогда I_{A_1}, I_{A_2}, \dots — последовательность независимых случайных величин. Частота ν_n события A в последовательности n первых испытаний, очевидно, равна $\frac{1}{n} \sum_{i=1}^n I_{A_i}$. Закон больших чисел в этом частном случае утверждает, что при любом сколь угодно малом $\varepsilon > 0$

$$P(|\nu_n - p| \geq \varepsilon) \rightarrow 0,$$

т. е. частота события A сходится по вероятности к величине вероятности события A при отдельном испытании. Значит, при очень большом n событие $\{|\nu_n - p| \geq \varepsilon\}$ будет событием очень малой вероятности.

Мы уже говорили (с. 25), что теорию вероятности с практикой сближают события очень малой и очень большой (близкой к единице) вероятности. В этом отношении закон больших чисел очень практичен, его выводы можно проверить, и они многократно проверялись. И если в какой-либо ситуации получаются выводы, противоречащие закону больших чисел, то в большинстве случаев причина этого не в том, что осуществляется событие малой вероятности, которое возможно теоретически, а в том, что в действительности оказываются не выполненными исходные предположения этого закона.

Закон больших чисел обосновывает частотную интерпретацию вероятности. Типичным статистическим ансамблем при такой интерпретации является последовательность независимых случайных величин. Определение независимости в свою очередь производится на основании свойств совместных распределений вероятностей. Таким образом, при построении теории вероятностей по Колмогорову понятие «вероятность» является первичным. Этот подход коренным образом отличается от подхода, предложен-

ного Мизесом (1930), который предлагал в качестве исходного понятия взять статистический ансамбль со всеми свойствами, которые в нашем понимании свойственны последовательности независимых случайных величин. Одна из характеристик этого ансамбля — частота некоторого события в бесконечной последовательности, взятой из этого ансамбля, — объявлялась вероятностью этого события. Теория Мизеса, несмотря на внешнюю привлекательность, содержит ряд противоречий, а ее исходные предпосылки в действительности трудно проверяются.

П.7.2. Центральная предельная теорема

При условиях закона больших чисел последовательность средних арифметических из случайных величин сходится по вероятности к неслучайной величине или, как говорят, к случайной величине, но с вырожденным распределением, сосредоточенным в одной точке. Если эта точка равна нулю, то оказывается, что, домножая на \sqrt{n} среднее арифметическое из n случайных величин, можно получить последовательность случайных величин, стремящихся к случайной величине с невырожденным распределением. Это невырожденное распределение во многих случаях оказывается нормальным, каковы бы ни были распределения исходных случайных величин. Эта сходимость настолько быстрая, а предельный закон настолько устойчив к различным изменениям исходных данных, что факт сходимости может быть проверен эмпирическим путем, а сам нормальный закон часто может служить хорошим описанием реальных распределений в тех случаях, когда по смыслу явления исследуемая случайная величина равна сумме нескольких независимых случайных величин.

Пусть $(\xi_n)_1^\infty$ — последовательность независимых, одинаково распределенных случайных величин с математическими ожиданиями $E\xi_i = a$ и дисперсиями $D\xi_i = \sigma^2$. Рассмотрим последовательность случайных величин $(\zeta_n)_1^\infty$, где

$$\zeta_n = \frac{\sum_{i=1}^n \xi_i - an}{\sigma \sqrt{n}}.$$

Очевидно, $E\zeta_n = 0$ и $D\zeta_n = 1$. Это так называемая *центрированная и нормированная* сумма случайных величин ξ_i . Необходимо показать, что последовательность функций распределения $(F_n)_1^\infty$, где $F_n(x) = P(\zeta_n < x)$, равномерно относительно x ($-\infty < x < \infty$) сходится к функции распределения Φ нормального закона с параметрами $(0, 1)$.

Последовательность распределений с такими свойствами называется асимптотически нормальной. Наиболее короткое доказательство получается, если использовать теорему о непрерывном соответствии между распределениями и их характеристическими функциями (Боровков, 1972, с. 276). Согласно этой теореме, до-

статочно доказать, что последовательность значений в любой точке t характеристических функций φ_{ξ_n} , где $\varphi_{\xi_n}(t) = Ee^{it\xi_n}$, сходится к $e^{-t^2/2}$, что является характеристической функцией стандартного нормального распределения.

По независимости и одинаковой распределенности случайных величин ξ_i имеем

$$\varphi_{\xi_n}(t) = \varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right),$$

где $\varphi(t) = Ee^{it(\xi_1 - a)}$.

Справедливо разложение

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\varphi''(0) + o(t^2),$$

где $\frac{o(x)}{x} \rightarrow 0$ при $x \rightarrow \infty$.

Очевидно,

$$\varphi(0) = 1, \quad \varphi'(0) = 0 \quad \text{и} \quad \varphi''(0) = \sigma^2 \quad (\text{см. II. 4. 2 (в)}).$$

Отсюда

$$\varphi(t) = 1 - \frac{t^2\sigma^2}{2} + o(t^2).$$

Следовательно, при $n \rightarrow \infty$

$$\begin{aligned} \log \varphi_{\xi_n}(t) &= n \log \left(1 - \frac{\sigma^2}{2} \left(\frac{t}{\sigma\sqrt{n}} \right)^2 + o\left(\frac{t^2}{n}\right) \right) = \\ &= n \left(-\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right) = -\frac{t^2}{2} + n \cdot o\left(\frac{t^2}{n}\right) \rightarrow -\frac{t^2}{2}. \end{aligned}$$

Отсюда $\varphi_{\xi_n}(t) \rightarrow e^{-t^2/2}$ и асимптотическая нормальность последовательности функций распределения $(F_n)_{n=1}^{\infty}$ доказана.

Доказанное свойство последовательности случайных величин $(\xi_n)_{n=1}^{\infty}$ носит название *центральной предельной теоремы*. Под этим же названием известно много результатов, устанавливающих сходимость распределений центрированных и нормированных сумм независимых (и не только независимых) случайных величин к нормальному распределению. Иногда центральной предельной теоремой называют также свойство сходимости к другим законам. Например, естественный вариант центральной предельной теоремы при сложении углов на плоскости состоит в сходимости последовательности распределений на окружности к равномерному распределению (см. II.4.2 (б)).

II.8. СЛУЧАЙНЫЕ ПРОЦЕССЫ И ГЕОМЕТРИЧЕСКИЕ ВЕРОЯТНОСТИ

До сих пор мы изучали распределения конечномерных случайных векторов. В предыдущем параграфе мы имели дело с бесконечномерными случайными векторами — последовательностями не-

зависимых случайных величин. Но и в этом случае мы интересовались распределениями лишь первых n элементов последовательности. В этом параграфе мы определим и изучим свойства распределений бесконечномерных случайных векторов, которые в зависимости от вида параметрического множества называют случайными последовательностями, случайными процессами, случайными полями и т.д.

Наиболее типичным объектом в геологии является, по-видимому, случайное поле, т.е. такой бесконечномерный случайный вектор, параметрическим множеством которого является двух- или трехмерное пространство, с каждой точкой которого — параметром — сопоставлена случайная величина. Однако в некоторых случаях моделью геологических явлений является случайный процесс, развивающийся во времени. Наиболее просты случайные последовательности, которые встречаются в геологических моделях как в первичном виде, например, случайная последовательность слоев в разрезе осадочной породы, так и в виде сечений случайных полей, например, последовательность зерен гранита, расположенных вдоль линейного сечения трехмерного случайного поля, которое представляет собой гранит в рамках естественной математической модели (Vistelius, 1972).

П.8.1. Случайные последовательности

Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) задана последовательность случайных величин $(\xi_n)_1^\infty = (\xi_1, \xi_2, \dots)$, называемая также *случайной последовательностью*. Согласно общей концепции случайных векторов, $\xi = (\xi_n)_1^\infty$ — это отображение Ω в бесконечномерное пространство R^∞ , т.е. в пространство всех числовых последовательностей вида $x = (x_1, x_2, \dots)$ ($x_i \in R^1$).

Борелевской σ -алгебры B^∞ подмножеств множества R^∞ называют наименьшую σ -алгебру, содержащую все так называемые цилиндрические множества, т.е. множества вида $S_1 \times S_2 \times \dots \times S_m \times R^1 \times R^1 \times \dots$ ($m=1, 2, \dots$; $S_i \in \mathcal{B}^1$), где только конечное число «сторон» бесконечномерного «прямоугольника» отлично от R^1 , т.е. от максимально возможной «длины стороны». Отображение ξ предполагается измеримым относительно \mathcal{B}^∞ и \mathcal{F} . Тогда это отображение — случайный бесконечномерный вектор ξ — индуцирует вероятностную меру μ_ξ на измеримом пространстве $(R^\infty, \mathcal{B}^\infty)$, которая называется распределением случайной последовательности ξ :

$$\mu_\xi(A) = P(\xi \in A) \quad (A \in \mathcal{B}^\infty).$$

При изучении распределений случайных последовательностей на первый план выступают конечномерные маргинальные распределения, которые называют *частными распределениями*.

Пусть $\{i_1, i_2, \dots, i_k\}$ — подмножество чисел из множества $Z_+ = \{1, 2, \dots\}$. Проекцией $\pi_{(i_1, i_2, \dots, i_k)}$, соответствующей множеству индексов $\{i_1, \dots, i_k\}$, называется отображение R^∞ в R^k , при котором

с вектором $(x_1, x_2, \dots) \in R^\infty$ сопоставляется вектор $(x_{i_1}, x_{i_2}, \dots, x_{i_k}) \in R^k$, составленный из координат с номерами i_1, \dots, i_k . Очевидно, $\pi_{(i_1, i_2, \dots, i_k)}$ — измеримое отображение относительно \mathcal{B}^k и \mathcal{B}^∞ , и, следовательно, $\pi_{(i_1, \dots, i_k)}(\xi) = (\xi_{i_1}, \dots, \xi_{i_k})$ является случайным вектором размерности k . Распределение случайного вектора $(\xi_{i_1}, \dots, \xi_{i_k})$, т. е. совместное распределение случайных величин $\xi_{i_1}, \dots, \xi_{i_k}$, называется частным распределением данного случайного бесконечномерного вектора ξ , соответствующим данному набору индексов.

В отличие от конечномерных случайных векторов распределения бесконечномерных случайных векторов, кроме самых простых случаев, нельзя задать ни в виде плотности на пространстве R^∞ , ни в виде таблицы, дающей в соответствие некоторым элементам R^∞ положительные значения вероятности. Задают распределения бесконечномерных случайных векторов путем задания всех его частных распределений. Очевидно, все частные распределения заданы, если заданы частные распределения для возрастающей последовательности подмножеств множества Z_+ , стремящейся к Z_+ , например, для такой: $\{1\}, \{1, 2\}, \{1, 2, 3\}, \dots$

Распределение $\mu_{\pi_{(i_1, \dots, i_k)}(\xi)}$ является маргинальным для распределения $\mu_{\pi_{(1, 2, \dots, m)}(\xi)}$, если $\{i_1, \dots, i_k\} \subset \{1, 2, \dots, m\}$, и, следовательно, оно может быть по нему однозначно восстановлено.

Итак, для задания распределения случайного вектора ξ строится последовательность конечномерных распределений $\mu^{(1)}, \mu^{(2)}, \mu^{(3)}, \dots$, где $\mu^{(k)}$ — k -мерное распределение, которое соответствует набору индексов $\{1, 2, \dots, k\}$. Эти распределения не могут быть совсем произвольны. Колмогоров (1974, с. 48) доказал, что существует одно и только одно распределение μ_ξ случайного вектора ξ , для которого для любого k

$$\mu_\xi(\pi_{(1, \dots, k)} \in A) = \mu^{(k)}(A) \quad (A \in \mathcal{B}^k)$$

в том и только в том случае, если каждое предыдущее распределение является маргинальным для последующего, т. е.

$$\mu^{(k)}(A) = \mu^{(k+1)}(\pi_{(1, \dots, k)} \in A)$$

для любого $k=1, 2, \dots$

Рассмотрим некоторые случайные последовательности.

а. Последовательность независимых случайных величин. В этом случае считаются известными все одномерные частные распределения $P(\xi_i \in A) = \mu_{\xi_i}(A)$ ($A \in \mathcal{B}^1$). Значения конечномерных частных распределений $\mu^{(k)}$ определяются на «прямоугольных» множествах вида $S = S_1 \times S_2 \times \dots \times S_k \in \mathcal{B}^k$ по правилу $\mu^{(k)}(S) = \prod_{i=1}^k \mu_{\xi_i}(S_i)$, которое является следствием независимости случайных величин ξ_1, \dots, ξ_k . Из этого правила следует

$$\mu^{(k)}(S) = \mu^{(k+1)}(S \times R^1) = \mu^{(k+1)}(\pi_{(1, \dots, k)} \in S),$$

колмогоровское условие согласованности частных распределений.

Итак, существует одно и только одно распределение μ_{ξ_i} , частные распределения которого совпадают с заданными.

Если при любом i ξ_i принимает только два значения, например, 1 и 2 с одними и теми же вероятностями

$$\mu_{\xi_i}(\{1\}) = p \text{ и } \mu_{\xi_i}(\{2\}) = 1 - p \quad (0 < p < 1),$$

то случайная последовательность $\xi = (\xi_n)_{n=1}^{\infty}$, состоящая из независимых случайных величин ξ_i , называется последовательностью Бернулли. Частные распределения последовательности Бернулли являются дискретными распределениями, сосредоточенными на точках $(\alpha_1, \dots, \alpha_n)$, где $\alpha_i \in \{1, 2\}$, вероятность которых

$$\mu_{\pi_{(1, \dots, n)}(\xi)}((\alpha_1, \dots, \alpha_n)) = \prod_{i=1}^n \mu_{\xi_i}(\{\alpha_i\}).$$

Мы называем последовательностью Бернулли также такую последовательность независимых случайных величин, где каждое ξ_i принимает значения из конечного множества чисел, скажем, из множества $\{1, 2, \dots, k\}$, где k одно и то же для всех i , с вероятностями

$$\mu_{\xi_i}(\{l\}) = p_l \quad (l = 1, 2, \dots, k), \quad p_l > 0 \text{ и } \sum_{l=1}^k p_l = 1.$$

Частные распределения такой случайной последовательности являются дискретными распределениями, определяемыми по той же формуле, что и в случае $k=2$, где $\alpha_i \in \{1, 2, \dots, k\}$.

Функция распределения суммы первых n членов последовательности независимых случайных величин, согласно II.6.1, является сверткой n функций распределения одномерных частных распределений:

$$F_{\sum_{i=1}^n \xi_i} = \overset{*}{\underset{*}{F}}_{\xi_i}.$$

Рассмотрим распределение суммы случайного числа первых членов случайной последовательности.

Пусть N — случайная величина, независимая от $\xi = (\xi_n)_{n=1}^{\infty}$, принимающая целые неотрицательные значения. По формуле полной вероятности имеем

$$\begin{aligned} F_{\sum_{i=1}^N \xi_i}(x) &= P\left(\sum_{i=1}^N \xi_i < x\right) = \sum_{n=1}^{\infty} P\left(\sum_{i=1}^n \xi_i < x \mid N = n\right) P(N = n) = \\ &= \sum_{n=1}^{\infty} P\left(\sum_{i=1}^n \xi_i < x\right) P(N = n) = \sum_{n=1}^{\infty} \left(\overset{*}{\underset{*}{F}}_{\xi_i}\right)(x) P(N = n). \end{aligned}$$

Математическое ожидание случайной величины $\sum_{i=1}^N \xi_i$ равно

$\sum_{n=1}^{\infty} \sum_{i=1}^n E\xi_i \cdot P(N=n)$, причем эта формула верна и в том случае, когда компоненты случайного вектора ξ зависимы между собой. В частности, если все $E\xi_i$ равны между собой, то

$$E\left(\sum_{i=1}^N \xi_i\right) = \sum_{n=1}^{\infty} E\xi_i \cdot n \cdot P(N=n) = E\xi_1 \cdot EN.$$

Аналогично могут быть подсчитаны дисперсия и другие моменты случайной величины $\sum_{i=1}^N \xi_i$.

б. Марковские последовательности. Это такие случайные последовательности, для которых все частные распределения $\mu_{(\xi_1, \dots, \xi_n)}$ являются марковскими (см. II. 5. 5). Это означает, что для любого $n \geq 3$

$$\mu_{\xi_n}(\cdot | \xi_1, \dots, \xi_{n-1}) = \mu_{\xi_n}(\cdot | \xi_{n-1})$$

с вероятностью единица. При этом

$$\begin{aligned} \mu^{(n)}(S_1 \times \dots \times S_n) &= \int_{S_1 \times \dots \times S_{n-1}} \mu_{\xi_1}(dx_1) \mu_{\xi_2}(dx_2 | \xi_1 = x_1) \times \\ &\times \mu_{\xi_3}(dx_3 | \xi_2 = x_2) \dots \mu_{\xi_{n-1}}(dx_{n-1} | \xi_{n-2} = x_{n-2}) \mu_{\xi_n}(S_n | \xi_{n-1} = x_{n-1}). \end{aligned}$$

Так как

$$\mu_{\xi_{n+1}}(R^1 | \xi_n = x) = 1,$$

то

$$\mu^{(n)}(S_1 \times \dots \times S_k) = \mu^{(n+1)}(S_1 \times \dots \times S_n \times R^1)$$

и, следовательно, частные распределения $\mu^{(n)}$ согласованы между собой. Частным случаем марковской последовательности является последовательность независимых случайных величин.

Марковские последовательности в том случае, если каждое ξ_i может принимать значения из заданного конечного множества, например из множества $\{1, \dots, k\}$, называются *марковскими цепями*. Наиболее проста теория однородных марковских цепей, т. е. таких марковских цепей, для которых условная вероятность

$$\mu_{\xi_n}(\{a\} | \xi_{n-1} = a') \quad (a, a' \in \{1, \dots, k\})$$

не зависит от n :

$$\mu_{\xi_n}(\{a\} | \xi_{n-1} = a') = \mu_{\xi_2}(\{a\} | \xi_1 = a') = p_{a', a}.$$

В этом случае распределение марковской цепи задается двумя параметрами:

вектором начальных вероятностей (p_1, p_2, \dots, p_k) , где $p_i = \mu_{\xi_1}(\{i\}) = P(\xi_1 = i)$,

и матрицей так называемых переходных вероятностей $(p_{ij})_{k \times k}$.

в. Гауссовские последовательности. Так называются случайные последовательности, для которых все частные распределения являются нормальными. Как известно (см. П.4.5 (б)), нормальное распределение порядка n характеризуется вектором математических ожиданий (a_1, \dots, a_n) и ковариационной матрицей Λ порядка $n \times n$. Нетрудно показать, что частные распределения $\mu^{(k)}$ гауссовской последовательности согласованы тогда и только тогда, когда вектор математических ожиданий распределения $\mu^{(k)}$ является частью соответствующего вектора для $\mu^{(k+1)}$ (если $E(\xi_1, \dots, \xi_{k+1}) = (a_1, \dots, a_{k+1})$, то $E(\xi_1, \dots, \xi_k) = (a_1, \dots, a_k)$), а ковариационная матрица распределения $\mu^{(k)}$ является подматрицей ковариационной матрицы распределения $\mu^{(k+1)}$ (если $\Lambda(\xi_1, \dots, \xi_{k+1}) = (\lambda_{ij})_{(k+1) \times (k+1)}$, то $\Lambda(\xi_1, \dots, \xi_k) = (\lambda_{ij})_{k \times k}$).

Таким образом, гауссовская последовательность характеризуется бесконечным вектором математических ожиданий $\mathbf{a} = (a_1, a_2, \dots)$, где $a_i = E\xi_i$, и бесконечной ковариационной матрицей $\Lambda(\xi) = (\lambda_{ij})_{\infty \times \infty}$, где $\lambda_{ij} = \text{cov}(\xi_i, \xi_j)$.

Гауссовская последовательность может быть одновременно марковской последовательностью, в том числе и последовательностью независимых случайных величин. Это добавочное свойство гауссовской последовательности отражается на виде ее ковариационной матрицы.

Гауссовским процессам, частным случаем которых является гауссовская последовательность, посвящена большая литература (Ибрагимов, Розанов, 1970, и др.). Эти процессы и соответствующие поля могут служить моделью различных геологических явлений, которые в этой книге не рассматриваются.

П.8.2. Стационарные случайные последовательности и эргодичность

Случайная последовательность $\xi = (\xi_1, \xi_2, \dots)$ называется стационарной, если ее частные распределения для наборов координат, различающихся только сдвигом, равны между собой:

$$\mu(\xi_1, \dots, \xi_n) = \mu(\xi_{k+1}, \dots, \xi_{k+n}) \quad (k, n = 1, 2, \dots).$$

Стационарные последовательности можно продолжить с сохранением стационарности на все целые числа $Z = \{0, \pm 1, \pm 2, \dots\}$. Для этого на множестве всех последовательностей вида

$$x = (\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots) \quad (x_i \in R^1),$$

которое мы обозначим R^Z , определяется σ -алгебра подмножеств \mathcal{B}^Z — наименьшая σ -алгебра, содержащая все цилиндрические подмножества множества R^Z (множества вида

$$\dots R^1 \times R^1 \times S_{-m} \times \dots \times S_{-1} \times S_0 \times S_1 \times \dots \times S_m \times R^1 \times R^1 \times \dots,$$

где $S_i \in \mathcal{B}^1$). Далее на этой σ -алгебре задаем последовательность конечномерных распределений $(\tilde{\mu}^{(n)})_1^\infty$, где $\tilde{\mu}^{(n)}$ является распределением на множестве векторов $(x_{-n}, \dots, x_{-1}, x_0, x_1, \dots, x_n)$, причем $\tilde{\mu}^{(n)}$ совпадает с частным распределением $\mu_{(\xi_1, \dots, \xi_{2n+1})}$ исходного процесса. По условию стационарности для исходного процесса семейство распределений $(\tilde{\mu}^{(n)})^\infty$ будет согласовано, и по теореме Колмогорова существует одно и только одно вероятностное распределение $\tilde{\mu}$ на множестве R^Z всех бесконечных в обе стороны последовательностей. Чтобы наше построение соответствовало принятой схеме, по которой случайная последовательность — это последовательность случайных величин, заметим, что в качестве Ω можно взять пространство R^Z , в качестве \mathcal{F} — σ -алгебру \mathcal{B}^Z , в качестве P — распределение $\tilde{\mu}$, в качестве ξ_i — функцию π_i , где $\pi_i(x) = x_i$ ($x \in R^Z$).

Одно из самых замечательных свойств стационарных последовательностей — это своеобразный аналог закона больших чисел.

Пусть существует $E\xi_i = a$ для всех i . Тогда из знаменитой теоремы Биркгофа—Хинчина (Гихман, Скороход, 1, 1971, с. 170) следует, что с вероятностью единица существует

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i = h.$$

В общем случае h является случайной величиной. Однако если h постоянна, то она равна a . Стационарные случайные последовательности, обладающие таким свойством, называются эргодическими. Для эргодических последовательностей при любой измеримой функции f от k переменных, для которой существует $Ef(\xi_1, \dots, \xi_k)$, с вероятностью единица

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(\xi_{i+1}, \xi_{i+2}, \dots, \xi_{i+k}) = Ef(\xi_1, \xi_2, \dots, \xi_k).$$

В частности, если

$$f(x) = I_A(x) \quad (A \in \mathcal{B}^1, x \in R^1),$$

т. е. $I_A(x) = 1$, если $x \in A$, и равно 0, если $x \notin A$, то

$$\frac{1}{n} \sum_{i=1}^n f(\xi_i) = \nu_{\{\xi_i \in A\}}.$$

Последнее выражение является аналогом частоты события $\{\xi_i \in A\}$ в серии из n независимых испытаний. При этом

$$\lim_{n \rightarrow \infty} \nu_{\{\xi_i \in A\}} = EI_A(\xi_i) = P(\xi_i \in A).$$

Приведем одно достаточное условие эргодичности стационарной случайной последовательности. Предположим, что существуют

вторые моменты случайных величин ξ_i . При этом элементы ковариационной матрицы

$$\lambda_{ij} = \text{cov}(\xi_i, \xi_j) \quad (i, j \in Z)$$

зависят только от разности индексов i и j . Функция

$$R(n) = \frac{1}{D\xi_1} \text{cov}(\xi_1, \xi_{1+n}) \quad (n=0, 1, 2, \dots)$$

называется *корреляционной функцией стационарной случайной последовательности*. Пример графика этой функции, наблюдавшегося при конкретных геологических исследованиях, приведен на

Рис. II.7. Корреляционная функция (коррелограмма) распределения пористости по разрезу верхнедевонских и нижнекаменноугольных отложений в скв. I у сел. Тепловка Саратовской области (Вистелиус, 1963).

При сохранении в дальнейшем того же характера затухания функции подобная коррелограмма указывала бы на эргодичность стационарного процесса изменения пористости по разрезу.

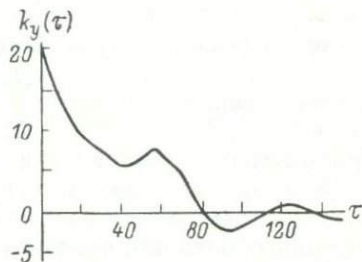


рис. II.7. Докажем, что последовательность эргодическая, если $R(n) \rightarrow 0$ при $n \rightarrow \infty$. Имеем

$$\begin{aligned} D\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) &= E\left(\frac{1}{n} \sum_{i=1}^n (\xi_i - a)\right)^2 = \frac{1}{n^2} \left(\sum_{i=1}^n E(\xi_i - a)^2 + \right. \\ &+ \left. \sum_{i \neq j} E(\xi_i - a)(\xi_j - a) \right) = \frac{1}{n^2} D\xi_i \left(n + 2 \sum_{1 \leq i < j \leq n} R(j-i) \right) = \\ &= \frac{1}{n^2} D\xi_i \left(n + 2 \sum_{i=1}^{n-1} (n-i) R(i) \right) = \\ &= \frac{1}{n^2} D\xi_i \left(n + 2 \sum_{i=1}^m (n-i) R(i) + 2 \sum_{i=m+1}^{n-1} (n-i) R(i) \right). \end{aligned}$$

Так как $|R(i)| \leq 1$ и при любом $\varepsilon > 0$ при всех достаточно больших m $|R(i)| < \varepsilon$, то для таких m и $n > m+1$ будем иметь

$$D\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) \leq \frac{1}{n^2} D\xi_i (n + 2m(n-1) + \varepsilon(n-m-1)(n-m)).$$

Следовательно, при $n \rightarrow \infty$ оцениваемая дисперсия становится и остается меньше любой сколь угодно малой положительной величины, т. е. стремится к нулю. Отсюда по неравенству Чебышева

следует, что $\frac{1}{n} \sum_{i=1}^n \xi_i$ сходится по вероятности к a и, значит, событие

$\left\{ \frac{1}{n} \sum_{i=1}^n \xi_i \rightarrow a \right\}$ имеет вероятность единица (о вероятностях таких событий см. Гнеденко, 1961, с. 211), т. е. последовательность $(\xi_n)_{n \in \mathbb{Z}}$ является эргодической.

Простой пример неэргодической последовательности можно получить следующим образом. Имеются две неслучайные последовательности $x' = (1, 1, 1, \dots)$ и $x'' = (0, 0, 0, \dots)$.

Пусть $A \in \mathcal{F}$ и $0 < P(A) < 1$. Составим третью последовательность $\xi = I_A x' + I_{\bar{A}} x'' = (I_A x'_1 + I_{\bar{A}} x''_1, I_A x'_2 + I_{\bar{A}} x''_2, \dots)$, которая, как нетрудно показать, будет стационарной случайной последовательностью.

Однако $h = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i$ является случайной величиной, принимающей значения 0 и 1, причем $P(h=1) = P(A)$.

Для эргодических последовательностей, так же как и для последовательностей независимых и одинаково распределенных случайных величин, частные распределения любой размерности могут быть в принципе определены по одной «типичной» траектории, причем вероятность «типичной» траектории равна единице. Если стационарная случайная последовательность не является эргодической, то при некоторых условиях, которые практически всегда выполняются, ее можно разложить на эргодические составляющие. При этом пространство Ω разбивается на некоторые измеримые пересекающиеся множества, образующие разбиение \mathfrak{B} так, что

$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i$ равен $E(\xi_i | \mathfrak{B})$ — условному математическому ожиданию ξ_i относительно разбиения \mathfrak{B} . Так что, грубо говоря, если фиксирован элемент разбиения $B \in \mathfrak{B}$, то последовательность является эргодической по отношению к новому вероятностному пространству $(\Omega, \mathcal{F}, P(\cdot | B))$.

В приведенном выше примере разбиению на эргодические составляющие соответствует разбиение $\mathfrak{B} = \{A, \bar{A}\}$. По отношению к любому из условных распределений вероятностей $P(\cdot | A)$ и $P(\cdot | \bar{A})$ эта последовательность является эргодической.

На практике часто возникает необходимость сопоставлять две стационарные последовательности. Их можно представить как стационарную последовательность пар $(\xi_n, \eta_n)_{n=1}^{\infty}$, заданную на одном и том же вероятностном пространстве. Помимо обычных характеристик для такой последовательности рассматривают *функцию взаимной корреляции*, определяемую по формуле

$$R_{12}(n) = \frac{\text{cov}(\xi_1, \eta_{1+n})}{\sqrt{D\xi_1 \cdot D\eta_{1+n}}} \quad (n=0, 1, 2, \dots).$$

Для независимых последовательностей $(\xi_n)_1^\infty$ и $(\eta_n)_1^\infty$ $R_{12}(n) = 0$ для всех n . В общем случае значения $R_{12}(n)$ могут служить мерой их зависимости. Эргодичность такой последовательности определяется аналогично. Для эргодической последовательности пар функцию взаимной корреляции можно оценить путем усреднения «вдоль траектории» по формуле

$$R_{12}(n) \sim \frac{1}{m} \sum_{i=1}^m \frac{(\xi_{i+1} - a)(\eta_{i+n} - b)}{\sigma_1 \sigma_2},$$

где $a = E\xi_1$, $b = E\eta_1$, $\sigma_1^2 = D\xi_1$, $\sigma_2^2 = D\eta_1$, которые также могут быть оценены по одной траектории с точностью тем большей, чем больше m . С анализом эргодических последовательностей можно познакомиться в работе Хеннана (1964). С помощью такого анализа случайных последовательностей можно обнаружить зависимость на первый взгляд совершенно не связанных друг с другом последовательностей. Простая идея взаимной корреляции во многих случаях кажется очень соблазнительной при решении геологических задач. Однако опыт работы по сопоставлению разрезов (метод скользящей корреляции; Вистелиус, Романова, 1962) показал, что могут возникнуть очень большие статистические трудности при оценке значимости $R_{12}(n)$.

II.8.3. Случайные процессы

Пусть T — некоторое множество значений параметра t . Рассмотрим функцию двух аргументов $\xi(\cdot, \cdot)$, определенную на множестве $\Omega \times T$, со значениями в R^1 . При фиксированном $\omega \in \Omega$ $\xi(\omega, \cdot)$ является обычной неслучайной функцией, определенной на T , со значениями в R^1 . При фиксированном $t \in T$ $\xi(\cdot, t)$ является функцией от ω , которую мы предполагаем измеримой.

Таким образом, $\xi(\cdot, t)$ — это случайная величина. Такую функцию двух аргументов $\xi(\cdot, \cdot)$ называют случайной функцией. Ее можно трактовать как семейство $(\xi(\cdot, t))_{t \in T}$ случайных величин. Если $T = Z = \{1, 2, \dots\}$ или $T = Z = \{0, \pm 1, \pm 2, \dots\}$, то в соответствии с II.8.1 случайная функция называется случайной последовательностью; если $T = R_+^1 = [0, \infty)$ или $T = R^1 = [-\infty, \infty]$, то случайная функция называется случайным процессом; если $T = R^n$ ($n \geq 2$), то случайную функцию называют случайным полем. Рассматривают также векторные случайные функции $\xi(\cdot, \cdot)$, для которых при любых $\omega \in \Omega$ и $t \in T$ $\xi(\omega, t) \in R^m$, где $m \geq 2$.

Реализациями случайной функции, т. е. значениями $\xi(\omega, \cdot)$, являются функции $x = x(\cdot)$, где $x(t) \in R^1$ ($t \in T$). Множество всех таких функций x мы обозначим R^T . Если на этом множестве определить σ -алгебру подмножеств \mathcal{F}^T как наименьшую σ -алгебру, содержащую все конечномерные цилиндрические множества, т. е. множества вида

$$\{x \in R^T : \pi_{t_1}(x) \in S_1, \dots, \pi_{t_n}(x) \in S_n\},$$

где $n=1, 2, \dots$; $t_i \in T$, $S_i \in \mathcal{B}^1$ и $\pi_i(x) = x(t)$ (значение функции $x(\cdot)$ в точке t), то случайную функцию можно трактовать как измеримое отображение Ω в R^T . Для такой σ -алгебры при любом $A \in \mathcal{B}^T$ множество $\{\omega : \xi(\omega, \cdot) \in A\}$ принадлежит \mathcal{F} . Следовательно, $\xi = \xi(\cdot, \cdot)$ индуцирует распределение μ_ξ на множестве R^T , которое называется *распределением случайной функции* (случайной последовательности, процесса, поля). Обычный (но не единственный) способ задания распределения μ_ξ , в случае если $T = R^n$ ($n \geq 1$) (или $T \subset R^n$), состоит в конструировании согласованной системы частных конечномерных распределений, которая в соответствии с теоремой Колмогорова определяет одно и только одно распределение на R^T с данными частными распределениями.

Рассмотрим несколько примеров.

а. **Белый шум.** Пусть $T = R^1$ и для любых $n=2, 3, \dots$ и $t_i \in T$ ($-\infty < t_1 < \dots < t_n < \infty$) случайные величины $\xi(\cdot, t_1), \dots, \xi(\cdot, t_n)$ взаимно независимы и одинаково распределены. При этом случайные величины $\xi(\cdot, t)$ называются независимыми в совокупности, а сам случайный процесс называется стационарным белым шумом. Частные распределения белого шума вычисляются по формуле

$$\mu_{\pi(t_1, \dots, t_n)}(\xi)(S_1 \times \dots \times S_n) = \prod_{i=1}^n \mu_{\pi t_i}(\xi)(S_i).$$

Траектории белого шума крайне нерегулярны, поэтому он редко используется в качестве модели реальных явлений.

б. **Стационарный гауссовский процесс.** Так называется случайный процесс на R^1 , все конечномерные частные распределения которого являются нормальными. Можно показать, что эти распределения согласованы в том и только в том случае, если для любых $n=1, 2, \dots$ и $t_i \in R^1$ ($-\infty < t_1 < \dots < t_n < \infty$) вектор математических ожиданий распределения $\mu_{\pi(t_1, \dots, t_n)}(\xi)$ состоит из одинаковых компонент $E\xi(\cdot, t_i) = a$, а члены ковариационной матрицы $\lambda_{ij} = \text{cov}(\xi(\cdot, t_i), \xi(\cdot, t_j))$ зависят лишь от разности параметров t_i и t_j , т. е.

$$\lambda_{ij} = \lambda(t_j - t_i) \quad (t_i < t_j),$$

где $\lambda(\cdot)$ — положительно определенная функция, заданная на $[0, \infty)$,

такая, что для любых n , t_i $\sum_{i,j=1}^n \lambda(|t_i - t_j|) t_i t_j > 0$ (о гауссовских

процессах см. Ибрагимов и Розанов, 1970).

в. **Марковские процессы.** Так же как и марковские последовательности, марковские процессы можно определить как такие случайные процессы, для которых любое конечномерное частное распределение является марковским. Если процесс определен на $[0, \infty)$, то для задания его распределения достаточно знать так называемое начальное распределение $\mu_{\pi_0}(\xi)$ и все условные

распределения $\mu_{\pi_{t_2}(\xi)}(\cdot | \pi_{t_1}(\xi))$ ($0 \leq t_1 < t_2$), которые в силу марковского свойства должны быть согласованы между собой уравнением

$$\mu_{\pi_{t_2}(\xi)}(S | \pi_{t_1}(\xi)) = \int_{R^1} \mu_{\pi_{t_2}(\xi)}(S | \pi_{t_2}(\xi) = x) \cdot \mu_{\pi_{t_2}(\xi)}(dx | \pi_{t_1}(\xi)),$$

где $t_1 < t_2 < t_3$. Это уравнение называется *уравнением Колмогорова—Чепмена* или обобщенным уравнением Маркова. Марковский процесс называется *однородным*, если условная вероятность $\mu_{\pi_{t_2}(\xi)}(S | \pi_{t_1}(\xi) = x)$ зависит лишь от разности $t_2 - t_1$.

Если обозначить в этом случае

$$\mu_{\pi_{t+h}(\xi)}(S | \pi_t(\xi) = x) = P_h(S | x),$$

то уравнение Колмогорова—Чепмена переписется в виде

$$P_{t_1+t_2}(S | x) = \int_{R^1} P_{t_2}(S | x_1) P_{t_1}(dx_1 | x) (t_1, t_2 > 0, S \in \mathcal{S}^1, x \in R^1).$$

Конечномерные частные распределения марковского процесса находятся так же, как и у марковской последовательности. В частности, для однородного марковского процесса

$$\begin{aligned} & \mu_{\pi_{(0, t_1, \dots, t_n)}(\xi)}(S_0 \times S_1 \times \dots \times S_n) = \\ = & \int_{S_0 \times S_1 \times \dots \times S_n} \mu_{\pi_0(\xi)}(dx_0) P_{t_1}(dx_1 | x_0) P_{t_2-t_1}(dx_2 | x_1) \dots P_{t_n-t_{n-1}}(dx_n | x_{n-1}). \end{aligned}$$

Определение стационарности случайного процесса дословно повторяет подобное определение для случайной последовательности. Однородный марковский процесс стационарен, если его одномерные частные распределения одинаковы для всех $t \geq 0$. Следовательно, его начальное распределение $\mu_{\pi_0(\xi)}$ должно обладать свойством

$$\mu_{\pi_0(\xi)}(S) = \int_{R^1} \mu_{\pi_0(\xi)}(dx) P_t(S | x)$$

при любом $t > 0$ и $S \in \mathcal{S}^1$.

Изложение общей теории марковских процессов читатель найдет в книгах Дынкина (1959, 1963), а также почти в любом курсе теории случайных процессов.

Более подробно мы изучим один частный случай однородных марковских процессов — процесс Пуассона, который интересен и сам по себе, так как обладает рядом замечательных свойств. Кажется, что он может служить моделью геологических явлений. Многомерные обобщения пуассоновского процесса служат обоснованием некоторых построений теории геометрических вероятностей,

которая также находит применение в геологических исследованиях.

г. Процесс Пуассона. Процессом Пуассона называется однородный марковский процесс ξ на полупрямой $T = [0, \infty)$, траектории которого могут принимать только неотрицательные целые значения и условные вероятности которого $P_t(S|x)$, называемые также переходными функциями, имеют вид

$$P_t(S|n) = \sum_{k \in S} P_t(k|n),$$

где

$$P_t(k|n) = \begin{cases} 0, & \text{если } k < n, \\ e^{-\lambda t} \frac{(\lambda t)^{k-n}}{(k-n)!}, & \text{если } k \geq n, \end{cases}$$

$\lambda > 0$, $k, n = 0, 2, 1, \dots$. Это означает, что распределение приращения траектории на интервале длины t является распределением Пуассона с параметром λt . Следовательно, траектории процесса с вероятностью единица не убывают. Так как переходные вероятности $P_t(k|n)$ зависят лишь от разности $k-n$, то приращения траектории на интервале $(a, a+t)$ и значение траектории в момент a являются независимыми случайными величинами.

Таким образом, процесс Пуассона относится к классу процессов с независимыми приращениями, для которых приращения не зависят от предшествующей траектории до момента a . Обычно исследуют процесс Пуассона при начальном распределении $\mu_{\pi_0(\xi)}(\{0\}) = 1$. Тогда случайная величина $\xi(t)$ совпадает со случайным числом скачков траектории процесса на интервале $(0, t]$. Чтобы показать это, достаточно проверить, что с вероятностью единица траектория не имеет скачков с величиной больше единицы.

Пусть $A(a, b]$ — событие, состоящее в том, что на интервале $(a, b]$ отсутствуют скачки с величиной больше единицы. Тогда

$$A(0, t] = \prod_{k=0}^{n-1} A\left(t \frac{k}{n}, t \frac{k+1}{n}\right].$$

Из свойств независимости приращений процесса ξ следует, что события $A\left(t \frac{k}{n}, t \frac{k+1}{n}\right]$ ($k = 0, 1, \dots, n-1$) взаимно независимы.

Следовательно,

$$P(A(0, t]) = \prod_{k=0}^{n-1} P\left(A\left(t \frac{k}{n}, t \frac{k+1}{n}\right)\right),$$

но

$$\begin{aligned} P\left(A\left(t \frac{k}{n}, t \frac{k+1}{n}\right)\right) &= P\left(A\left(0, \frac{t}{n}\right)\right) \geq \\ &\geq P\left(\xi\left(\cdot, \frac{t}{n}\right) \leq 1\right) = e^{-\lambda \frac{t}{n}} + e^{-\lambda \frac{t}{n}} \lambda \frac{t}{n} \end{aligned}$$

$$P(A(0, t]) \geq e^{-\lambda t} \left(1 + \lambda \frac{t}{n}\right)^n.$$

Так как $(1 + \lambda t/n)^n \rightarrow e^{\lambda t}$ при $n \rightarrow \infty$, то $1 \geq P(A(0, t]) \geq 1$. Следовательно, с вероятностью единица значение реализации в момент t совпадает с числом разрывов реализации на интервале $(0, t]$. Доказанное свойство называют также свойством ординарности процесса.

Математическое свойство ординарности выражается следующим образом:

$$\frac{1}{t} P_t((n+2, n+3, \dots) | n) \rightarrow 0 \text{ при } t \rightarrow 0.$$

Интересно отметить, что три свойства целочисленного случайного процесса — независимость приращений, однородность и ординарность — полностью характеризуют класс процессов Пуассона (Гнеденко, 1961, с. 295).

Обозначим τ_1, τ_2, \dots — моменты скачков процесса Пуассона. Это положительные случайные величины, причем $\tau_1 < \tau_2 < \dots$. Распределение случайной величины τ_1 определяется просто:

$$[P(\tau_1 > t) = P(\xi(\cdot, t) = 0) = e^{-\lambda t}$$

и, следовательно, τ_1 имеет экспоненциальное распределение с параметром λ . Можно доказать, что для пуассоновского процесса любая разность $\tau_n - \tau_{n-1}$ независима от траектории до момента τ_n и также имеет экспоненциальное распределение, т. е. после каждого скачка приращения процесса ведут себя точно так же, как если бы он начинался с момента нуля. В этом проявляется так называемое строго марковское свойство пуассоновского процесса (Дынкин, 1963, с. 142). Траектория пуассоновского процесса, очевидно, будет задана, если будут указаны моменты скачков.

Таким образом, пуассоновский процесс может быть интерпретирован как последовательность случайных моментов времени τ_1, τ_2, \dots , где $0 < \tau_1 < \tau_2 < \dots$, причем случайные величины $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2$ взаимно независимы и имеют одно и то же экспоненциальное распределение. Приращение процесса Пуассона на интервале $(a, b]$, очевидно, равно числу моментов τ_i , попадающих в этот интервал. Из свойства независимости приращений процесса Пуассона следует, что распределение числа моментов τ_i , попадающих в любое объединение конечного числа непересекающихся интервалов, является распределением Пуассона с параметром, равным λ , умноженным на сумму длин этих интервалов. То же будет для любого борелевского множества, принадлежащего $[0, \infty)$. Число попадающих в него точек τ_i имеет распределение Пуассона с параметром, пропорциональным «длине» этого множества. Кроме того, числа точек τ_i , попадающие в два непересекающихся борелевских множества, являются независимыми случайными величинами. Это свойство пуассоновского процесса на полупрямой лежит

в основе всех его обобщений. Во-первых, с выполнением этого свойства можно продолжить процесс Пуассона на всю прямую — множество R^1 . Во-вторых, можно определить пуассоновское поле (называемое также многомерным пуассоновским процессом) как такое случайное распределение меченых точек в пространстве R^n ($n \geq 2$), что числа меченых точек $N(\cdot, A)$, попадающих в борелевское множество A , являются случайными величинами, имеющими пуассоновское распределение с параметром, пропорциональным «объему» множества A , причем для непересекающихся борелевских множеств A_1 и A_2 случайные величины $N(\cdot, A_1)$ и $N(\cdot, A_2)$ независимы.

д. Винеровский процесс. Другим случайным процессом, который, так же как и пуассоновский процесс, может найти применение при построении математических моделей в геологии, является винеровский процесс. Рассмотрим только одномерный случай.

Винеровский процесс можно определить как однородный марковский процесс ξ на полупрямой $T=[0, \infty)$, условное распределение которого $P_t(\cdot|x)$ при фиксированном $x \in R^1$ и $t > 0$ совпадает с нормальным распределением с математическим ожиданием x и с дисперсией t . Отсюда следует, чторащения винеровского процесса на любом интервале $(a, b]$ независимы от предшествующей траектории до момента a . Обычно исследуют винеровский процесс при начальном распределении $\mu_{\pi_0(\xi)}(\{0\})=1$. При этом условии винеровский процесс можно отнести к классу гауссовских процессов, для которых, согласно определению, все конечномерные частные распределения нормальны. Но он не является стационарным гауссовским процессом. Дисперсия распределения $\mu_{\pi_t(\xi)}$ пропорциональна времени t . Вероятность как бы «растекается» из начальной точки $x=0$ по всей прямой.

Траектории винеровского процесса непрерывны, поэтому он может служить моделью движения материальных тел. Именно для этой цели — для описания броуновского движения — такой процесс был впервые определен и исследован Эйнштейном и Смолуховским (1936). Однако траектории винеровского процесса очень нерегулярны. Например, с вероятностью единица они не имеют производных ни в одной точке множества T . На практике чаще используется преобразованный винеровский процесс $\xi_1(t) = at + b\xi(t)$, где $a \in R^1$ и $b > 0$. Параметр a называется сносом, а b — локальной дисперсией.

Рассматривают также аналоги винеровского процесса с переменным сносом и локальной дисперсией, называемые диффузионными процессами. Как следует из названия, эти процессы хорошо описывают диффузию частиц одного вещества внутри другого.

Подробнее о винеровских и диффузионных процессах см. работу Ито и Маккина (1968), а также любой курс теории случайных процессов.

§ 8.4. Точечные процессы

При изучении процесса Пуассона мы встречаемся со случайным распределением точек на прямой или в пространстве R^n ($n \geq 2$). Исследованием таких случайных распределений занимается теория случайных *точечных процессов* и случайных *точечных полей* (последние называются также случайными многомерными точечными процессами, причем слова «случайный» и «многомерный» иногда опускаются). Наиболее простое определение точечных процессов получится, если интерпретировать их как случайные целочисленные меры.

Пусть в пространстве R^n некоторым неслучайным образом выбраны точки так, что в любом параллелепипеде $[x_1, y_1] \times [x_2, y_2] \times \dots \times [x_n, y_n]$ содержится только конечное число выбранных точек. Рассмотрим систему \mathcal{B}_0^n всех ограниченных борелевских множеств пространства R^n . Ограниченным называется множество, включенное в некоторый параллелепипед. Для любого $A \in \mathcal{B}_0^n$ определено и конечно число $\nu(A)$ выбранных точек, принадлежащих A . Очевидно, если $A_1 \cap A_2 = \emptyset$, то $\nu(A_1 \cup A_2) = \nu(A_1) + \nu(A_2)$, и более того, если $(A_n)_{n=1}^{\infty}$ — система попарно непересекающихся множеств, принадлежащих \mathcal{B}_0^n , то

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \nu(A_n).$$

Мы видим, что функция множеств $\nu(\cdot)$ обладает всеми свойствами меры, с которыми мы имели дело до сих пор, за исключением, может быть, одного, а именно: мера всего пространства R^n — число $\nu(R^n)$ — не обязано быть конечным. Говорят, что ν является *целочисленной σ -конечной мерой*. Очевидно, эта мера принимает только целые значения на ограниченных множествах. Точечный процесс (поле) — это случайная целочисленная σ -конечная мера, т. е. такая функция от двух аргументов $N = N(\cdot, \cdot)$, заданная на декартовом произведении множеств $\Omega \times \mathcal{B}_0^n$, что для любого $\omega \in \Omega$ $N(\omega, \cdot)$ — целочисленная σ -конечная мера на \mathcal{B}_0^n , а при любом $A \in \mathcal{B}_0^n$ $N(\cdot, A)$ — случайная величина, принимающая неотрицательные целые значения.

Пусть \mathcal{M} — множество всех целочисленных σ -конечных мер (неслучайных) на \mathcal{B}_0^n . Если определить σ -алгебру \mathcal{M} подмножеств множества \mathcal{M} как наименьшую σ -алгебру, содержащую все «цилиндрические» множества, т. е. множества вида

$$\bigcap_{i=1}^m \{\nu \in \mathcal{M} : \nu(A_i) = k_i\},$$

где $m = 1, 2, \dots$, $A_i \in \mathcal{B}_0^n$, $k_i \in \{0, 1, 2, \dots\}$, то нетрудно заметить, что N является измеримым отображением Ω в \mathcal{M} относительно σ -алгебр \mathcal{F} и \mathcal{M} .

Следовательно, N индуцирует распределение μ_N на множестве \mathcal{N} :

$$\mu_N(S) = P(N \in S),$$

где $S \in \mathcal{M}$.

Конечномерными частными распределениями распределения μ_N называются совместные распределения случайных величин $N(\cdot, A_1), \dots, N(\cdot, A_m)$, где m — любое целое положительное число и $A_i \in \mathcal{F}_0^n$. Для случайных целочисленных σ -конечных мер — точечных процессов и полей, — также как и для случайных функций, существуют условия согласования конечномерных частных распределений, необходимые и достаточные для существования некоторого распределения μ_N на множестве реализаций точечных процессов (Kallenberg, 1974).

Точечный процесс (поле) называется стационарным, если все его конечномерные частные распределения не изменяются при одновременном сдвиге всех параметров — множеств A_1, \dots, A_m — на один и тот же вектор $x \in R^n$:

$$\mu(N(\cdot, A_1), \dots, N(\cdot, A_m)) = \mu(N(\cdot, A_1+x), \dots, N(\cdot, A_m+x)),$$

где $A+x = \{x' + x : x' \in A\}$.

Точечный процесс (поле) называется изотропным, если его конечномерные частные распределения не изменяются при замене параметров A_1, \dots, A_m на их образы DA_1, \dots, DA_m при любом вращении D пространства R^n . Стационарные и изотропные точечные процессы (поля) называются инвариантными относительно евклидовых движений. В этом случае «объем» преобразуемых областей не меняется.

Важнейшей характеристикой точечного процесса (поля) является его математическое ожидание, т. е. мера $M(\cdot)$, заданная на \mathcal{F}_0^n соотношением $M(A) = EN(\cdot, A)$ ($A \in \mathcal{F}_0^n$). Мера $M(\cdot)$ называют мерой интенсивности точечного процесса (поля). Мера интенсивности называется абсолютно непрерывной, если существует такая функция $m(\cdot)$, что для любого $A \in \mathcal{F}_0^n$ $M(A) = \int_A m(x) dx$.

Функция $m(\cdot)$ называется *интенсивностью* точечного процесса (поля).

Другой важной характеристикой точечного процесса является аналог ковариационной функции — моментная функция множеств второго порядка $V(\cdot, \cdot)$, определяемая равенством

$$V(A_1, A_2) = \text{cov}(N(\cdot, A_1), N(\cdot, A_2)).$$

Рассмотрим некоторые частные случаи.

а. Пуассоновское поле. Пуассоновским (однородным) полем называется случайная σ -конечная целочисленная мера N , значения которой на непересекающихся множествах являются независимыми случайными величинами, причем случайная величина $N(\cdot, A)$ ($A \in \mathcal{F}_0^n$) имеет распределение Пуассона с парамет-

ром, пропорциональным объему $|A|$ множества A . В результате мы имеем

$$P(N(\cdot, A) = k) = e^{-\lambda|A|} \frac{(\lambda|A|)^k}{k!}.$$

Следовательно,

$$EN(\cdot, A) = \lambda|A| \text{ (см. 2.4.1 (б))},$$

т. е. для пуассоновского поля определена интенсивность, и эта интенсивность равна λ . Отсюда же следует, что пуассоновское поле инвариантно относительно всех евклидовых движений пространства R^n . Из определения следует, что

$$V(A_1, A_2) = 0,$$

если $A_1 \cap A_2 = \emptyset$ и $V(A, A) = \lambda|A|$. Прочие значения функции множеств $V(\cdot, \cdot)$ легко могут быть подсчитаны с использованием свойства аддитивности $N(\cdot, A_1 \cup A_2) = N(\cdot, A_1) + N(\cdot, A_2)$ ($A_1 \cap A_2 = \emptyset$) и независимости $N(\cdot, A_1)$ и $N(\cdot, A_2)$.

Пуассоновское поле на пространстве $R_+^1 = [0, \infty)$ является процессом Пуассона, как это следует из интерпретации процесса Пуассона в виде последовательности $\{\tau_n\}_1^\infty$ — случайных моментов скачков траектории процесса. Для двухмерного и больших размерностей пуассоновского поля трудно указать аналогичное представление в виде случайной функции. Однако, как и для любого случайного точечного процесса (поля), реализацию пуассоновского процесса (поля) можно задать, указав отмеченные точки (или, как говорят, «атомы» меры). Для пуассоновского процесса условное распределение отмеченных точек в данном параллелепипеде при условии, что число этих точек фиксировано и равно k , совпадает с распределением, которое получается, если каждая из k точек выбрана в этом параллелепипеде случайно и независимо от других в соответствии с равномерным распределением. Мера интенсивности однородного пуассоновского поля пропорциональна объему или, как говорят, лебеговой мере в n -мерном пространстве. Для любого параллелепипеда $A = [a_1, b_1] \times \dots \times [a_n, b_n]$ лебегова

мера $|A|$ равна $\prod_{i=1}^n (b_i - a_i)$. Лебегова мера является примером бесконечной на всем пространстве, но σ -конечной меры. Пусть $\Lambda(\cdot)$ — любая другая σ -конечная мера на пространстве R^n . С этой мерой можно связать точечный процесс, для которого целочисленная случайная σ -конечная мера N имеет независимые значения на пересекающихся множествах, причем случайная величина $N(\cdot, A)$ ($A \in \mathcal{B}_n$) имеет распределение Пуассона с параметром $\Lambda(A)$. Такое точечное поле также называется пуассоновским (неоднородным, если мера $\Lambda(\cdot)$ не пропорциональна мере Лебега). Очевидно, мера Λ является мерой интенсивности пуассоновского поля (процесса).

б. Процесс восстановления. Для процесса Пуассона на полупрямой длины интервалов между точками $0, \tau_1,$

τ_2, \dots являются независимыми, случайными величинами с одним и тем же экспоненциальным распределением. Процессы восстановления отличаются от процессов Пуассона лишь тем, что для них допускается любое, а не только экспоненциальное распределение длин интервалов между скачками. Простейшее преобразование процесса Пуассона — последовательность четных точек $(\tau_{2n})_1^\infty$ — является процессом восстановления с функцией распределения $F_2 = F * F$, где $F(x) = 1 - e^{-\lambda x}$ ($x > 0$). Нетрудно выписать формулу для распределений значений на интервалах случайной меры, связанной с процессом восстановления. В данном случае — это распределение числа точек τ_i , попадающих в некоторый интервал, скажем, в интервал (a, b) , которое определяется по формуле

$$P\{N(\cdot, \underline{x}(a, b)) \geq k\} = \sum_{n=0}^{\infty} P(\{N(\cdot, (0, a)) = n\} \cap \{N(\cdot, (0, b)) \geq n + k\}) = \\ = \sum_{n=0}^{\infty} \int_0^a F_n(dt_1) F_1(dt_2) F_{k-1}(b - t_1 - t_2)$$

(интеграл берется по области $0 < t_1 \leq a$, $a < t_1 + t_2 < b$), где $k \geq 1$, F_1 — функция распределения длины интервала между точками τ_1 процесса восстановления, $F_n = F_1^{*n}$ — n -кратная свертка функций распределения F_1 , F_0 — вырожденная функция распределения, для которой $F(x) = 0$ при $x \leq 0$ и $F(x) = 1$ при $x > 0$. Процесс восстановления, для которого τ_1 имеет то же распределение, что и длина каждого из последующих интервалов, в общем случае не является стационарным точечным процессом. Чтобы сделать его стационарным точечным процессом, нужно изменить распределение длины начального интервала. При не слишком ограничительных условиях это стационарное начальное распределение равно пределу при $t \rightarrow \infty$ распределений случайных величин $\tau(t) - t$, где $\tau(t)$ — ближайшая справа к t точка τ_i . Очевидно,

$$P(\tau(t) - t \leq x) = P(N(\cdot, \underline{x}(t, t+x)) \geq 1).$$

Свое название процессы восстановления получили по первоначальному применению их в технике, где точки τ_i соответствовали моментам замены вышедшей из строя детали новой, причем время безотказной работы новой детали — случайная величина с функцией распределения F_1 .

Пример II.5. При исследовании соотношения между периодами активности вулкана ξ_1 и длительностью пауз между этими периодами ξ_2 было установлено, что случайные величины ξ_1 и ξ_2 независимы (Wickman, El-Hinnawi, 1963). Каждая из этих величин имеет экспоненциальное распределение с параметрами λ_1 и λ_2 соответственно.

Таким образом, последовательность моментов начал периодов активности образует процесс восстановления с функцией распределения длин интервалов $F = F_1 * F_2$, где $F_1(x) = 1 - e^{-\lambda_1 x}$, $F_2(x) = 1 - e^{-\lambda_2 x}$.

Можно показать, что стационарная функция распределения F_0 начального интервала в этом процессе восстановления представляет собой суперпозицию функций распределения F_1 и F_2 вида

$$F_0 = \frac{\lambda_2}{\lambda_1 + \lambda_2} (F_1 * F_2) + \frac{\lambda_1}{\lambda_1 + \lambda_2} F_2,$$

где $\frac{\lambda_2}{\lambda_1 + \lambda_2}$ — стационарная вероятность активности вулкана в момент t , которая от t не зависит.

Различные аспекты теории процессов восстановления исследуются в книгах Кокса и Смита (1967), Феллера (2, 1967) и др.

Иногда различают еще один тип точечных процессов — точечные процессы с маркированными точками. В этом случае с каждым моментом времени τ , сопоставляется марка (или метка) x_i — случайная точка, выбранная из некоторого множества X . Если $X = R^n$, то маркированный точечный процесс, очевидно, является частным случаем точечного процесса в R^{n+1} . Обычно X — конечное множество. Тогда маркированный точечный процесс можно представить как систему $(N_x)_{x \in X}$ точечных процессов и каждый из них будет состоять из точек с метками только одного вида.

Пример II.6. О разносе обломочного материала в седиментационном бассейне

Как было отмечено выше, различные виды случайных процессов могут служить моделью геологических явлений. Модель процесса накопления обломочного материала, в которой используются маркированные точечные процессы, диффузионные процессы, остановленные в случайные моменты времени, и случайные меры, исследуется в ряде работ (Вистелиус и др., 1976; Харламов, 1978).

Рассмотрим процесс концентрации аксессуарного минерала A в терригенных отложениях — в слое конгломерата, мощном песчаном пласте и т. п.

Пусть в некоторой точке в пределах области питания имеется источник интересующего нас вещества A , из которого оно поступает в бассейн осадконакопления и фиксируется в осадке вместе с остальным обломочным материалом. При этом предполагается, что накопление материала компенсирует прогибание дна бассейна. Нас будет интересовать содержание вещества A по разрезу и по площади бассейна. Типичная геологическая ситуация может быть заимствована, скажем, из описания исследования спессартина в осадках апт-сеноманского возраста на Юго-Востоке СССР (Вистелиус и др., 1976).

Рассмотрим следующий случайный механизм образования меры вещества A . Источник вещества характеризуется случайной последовательностью

$$(t_i, z_i)_1^\infty,$$

где z_i — масса i -той частицы ($z_i \geq 0$), а t_i — момент отрыва частицы от источника. Для определенности мы отсчитываем время от настоящего в прошлое, т. е. t_i равно «возрасту» частицы — разности между настоящим временем и моментом отрыва частицы от источника. Большим номерам соответствует больший возраст частиц:

$$0 \leq t_1 < t_2 < t_3 < \dots$$

Каждая частица блуждает в окружающем пространстве в течение случайного отрезка времени τ_i ($\tau_i \geq 0$). Это блуждание представляет собой случайный процесс с фазовым пространством R^2 (точнее — блуждание вдоль земной поверхности), со случайными моментами начала и остановки. Предполагается, что мощность накопленного осадка пропорциональна времени его накопления. В момент остановки блуждания частица вещества A остается на уровне осадка, соответствующего данному моменту.

Таким образом, каждая частица данного вещества, распределенного в исследуемом осадке, характеризуется пятеркой величин:

$$(t_i, z_i, \tau_i, x_i, y_i),$$

где (x_i, y_i) — точка на плоскости, соответствующая i -той частице в момент ее остановки; если время измеряется мощностью накопления осадка, то три координаты i -той частицы, погребенной в осадочной толще, равны $(x_i, y_i, t_i - \tau_i)$ (условие $t_i < \tau_i$ означает, что i -тая частица не содержится в осадочной толще, она еще продолжает блуждать).

Случайная мера \mathfrak{M} накопленного вещества A в слое $R^2 \times [0, \infty)$ определяется равенством

$$\mathfrak{M}(S) = \sum_{i=1}^{\infty} z_i I(S | x_i, y_i, t_i - \tau_i),$$

где $S \subset R^2 \times [0, \infty)$ — измеримое множество.

Первая характеристика случайной меры \mathfrak{M} — это мера интенсивности

$$M(S) = E\mathfrak{M}(S),$$

при этом

$$M(S) = \sum_{i=1}^{\infty} E(z_i I(S | x_i, y_i, t_i - \tau_i)).$$

Согласно свойству условного математического ожидания (Феллер, 1967, а также II. 5. 5),

$$E(z_i I(S | x_i, y_i, t_i - \tau_i)) = E(z_i E(I(S | x_i, y_i, t_i - \tau_i) | t_i, z_i)) = \\ = E(z_i P((x_i, y_i, t_i - \tau_i) \in S | t_i, z_i)),$$

откуда

$$M(S) = E \sum_{i=1}^{\infty} z_i P((x_i, y_i, t_i - \tau_i) \in S | t_i, z_i).$$

Для любой измеримой функции f

$$\sum_{i=1}^{\infty} z_i f(t_i, z_i) = \int_{R_+^2} z f(t, z) \nu(dt, dz),$$

где $\nu(B)$ — число таких i , что $(t_i, z_i) \in B$; ν — случайная целочисленная мера на R_+^2 ($R_+^2 = [0, \infty) \times [0, \infty)$). Эта мера характеризует продуктивность источника вещества. Ее задание эквивалентно заданию случайной последовательности $(t_i, z_i)_{i=1}^{\infty}$. Другая мера, также характеризующая случайную последовательность, это мера μ :

$$\mu(dt, dz) = z \cdot \nu(dt, dz),$$

при этом

$$\sum_{i=1}^{\infty} z_i f(t_i, z_i) = \int_{R_+^2} f(t, z) \mu(dt, dz).$$

Пусть $m(B) = E\mu(B)$ — математическое ожидание массы вещества, испускаемого источником в области $B \subset R_+^2$. Тогда

$$M(S) = \int_{R_+^2} p(S|t, z) m(dt, dz), \quad (*)$$

где

$$p(S|t, z) = P((x_i, y_i, t_i - \tau_i) \in S | z_i = z, t_i = t).$$

При этом мы предполагаем, что последняя условная вероятность непосредственно от i не зависит. Следовательно, для вычисления меры M достаточно знать меру интенсивности источника (меру m) и условное распределение тройки (x_i, y_i, τ_i) при фиксированной паре (t_i, z_i) . Заметим, что для вывода формулы никаких предположений о стохастической независимости не делается.

Пусть

$$D(S_1, S_2) = E\mathfrak{M}(S_1) \mathfrak{M}(S_2),$$

где $S_1, S_2 \subset R^2 \times [0, \infty)$ и имеет место следующее условие условной стохастической независимости различных блуждающих частиц:

$$P((x_i, y_i, t_i - \tau_i) \in S_1, (x_j, y_j, t_j - \tau_j) \in S_2 | t_i = t, z_i = z, t_j = t', z_j = z') = p(S_1|t, z) p(S_2|t', z').$$

Тогда, как нетрудно показать,

$$D(S_1, S_2) = M(S_1) M(S_2) + \int_{R_+^2} (p(S_1 \cap S_2|t, z) - p(S_1|t, z) p(S_2|t, z)) z m(dt, dz).$$

Наибольший интерес представляет мера интенсивности случайной меры \mathfrak{M} , которую легко вычислить при некоторых естественных предположениях о характеристиках источника и блуждания частиц. Пусть, например, все частицы обладают одинаковой массой, интенсивность источника постоянна во времени, каждая

частица блуждает относительно источника по закону двумерного диффузионного процесса с параметрами $\mathbf{a}=(a_1, a_2)$ (локальный снос) и $\mathbf{b}=\begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix}$ (локальная ковариационная матрица; Феллер, 1967, с. 407); момент остановки τ не зависит от предшествующей траектории блуждания и распределен экспоненциально с плотностью $\lambda e^{-\lambda t}$.

Тогда интенсивность меры $M'(x, y, s)$ случайной меры \mathfrak{M} равна $M_1(x, y)$, т. е. постоянна по глубине, а функция M_1 подчиняется дифференциальному уравнению

$$\lambda M_1 - \delta = -\frac{\partial}{\partial x}(a_1 M_1) - \frac{\partial}{\partial y}(a_2 M_1) + \\ + \frac{1}{2} \frac{\partial^2}{\partial x^2}(b_{11} M_1) + \frac{\partial^2}{\partial x \partial y}(b_{12} M_1) + \frac{1}{2} \frac{\partial^2}{\partial y^2}(b_{22} M_1) \quad \S$$

где δ — двумерная δ -функция (для любой непрерывной и ограниченной функции $f \int_{R^2} f(x, y) \delta(x, y) dx dy = f(0, 0)$).

Это уравнение имеет единственное ограниченное решение в классе обобщенных функций, если функции $a_i(x, y)$, $b_{ij}(x, y)$ кусочно-непрерывны и ограничены (Гельфанд, Шилов, 1958). В ряде конкретных случаев решение можно найти с наперед заданной точностью с помощью ЭВМ, используя тот или иной стандартный вычислительный алгоритм. При этом все особенности распределения обломочного материала относительно источника полностью определяются параметрами диффузионного блуждания \mathbf{a} и \mathbf{b} .

Даже эта явно идеализированная модель позволяет выявить некоторые особенности накопления материала. Так, например, в рамках этой модели могут образовываться локальные максимумы концентрации вещества в точках разрыва или быстрого изменения параметров. В некоторых случаях эти локальные максимумы могут превышать локальный максимум, который обычно образуется на месте источника. Это обстоятельство заставляет относиться с осторожностью к интерпретации реально обнаруживаемых максимумов как истинных положений источников вещества. Реальность построения иллюстрируется рис. II. 8 (см. вкл.).

II.8.5. Геометрические вероятности

Задачи теории вероятности, в которых в качестве пространства элементарных событий можно взять ограниченную часть евклидова пространства с равномерным распределением вероятностей на этой части, принято называть задачами на геометрическую вероятность. Для определения вероятностей различных событий

в этом случае требуется измерять длины, площади или объемы различных фигур, расположенных в евклидовом пространстве. К задачам на геометрическую вероятность относят также задачи, в которых вероятности различных событий связаны с распределением пуассоновских процессов и полей, множество реализаций которых можно принять за исходное пространство элементарных событий, а распределение пуассоновского процесса (поля) — за исходное вероятностное распределение. Основная трудность при решении задач на геометрическую вероятность состоит в нахождении граничных областей, объем которых пропорционален искомой вероятности, а также в переводе исходных терминов задачи на язык равномерных распределений или пуассоновских полей.

Рассмотрим задачу Бюффона. На плоскость, расчерченную сетью параллельных прямых, проходящих на одинаковом расстоянии одна от другой, равном единице, случайным образом бросается игла длиной L . Найти вероятность того, что игла пересечет хотя бы одну прямую. В условии задачи имеется неопределенность, типичная для задач на геометрическую вероятность. Фиксировав одну из прямых, мы видим, что задача определена в том и только в том случае, когда задано распределение расстояния центра иглы от этой прямой и угла между этой прямой и направлением иглы. Среди всего множества возможных распределений центра иглы и угла ориентации иглы при отсутствии уточнений в условии задачи принято брать распределение, *инвариантное относительно евклидовых движений плоскости*. Одно из таких инвариантных распределений — равномерное распределение угла на интервале $(0, 2\pi]$ и равномерное распределение центра иглы по всему пространству R^2 при независимости угла и положения центра. Это распределение не является вероятностным, так как объем всего пространства равен бесконечности, и поэтому интеграл от постоянной положительной плотности по всему пространству равен бесконечности.

Можно показать, что это единственное инвариантное распределение в данной задаче, т. е. вероятностных инвариантных распределений не существует. В то же время интуиция подсказывает нам, что при таком инвариантном распределении центр иглы должен быть распределен равномерно между двумя параллельными прямыми. Для придания инвариантному решению вероятностного смысла используют или условные вероятности, или пуассоновские поля игл.

В первом случае рассматривают последовательность равномерных распределений центра иглы на частях пространства R^2 , стремящуюся к «равномерному» распределению на всей плоскости. При каждом распределении из этой последовательности определяется условное распределение расстояния центра иглы от фиксированной прямой при условии, что это расстояние не более $1/2$. Предел условных распределений, который в рассматриваемом случае будет равномерным распределением на интервале $[-1/2, +1/2]$, принимается в качестве обоснованного предположения о рас-

пределении центра иглы при отсутствии других указаний о характере случайности. Равномерное распределение угла (на интервале $(0, 2\pi)$) и независимость угла от положения центра не нуждается в подобном обосновании.

Во втором случае рассматривается пуассоновское поле игл. Положение иглы на плоскости определяется двумя параметрами: точкой $x \in R^2$ (центр иглы) и углом $\varphi \in (0, 2\pi]$ (направление конца иглы). В пространстве $R^2 \times (0, 2\pi]$ строится пуассоновское поле с некоторой интенсивностью λ . Построенное пуассоновское поле, которое интерпретируется как пуассоновское поле игл, инвариантно относительно всех евклидовых движений пространства R^2 . Искомая вероятность теперь интерпретируется как математическое ожидание отношения числа игл, пересекающих прямые, к общему числу игл на достаточно большой площади. Если ввести усредненную по всему пространству интенсивность λ' точечного поля игл, пересекающих прямые, то искомая вероятность будет равна λ'/λ . Пользуясь периодичностью множества A параметров, соответствующих иглам, пересекающим прямые, это отношение можно найти, исследуя часть объема множества A в пределах одного периода. И так как множество $A \subset R^2 \times (0, 2\pi]$ имеет вид бесконечного цилиндра, вытянутого вдоль оси, соответствующей направлению параллельных прямых, то достаточно исследовать перпендикулярное сечение этого цилиндра в пределах одного периода. Так снова косвенным образом вводится равномерное распределение центра иглы на отрезке, соединяющем две соседние прямые, и равномерное распределение угла на интервале $(0, 2\pi]$ при независимости положения центра и угла.

В задаче Бюффона после ее формализации полагают, что точка (x, φ) , представляющая иглу, распределена равномерно на прямоугольнике $\left[-\frac{1}{2}, \frac{1}{2}\right] \times (0, 2\pi]$. Чтобы решить задачу, достаточно выбрать в этом прямоугольнике область таких (x, φ) , которые обеспечивают пересечение иглой сетки параллельных линий.

Пусть, например, $L < 1$. Тогда область таких точек состоит из всех (x, φ) , для которых $0 < \varphi \leq 2\pi$ и $-\frac{L}{2} \sin \varphi \leq x \leq \frac{L}{2} \sin \varphi$. Ее площадь равна

$$\int_0^{2\pi} L |\sin \varphi| d\varphi = 4$$

и, следовательно, искомая вероятность равна $2/\pi$.

Оба приема, продемонстрированных на примере задачи Бюффона, применяются для обоснования равномерного распределения в том случае, когда оно не указано явно. В этой задаче построение пуассоновского поля выступает как вспомогательный прием. Однако получаемые таким образом системы геометрических объектов с пуассоновским характером распределения в пространстве часто

интересны сами по себе. Таковы, например, пуассоновские поля прямых в пространстве R^n ($n \geq 2$), пуассоновские поля плоскостей в пространстве R^n ($n \geq 3$) и т. д.

Общая схема построения таких полей состоит в следующем. В пространстве R^n рассматривается множество L геометрических объектов, т. е. класс таких подмножеств, который отображается взаимно однозначно на себя при любом евклидовом движении пространства.

Пусть, например, L — это множество всех прямых в пространстве R^n . При движении пространства прямые переходят в прямые, т. е. класс L инвариантен относительно евклидовых движений.

Пусть каждый элемент $l \in L$ характеризуется k -числами — параметрами.

Таким образом, имеется взаимно однозначное соответствие между L и частью A пространства R^k возможных значений параметров. На множестве A задается пуассоновское поле, может быть, неоднородное, которое однозначно определяется своей мерой интенсивности Λ . Пуассоновское поле случайных геометрических объектов из класса L тем самым задано. Но если Λ — произвольная мера, то она не обязана быть инвариантной относительно евклидовых движений пространства R^n или, что то же самое, пространства L .

Пусть F_1 — отображение $A \rightarrow L$ и F_2 — отображение $L \rightarrow A$, $F_1 = F_2^{-1}$, где F_1 и F_2 — отображения, определяющие прямое и обратное соответствие между параметрами и геометрическими объектами.

Пусть $D : L \rightarrow L$ — движение, переводящее L в L .

Не заботясь о точном определении, предположим, что все эти отображения измеримы и, следовательно, отображение $S : A \rightarrow A$, состоящее из последовательных отображений $A \xrightarrow{(F_1)} L, L \xrightarrow{(D)} L, L \xrightarrow{(F_2)} A$, где $S = F_2 \cdot D \cdot F_1$, также измеримо. Это отображение преобразует меру Λ в меру Λ' по правилу $\Lambda'(B) = \Lambda(S^{-1}(B)) = \Lambda(F_1^{-1}(D^{-1}(F_2^{-1}(B))))$.

Мера Λ называется инвариантной относительно евклидовых движений в пространстве L , если при любом измеримом $B \subset A$ и любом евклидовом движении D $\Lambda(B) = \Lambda'(B)$. Соответственно и пуассоновское поле геометрических объектов из класса L называется инвариантным, если мера Λ инвариантна. В общем случае инвариантная мера Λ не пропорциональна лебеговой мере на множестве A , т. е. пуассоновское поле на множестве A неоднородно.

Среди всех возможных параметризаций множества L наиболее естественны те, для которых инвариантная мера Λ пропорциональна лебеговой мере, т. е. $\Lambda(B) = |B| \cdot k$, где $|B|$ — объем множества B . Нахождение и доказательство их единственности обычно представляет собой трудную математическую задачу.

Рассмотрим инвариантное пуассоновское поле прямых на плоскости. Из двух рассматриваемых параметризаций одна пред-

ставляется более естественной, так как пуассоновское поле в пространстве параметров для нее является однородным.

а). Любая прямая в плоскости, не проходящая через начало координат, представляет собой множество точек (x, y) , связанных уравнением

$$ux + vy + 1 = 0,$$

где u, v — некоторые числа, не равные одновременно нулю. Между множеством L всех прямых, не проходящих через начало координат, и множеством A , совпадающим с плоскостью R^2 , из которой исключена точка $(0, 0)$, таким образом установлено взаимно однозначное соответствие. Любое евклидово движение D в плоскости задается системой двух уравнений, связывающей координаты точек до и после преобразования.

Пусть $D(x, y) = (X, Y)$.

Тогда

$$X = a + x \cos \alpha - y \sin \alpha,$$

$$Y = b + x \sin \alpha + y \cos \alpha,$$

где $(a, b) \in R^2$, $0 \leq \alpha < 2\pi$; a, b и α — параметры евклидова движения в плоскости. Если фиксировано преобразование D , точка $(u, v) \in A$ при преобразовании $S = F_2 \cdot D \cdot F_1$ переходит в некоторую точку (U, V) . При этом F_1 переводит (u, v) в прямую l с уравнением $ux + vy + 1 = 0$, D переводит прямую l в прямую l' с уравнением $uX' + vY' + 1 = 0$, где $(X', Y') = D^{-1}(x, y)$, т. е.

$$X' = -(a \cos \alpha + b \sin \alpha) + x \cos \alpha + y \sin \alpha,$$

$$Y' = (a \sin \alpha - b \cos \alpha) - x \sin \alpha + y \cos \alpha.$$

Отсюда получаем уравнение прямой l' :

$$\frac{x(u \cos \alpha - v \sin \alpha) + y(u \sin \alpha + v \cos \alpha)}{1 - u(a \cos \alpha + b \sin \alpha) + v(a \sin \alpha - b \cos \alpha)} + 1 = 0.$$

Отображение F_2 переводит l' в точку (U, V) , где

$$U = \frac{u \cos \alpha - v \sin \alpha}{1 - u(a \cos \alpha + b \sin \alpha) + v(a \sin \alpha - b \cos \alpha)},$$

$$V = \frac{u \sin \alpha + v \cos \alpha}{1 - u(a \cos \alpha + b \sin \alpha) + v(a \sin \alpha - b \cos \alpha)}. \quad (*)$$

Следовательно, отображение S переводит точку (u, v) в точку (U, V) , где U и V связаны с u и v предыдущими формулами.

Теперь требуется найти такую меру Λ на множестве A , чтобы для любого измеримого $B \subset A$ $\Lambda(B) = \Lambda(S^{-1}(B))$. Мы будем искать такую меру среди абсолютно непрерывных мер, т. е. среди таких мер, для которых существует плотность λ . При этом

$$\Lambda(B) = \int_B \lambda(u, v) du dv$$

$$\Delta(S^{-1}(B)) = \int_{S^{-1}(B)} \lambda(u, v) du dv.$$

Используя правило замены переменных для гладких преобразований $R^2 \rightarrow R^2$, каким является преобразование S^{-1} , получаем

$$\int_{S^{-1}(B)} \lambda(u, v) du dv = \int_B \lambda(S^{-1}(U, V)) |J_{S^{-1}}(U, V)| dU dV,$$

где $J_{S^{-1}}$ — якобиан преобразования S^{-1} (см. II.6) и $S^{-1}(U, V) = (u, v)$.

Решая систему (*), получаем

$$u = \frac{U \cos \alpha + V \sin \alpha}{1 + aU + bV},$$

$$v = \frac{V \cos \alpha - U \sin \alpha}{1 + aU + bV}.$$

Следовательно, якобиан

$$\begin{vmatrix} \frac{\partial u}{\partial U} & \frac{\partial u}{\partial V} \\ \frac{\partial v}{\partial U} & \frac{\partial v}{\partial V} \end{vmatrix} = (1 + aU + bV)^{-3}.$$

Так как B — произвольное множество, то из равенства интегралов следует:

$$\lambda(U, V) = \lambda(S^{-1}(U, V)) \frac{1}{(1 + aU + bV)^3}.$$

Замечаем, что

$$u^2 + v^2 = \frac{U^2 + V^2}{(1 + aU + bV)^2},$$

откуда

$$(1 + aU + bV)^3 = \frac{(U^2 + V^2)^{3/2}}{(u^2 + v^2)^{3/2}}$$

и, значит,

$$\frac{\lambda(u, v)}{\lambda(U, V)} = \frac{(u^2 + v^2)^{-3/2}}{(U^2 + V^2)^{-3/2}}.$$

Отсюда следует, что интенсивность пуассоновского поля $\lambda(u, v)$, инвариантного относительно евклидовых движений плоскости, которой принадлежат прямые, пропорциональна функции $(u^2 + v^2)^{-3/2}$. Например, $\lambda(u, v) = (u^2 + v^2)^{-3/2}$. Плотность не постоянна на A , откуда следует, что пуассоновское поле в пространстве A не однородно, в то время как соответствующее пуассоновское поле прямых на плоскости однородно — инвариантно относительно евклидовых движений плоскости.

Оправданием нашего пренебрежения прямыми, проходящими через начало координат, служит сходимость интеграла

$$\int_{R^2 \setminus S_0} \frac{dudv}{(u^2 + v^2)^{3/2}} = \frac{2\pi}{r_0},$$

где S_0 — круг с центром в начале координат и радиусом r_0 , так как множеству прямых, пересекающих окрестность начала координат, соответствует окрестность бесконечной точки в пространстве параметров.

б). Рассмотрим теперь другую параметризацию множества прямых. Любая прямая, не проходящая через начало координат, однозначно характеризуется длиной перпендикуляра p , опущенного из начала координат на эту прямую, и углом θ , дающим направление перпендикуляра. Напишем уравнение этой прямой:

$$x \frac{\cos \theta}{-p} + y \frac{\sin \theta}{-p} + 1 = 0.$$

Переход к параметрам (p, θ) соответствует замене переменных

$$u = -\frac{\cos \theta}{p}, \quad v = -\frac{\sin \theta}{p},$$

которая определяет взаимно однозначное соответствие между множеством $A_1 = (0, \infty) \times (0, 2\pi]$ параметров (p, θ) и множеством $A = R^2 \setminus \{(0, 0)\}$ параметров (u, v) .

Пуассоновскому полю в пространстве A соответствует пуассоновское поле в пространстве A_1 , которое будет инвариантно относительно евклидовых движений в пространстве прямых тогда и только тогда, когда этим же свойством обладает пуассоновское поле в пространстве A . Покажем, что пуассоновское поле в пространстве A_1 однородно и интенсивность его равна единице, если интенсивность пуассоновского поля в пространстве A равна $\lambda(u, v) = (u^2 + v^2)^{-3/2}$.

Если отображение $F: A_1 \rightarrow A$ переводит точку $(p, \theta) \in A_1$ в точку $(-\frac{\cos \theta}{p}, -\frac{\sin \theta}{p}) \in A$, то мера интенсивности Λ_1 пуассоновского поля в пространстве A_1 , индуцируемая обратным отображением $F^{-1}: A \rightarrow A_1$ и мерой Λ , определяется соотношением

$$\Lambda_1(B_1) = \Lambda(F(B_1)) \quad (B_1 \subset A_1).$$

При этом

$$\Lambda_1(B_1) = \int_{F(B_1)} (u^2 + v^2)^{-3/2} dudv.$$

Последний интеграл по правилу замены переменных

$$\int_{B_1} \left(\left(-\frac{\cos \theta}{p} \right)^2 + \left(-\frac{\sin \theta}{p} \right)^2 \right)^{-3/2} |J_F(p, \theta)| dp d\theta.$$

Так как якобиан

$$\left| \begin{array}{cc} \frac{\partial u}{\partial \theta} & \frac{\partial u}{\partial p} \\ \frac{\partial v}{\partial \theta} & \frac{\partial v}{\partial p} \end{array} \right| = p^{-3},$$

то $\int_{B_1} dp d\theta = |B_1|$, т. е. интеграл равен «площади» множества B_1 .

Следовательно,

$$\Lambda_1(B_1) = |B_1|,$$

т.е. интенсивность пуассоновского поля в пространстве параметров A_1 , инвариантного относительно евклидовых движений в пространстве прямых, равна единице. Можно показать, что параметризация этого вида — единственная из параметризаций прямых на плоскости, для которой инвариантное пуассоновское поле в пространстве параметров однородно (Кендал, Моран, 1972).

Много приложений имеют инвариантные пуассоновские поля прямых и плоскостей в пространстве R^3 . Первые служат, например, моделью волокнистых материалов с «чисто случайным» взаимным расположением волокон. Они могут служить также для обоснования выбора вероятностной меры при решении задач о случайном пересечении прямыми каких-либо тел в пространстве. Пуассоновские поля плоскостей интересны, например, свойствами ячеек пространства, на которые делят пространство «пуассоновские» плоскости.

Большой класс задач связан с оценкой свойств пуассоновских полей выпуклых тел по свойствам их сечений плоскостями или прямыми. Этими задачами занимается раздел теории геометрических вероятностей, называемый стереологией (Santaló, 1976). Использование методов стереологии уточнило представление о результатах количественно-минералогического (модального) анализа (Андерсен, 1978). Множество практических задач, связанных с геометрическими вероятностями, еще ждет своей точной постановки и решения. К их числу относится проблема построения такой случайной мозаики, как мозаика зерен гранита с ячейками, заполненными случайным образом веществом одного из k типов, и такой, что вдоль любого сечения мозаики прямой линией последовательность веществ в пересекаемых ячейках образует марковскую цепь, не являющуюся последовательностью независимых событий.

Практическая деятельность геолога протекает так, что в большинстве случаев ему приходится сравнивать повторяемость изучаемых им событий — частоту. С частотой появления события в большой серии испытаний связана возможность появления события в каждом отдельном испытании. Математическим понятием, дающим меру этой возможности, является вероятность. Оно охва-

тывает те случаи практической деятельности, когда нужно построить научные заключения о повторяемости явлений. Распределение вероятностей отражает условия, в которых реализовались явления (события). Наиболее исследовано поведение вероятностей при независимых испытаниях. Если испытания зависят от того, какой исход имело предыдущее испытание, или от того, каковы координаты точки в той области, где проходят испытания, то нахождение соответствующих вероятностей исходов испытаний значительно осложняется.

Независимые испытания встречаются в геологии весьма редко. Однако на материале независимых испытаний удобнее всего вводить различные понятия, необходимые для описания вероятностной меры в конкретных условиях. Введенные на материале независимых испытаний понятия используют для описания результатов зависимых испытаний. Вводятся здесь, конечно, и новые специфические понятия.

В связи с тем что в геологии особенное значение имеют исходы зависимых испытаний, а им отвечают концепции различных случайных (стохастических) процессов, для геологов очень важна теория этих процессов. При этом имеется большой опыт использования процессов со специфическим последствием, называемых марковскими (говоря о процессе, мы не различаем здесь непрерывный и дискретный случаи).

Вероятностные концепции носят чисто теоретический характер и никак не связаны с вопросом о том, как устанавливать практически их соответствие наблюдениям. Выяснению соотношений между теоретико-вероятностными построениями (моделями геологических процессов в математической геологии) и наблюдениями посвящается следующая глава.

Литература

- Андерсен Р. С. Новейшие результаты в анализе объемного состава. — В кн.: Исследования по математической геологии. Л., «Наука», 1978, с. 187—199.
- Боровков А. А. Курс теории вероятностей. М., «Наука», 1972. 288 с.
- Вистелиус А. Б. Об окатанности кварцевых песчинок Белинского банка (дельта Волги). — Докл. АН СССР, 1948, т. 63, № 1, с. 69—70.
- Вистелиус А. Б. О распространенности энантиоморфных типов кварца. — Зап. ВМО, 1950, ч. 79, № 3, с. 191—195.
- Вистелиус А. Б. Фазовая дифференциация палеозойских отложений Среднего Поволжья и Заволжья. Л., Изд-во АН СССР, 1963. 203 с.
- Вистелиус А. Б. Стохастические модели процессов осадконакопления и их роль в седиментологии. — В кн.: Физические и химические процессы и фации. М., «Наука», 1968, с. 7—14.
- Вистелиус А. Б., Демина М. Е., Харламов Б. П. Основная задача поисковой геохимии и палеогеографии по терригенным компонентам как задача о структуре случайных полей. — В кн.: Геологическая информация и математическая геология. МГК, XXV сессия, докл. сов. геол. М., «Недра», 1976, с. 37—47.
- Вистелиус А. Б., Романова М. А. Красноцветные отложения полуострова Челекен. Л., Изд-во АН СССР, 1962. 227 с.

- Гельфанд И. М., Шилов Г. Е. Обобщенные функции и действия над ними. М., «Наука», 1958. 347 с.
- Гихман И. И., Скороход А. В. Теория случайных процессов. Т. 1. М., «Наука», 1971. 664 с.
- Гнеденко Б. В. Курс теории вероятностей. М., Изд-во физ.-мат. лит-ры, 1961. 406 с.
- Дынкин Е. Б. Основания теории марковских процессов. М., Изд-во физ.-мат. лит-ры, 1959. 227 с.
- Дынкин Е. Б. Марковские процессы. М., Изд-во физ.-мат. лит-ры, 1963. 859 с.
- Ибрагимов И. А., Розанов Ю. А. Гауссовские случайные процессы. М., «Наука», 1970. 384 с.
- Иванов Д. Н., Подольский Ю. В. Об оценке связи между количественными характеристиками при геолого-геохимических исследованиях. — Советская геология, 1971, № 9, с. 137—141.
- Ито К. Вероятностные процессы, т. II. М., ИЛ, 1963. 135 с.
- Ито К., Маккин Г. Диффузионные процессы и их траектории. М., «Мир», 1968. 394 с.
- Кендалл М., Моран П. Геометрические вероятности. М., «Наука», 1972. 192 с.
- Кендалл М. Дж., Стьюарт А. Теория распределений. М., «Наука», 1966. 587 с.
- Кендалл М., Стьюарт А. Статистические выводы и связи. М., «Наука», 1973. 899 с.
- Кокс Д., Смит В. Теория восстановления. М., «Советское радио», 1967. 396 с.
- Колмогоров А. Н. Основные понятия теории вероятностей. М., «Наука», 1974. 120 с.
- Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. М., «Наука», 1972. 496 с.
- Крамер Г. Математические методы статистики. М., «Мир», 1975. 648 с.
- Кульбак С. Теория информации и статистика. М., «Наука», 1967. 408 с.
- Мизес Р. Вероятность и статистика. М., 1930. 208 с.
- Неве Ж. Математические основы теории вероятностей. М., «Мир», 1969. 309 с.
- Рао С. Р. Линейные статистические методы и их применение. М., «Наука», 1968. 547 с.
- Романова М. А. Современные песчаные отложения Центральных Каракумов. Л., «Наука», 1971. 256 с.
- Скитович В. П. Линейные формы от независимых случайных величин и нормальный закон распределения. — Изв. АН СССР, 1954, т. 18, с. 185—200.
- Тутубалин В. Н. Теория вероятностей. М., Изд-во МГУ, 1972. 229 с.
- Фаас А. В., Сарманов О. В. О распределении процентных величин. — В кн.: Вопросы математической геологии. Л., «Наука», 1968, с. 111—117.
- Феллер В. Введение в теорию вероятностей и ее приложения, т. II. М., «Мир», 1967. 752 с.
- Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления. 3. М., Изд-во физ.-мат. лит-ры, 1960. 656 с.
- Харламов Б. П. Об одной математической модели накопления акцессорных минералов в осадочных отложениях. — В кн.: Исследования по математической геологии. Л., «Наука», 1978, с. 80—89.
- Хеннан Э. Анализ временных рядов. М., «Наука», 1964. 216 с.
- Эйнштейн А., Смолуховский М. Броуновское движение. М., ОНТИ, 1936. 150 с.
- Bernstein S. N. Sur les courbes de distribution des probabilités. — Math. Z., 1925, No. 12, S. 75—85.
- Kallenberg O. Lectures on random measures. Inst. of Statistics, Goteborg, Mimeo Ser. Nov., 1974, No. 963. 180 p.

- Santalo L. Integral Geometry and Geometric Probability. — Encycl. Math. and Appl. v. 1. Ed. Gian-Carlo Rota. London e. a., Addison-Wesley Publ. Co., v. 17. 404 p.
- Vistelius A. B. Studies in Mathematical Geology. New York, Consultant Bureau, 1967. 294 p.
- Vistelius A. B. Ideal granite and its properties. I. Stochastic Model. — J. Int. Assoc. Math. Geol., 1972, v. 4, No. 2, p. 35—65.
- Wickman F. E., El-Hinnawi E. E. The time distribution of lava-fragment ejection from volcanoes. — Arkiv för Miner. och Geol., 1963, Bd 3, No. 16, p. 363—383.

Глава III

ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ПОНЯТИЯ — ВОПРОСЫ ОЦЕНИВАНИЯ И ПРОВЕРКА ГИПОТЕЗ

Положение статистики в геологии, ее задачи и методы. Вопросы оценивания параметров по выборочным данным. Необходимость исследования процедур оценивания на примере изучения парагенезисов. Информационные статистики, их смещение. Понятие достаточности. Проверка геологических построений с помощью статистических гипотез. Анализ распределения Na_2O в базальтах на основе изложенных методов.

Ключевые слова:

случайный выбор, оценка, проверка гипотез, ошибки первого и второго рода, мощность критерия.

III.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Независимо от того, какие методы применяет геолог в своей работе, ему всегда необходимо проверять по наблюдениям свои теоретические представления (согласно нашей терминологии, модель). Для этой цели он прибегает к различным техническим средствам, дающим, с его точки зрения, искомые результаты. В геологических науках, в частности, широко используются различные графические методы. При этом в спорах с представителями точных наук, отмечающих субъективный характер подобных методов, от геологов часто можно слышать, что доказательства заключаются в своеобразной логике геологических исследований, которая обеспечивает объективность заключений. Существуют, якобы, специфические геологические доказательства.

Подобные рассуждения имеют какой-то смысл до тех пор, пока задача не переведена на формальный язык математики, т.е. до тех пор, пока не дана *математическая модель* явления. Математическая модель должна содержать все существенные черты явления.

После того как модель сформулирована, ее необходимо сравнить с наблюдениями, для того чтобы выяснить, противоречит или нет модель наблюдениям. Этим вопросом занимается специальная математическая дисциплина — *математическая статистика*.

Решение вопроса в этом случае дается в оптимальной форме. Так как модель отражает суть геологической задачи, то статистическое решение является окончательным и целиком охватывает все вопросы, касающиеся проверки модели, в том числе и геологические доказательства.

В настоящей главе мы рассмотрим, как по наблюдениям оцениваются параметры распределения вероятностей, отвечающие той или иной модели, и каким образом выясняется непротиворечивость этого распределения вероятностей наблюдениям. Независимо от того, как проведены наблюдения и какова модель, оценка параметров и проверка непротиворечивости модели проводится методами математической статистики. Однако результат статистического исследования в сильнейшей степени зависит от того, как проведены наблюдения (опробование). Статистическую задачу невозможно решить, не зафиксировав метод опробования. При этом, говоря об опробовании, обычно придерживаются следующей терминологии, которой мы будем следовать и здесь. Результаты опробования, т.е. цифры анализов, замеров мощностей, совокупность экземпляров собранной фауны, называют *выборкой*, процесс сбора этих данных — *выбором*, а число наблюдений — *объемом выборки*. Выбор может быть осуществлен самыми разнообразными методами (Cochran, 1961). Однако математическая статистика в настоящее время разработана в основном для случая, называемого *простым случайным выбором*. Если наш материал не удовлетворяет случаю простого случайного выбора, то, для того чтобы получить правильные результаты, необходимо статистические вопросы разрабатывать в соответствии со спецификой задачи. В противном случае решения, полученные на основе предположения о том, что данные собраны с помощью простого случайного выбора, могут оказаться заведомо ошибочными.

В настоящей главе мы рассмотрим основные понятия статистики. Это делается для того, чтобы сориентировать в них геологов. Изложение иллюстрируется примерами. При этом во всех случаях предполагается, что осуществлен простой случайный выбор. Делается это не из-за того, что автор уверен в реальности такого выбора в задачах геологии, а лишь потому, что в этом случае достигается наибольшая простота исследования.

Что такое простой выбор, проще всего понять из следующего примера. Пусть мы имеем пробу кварцевого песка, состоящую из многих миллионов песчинок. Допустим, что изучаемый песок обладает очень высокой сортированностью, т.е. размер песчинок колеблется в узком интервале, и кроме того все песчинки имеют близкую степень окатанности. В силу того что проба взята из одного слоя, порожденного одним процессом накопления песчинок, мы принимаем, что любая песчинка является порождением одного и того же процесса с фиксированной функцией распределения характеристик песчинок. Функция распределения вероятностей характеристик песчинок одна и та же для всех песчинок.

Если отмеченную пробу мы тщательно перемешаем, придадим ей форму тонкого слоя и из точек этого слоя с координатами, определенными по таблице случайных чисел, выберем N зерен, беря в каждой точке по одному зерну, то полученные N зерен могут считаться извлеченными в результате простого выбора. Предположение о том, что N зерен извлечено в результате простого выбора, справедливо в том случае, если *функция распределения* изучаемых характеристик для всех зерен одна и та же, с *одними и теми же параметрами*, и если гарантирована *стохастическая независимость* характеристик извлекаемого зерна от таких же характеристик остальных зерен. Указанная черта выбора называется *независимостью наблюдений*. Термин этот общепринят, но крайне неудачен, так как создается впечатление, что речь идет о процессе наблюдения. Между тем термин подразумевает лишь *независимость результатов наблюдений*. Итак, *составление выборки из индивидов, несущих стохастически независимые значения изучаемой характеристики с одной и той же функцией распределения вероятностей этой характеристики для всех индивидов, называется простым случайным выбором*.

Естественно, что свойства геологических объектов таковы, что часто нет никакой гарантии в том, что возможен простой случайный выбор. Очень сложен вопрос и о возможности извлечения независимых наблюдений. Ярким примером этого является провал международного проекта по определению точности анализов силикатов (Vistelius, 1971). Если об объекте заранее ничего неизвестно и он подвергается опробованию, то зависимость или независимость наблюдений может быть оценена только после проведения опробования. Зависимость или независимость наблюдений неизвестного объекта в принципе определяется природой объекта, в соответствии с которой должны осуществляться манипуляции лица, производящего опробование.¹ Зная свойства объекта, соответствующими манипуляциями можно искусственно отбирать независимые наблюдения.

При работе с геологическими материалами приходится руководствоваться не столько принципами, сколько возможностью использования этих принципов. Соответственно при оценке совокупности возможны три ситуации.

Первая ситуация начинается с того, что, желая охарактеризовать совокупность, геолог собирает данные. При этом, пользуясь, скажем, химическими анализами, он оценивает точность примененных методов анализа вообще, репутацию лабораторий, данными которых он пользовался, квалификацию и добросовестность аналитиков. Однако он не располагает информацией о том, как собраны данные и какие свойства характерны для исследуемой совокупности. Выводы, полученные в таком случае о совокупности,

¹ В специальной литературе по математической статистике соответствующие манипуляции носят название эксперимента, а условия, в которых они осуществляются, — постановкой эксперимента.

весьма ограничены, не допускают оценки их реальности и пригодны только для самых грубых, предварительных прикидок. На таком уровне оценки находятся, как правило, различные подсчеты баланса, выполняемые геохимиками, равно как и данные о распространенности в земной коре различных химических элементов.

Вторая ситуация характеризуется тем, что имеются данные о сборе материала, однако хотя и известно, что для всех наблюдений функция распределения одна и та же, неизвестно, какова эта функция. В то же время тем или иным способом обеспечена независимость наблюдений. Возможности здесь тоже ограничены, и анализ таких данных требует специального подхода.

Наконец, возможен случай, когда геолог имеет модель, в которой функция распределения известна с точностью до параметра, и осуществлен простой случайный выбор. Наша задача — оптимальное оценивание параметров функции распределения и выяснение непротиворечивости модели наблюдениям. Это типичная задача, допускающая содержательное решение, на ней и будет сконцентрировано наше внимание.

Если мы пытаемся оценить параметры, то математическая статистика предлагает для этого ряд методов. Ниже мы рассмотрим смысл так называемого *точечного оценивания*, исследуем свойства оценок и покажем на примерах, почему этой теорией следует пользоваться. В дальнейшем будет дано также понятие об *интервальном оценивании*.

Проверка модели посредством сравнения отвечающей ей функции распределения вероятностей с наблюдениями в аспекте браковки или приемки модели осуществляется методами теории проверки гипотез. Ниже даются основные определения этой теории и приводится пример конкретного исследования, показывающий опасность стандартного подхода и необходимость углубленного анализа даже, казалось бы, в простых вопросах.

При решении задач часто приходится оперировать с понятием «параметр» — некоторой величиной, определяющей конкретный вид функции распределения совокупности. При этом имеются два основных подхода. При первом подходе предполагается, что параметр является постоянной величиной, неизвестной константой, и мы его оцениваем по выборке. Этим подходом мы будем пользоваться особенно часто.

Однако существует точка зрения на параметр, как на случайную величину. Такая точка зрения носит название *байесовской*. На ней мы подробно останавливаться не будем.

III.2. ТОЧЕЧНОЕ ОЦЕНИВАНИЕ

Проблема точечного оценивания очень обширна, и излагать ее систематически при наличии отличных руководств бессмысленно. Здесь мы попытаемся познакомить геолога с наиболее

важными определениями и показать, что при численном расчете оценок параметров нужно быть осторожным. Подробно излагаемые примеры дают представление, какая техника требуется для нахождения «хороших» оценок.

III.2.1. Основные идеи и определения

Допустим, что вид функции распределения $F(x; \theta)$ нам известен с точностью до параметра θ . Также известно, какие допустимые значения может принимать параметр θ этого распределения, но неизвестно, какое конкретное значение θ_0 для исследуемой совокупности он имеет. *Статистическая задача, заключающаяся в том, чтобы по данным выборки тем или иным способом оценить параметр как неизвестную константу (точку), называется задачей точечного оценивания.* Всякая функция, зависящая только от наблюдений, называется *статистикой*. Метод получения оценки называется *оцениванием*. Оценка параметра θ обозначается в дальнейшем $\bar{\theta}(x)$. Оценка является случайной величиной, функция распределения которой $P_n(\bar{\theta}; \theta)$ зависит от θ и объема выборки n .

Функцию $\bar{\theta}(x)$ для оценивания θ можно строить разными способами (по методу моментов, максимального правдоподобия и т.п.). В зависимости от выбранного способа оценка может иметь различные свойства, которые сказываются на качестве $\bar{\theta}(x)$ как приближении к параметру.

Если математическое ожидание оценки совпадает с истинным значением параметра при всех допустимых его значениях, то такую оценку называют несмещенной. Встречаются как несмещенные оценки, так и имеющие смещение. Так, например, если распределение случайной величины сосредоточено на конечном сегменте $[a, b]$, то длину этого сегмента $b - a$ естественно оценивать с помощью выборочного размаха изучаемой величины x , т.е. $x_{\max} - x_{\min}$. Ясно, что x_{\max} всегда меньше или равно b , а x_{\min} всегда равно или больше a . Таким образом, в среднем размах оценивает длину интервала (размах случайной величины) с недостатком, т.е. обладает смещением. Если оценка обладает смещением, то естественно вводить поправку. Как находится величина поправки, будет показано в примере.

Если имеются две разные оценки $\bar{\theta}_1(x)$ и $\bar{\theta}_2(x)$ одного и того же параметра θ , то оценка с меньшей дисперсией, при прочих равных условиях, представляется лучшей. *Оценку с меньшей дисперсией иногда называют более эффективной.*

Пусть имеется, например, равномерное распределение на $[0; \theta]$ и требуется оценить неизвестный параметр θ этого распределения. Если взять две оценки: оценку по методу моментов

$$\bar{\theta}_1(x) = \frac{2 \sum_{i=1}^n x_i}{n}$$

и оценку, найденную по методу максимального правдоподобия и исправленную на смещение,

$$\bar{\theta}_2(x) = \frac{n+1}{n} x_{\max},$$

то оказывается, что обе эти оценки несмещенные, а отношение их дисперсий

$$\frac{D[\bar{\theta}_1(x)]}{D[\bar{\theta}_2(x)]} = \frac{n+2}{3},$$

т.е. их эффективности различаются на порядок объема выборки.

С эффективностью связано число наблюдений, требуемое для достижения оценки параметра с заданной точностью. Низкая эффективность вызывает необходимость лишних наблюдений. Примером менее эффективной оценки может служить выборочная медиана по сравнению со средним арифметическим, если она используется для точечной оценки математического ожидания в нормальном распределении.

От хорошей оценки естественно требовать, чтобы при неограниченном росте выборки распределение $\bar{\theta}(x)$ стремилось к такому распределению, математическое ожидание которого совпадает с θ (*асимптотическая несмещенность*), а дисперсия стремилась бы к нулю (*асимптотическая эффективность*). В таком случае с вероятностью, сколь угодно близкой к единице, параметр будет отличаться от оценки произвольно мало при достаточно больших выборках. Указанное свойство оценок называется *состоятельностью*.

По выборке можно строить статистики различной размерности. В частности, сама выборка x_1, \dots, x_n представляет n -мерную статистику. Построенная по выборке статистика

$$\left(\begin{array}{c} \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^2 \end{array} \right) \quad (\text{III. 2. 4})$$

является двумерной.

Переходя от всей выборки к статистикам меньшей размерности, можно потерять часть информации о параметре данного распределения. Если этого не происходит, то статистика размерности меньшей, чем n , содержит столько же информации о параметре, сколько выборка. Такую статистику называют *нетривиальной достаточной статистикой параметра* в данном распределении (тривиальной достаточной статистикой для параметра является вся выборка). Так, например, если совместная плотность при простом выборе нормальна, т.е.

$$f(x) = 2^{-n/2} \pi^{-n/2} \sigma^{-n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2 \right],$$

то для двумерного параметра (a, σ^2) достаточной статистикой будет двухкомпонентный вектор (III.2.1), но не существует достаточной статистики размерности меньшей, чем два.

Ниже даются примеры вычисления оценки по наблюдениям, расчеты и практические рекомендации, относящиеся к смещенности, а также рассматривается влияние ситуации, в которой реализуется наблюдение, на свойства статистики.

Пример III.4. Оценки переходных вероятностей в предположении их пропорциональности

Цель исследования — оценивание переходных вероятностей в некоторых типах марковских матриц, встречаемых в геологических задачах. Вычисления проводятся методом *максимального правдоподобия*, который, как правило, дает «хорошие» оценки. Решая отмеченную задачу, мы поясним сначала метод максимального правдоподобия, а затем перейдем к самой задаче.

Пусть x_1, x_2, \dots, x_n — наблюдения, совместная вероятность появления которых

$$p(x_1, x_2, \dots, x_n; \theta_1, \dots, \theta_k). \quad (\text{III.2.2})$$

Функциональная форма p совместной вероятности (III.2.2) предполагается известной.² $\theta_1, \dots, \theta_k$ — параметры, относительно которых возможны три случая: либо они известны, либо известна часть из них, либо они неизвестны. В том случае, когда параметры известны, функция $p(x_1, \dots, x_n; \theta_1, \dots, \theta_k)$ характеризует вероятность появления выборочной точки x и задает распределение вероятностей в n -мерном пространстве $W = (x_1, \dots, x_n)$, называемым *выборочным*. В том случае, когда параметры неизвестны, функция $p(x_1, \dots, x_n; \theta_1, \dots, \theta_k)$ носит название функции правдоподобия.

Функция правдоподобия используется при проверке гипотез и точечном оценивании. Здесь мы рассмотрим ее применение для получения точечных оценок параметров.

Оценки максимального правдоподобия — это такие допустимые значения параметров $\theta_1, \dots, \theta_k$, как функций наблюдений x_1, x_2, \dots, x_n , подстановка которых на место $\theta_1, \dots, \theta_k$ в функцию правдоподобия дает возможно большее значение этой функции.

Таким образом, отыскание оценок максимального правдоподобия представляет задачу определения экстремальных точек — требуется найти такую точку

² Выражение (III.2.2) является вероятностью появления выборочной точки x_1, \dots, x_n , если распределение случайной величины x дискретно; если (III.2.2) представляет производную от функции распределения вероятностей, то она является плотностью вероятности.

$$\hat{\Theta}(x) = \begin{pmatrix} \hat{\theta}_1(x) \\ \hat{\theta}_2(x) \\ \vdots \\ \hat{\theta}_k(x) \end{pmatrix},$$

подстановка которой вместо

$$\Theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{pmatrix}$$

дала бы точную верхнюю границу значений этой функции.

Функция правдоподобия зависит от выборочной точки и в разных выборках является различной, поэтому оценка $\hat{\theta}(x)$ представляет собой случайный вектор.

С вопросами существования оценки максимального правдоподобия, ее единственности и качества оценок, получаемых по методу максимального правдоподобия, читатель может познакомиться по книге Рао (1968). В целом можно сказать, что при соблюдении некоторых условий, которые, как правило, практически имеют место, оценки максимального правдоподобия существуют, являются состоятельными, асимптотически несмещенными и обладают хорошей эффективностью в больших выборках. В малых же выборках метод максимального правдоподобия может иметь ряд недостатков: оценки оказываются смещенными, эффективность далека от оптимальной и т.д.

Таким образом, метод максимального правдоподобия не гарантирует оценок с оптимальными качествами. Он является стандартным рабочим методом, который обычно применим. Другие стандартные методы, используемые для получения точечных оценок, либо дают, как правило, не лучшие результаты (метод моментов), либо реже доступны для применения (несмещенные оценки с минимальной дисперсией), либо дают примерно те же результаты (оценки минимума χ^2). Вопрос выбора метода — это вопрос технической разрешимости конкретных задач. Рассматриваемый ниже пример содержит минимум вычислительных трудностей и одновременно решает нужный вопрос.

Переходим к интересующей нас задаче — к оценке переходных вероятностей, которая нам потребуется при разборе моделей.

Пусть имеются два распределения сгруппированных данных с неизвестными нам вероятностями

$$\begin{aligned} p_{11}, p_{12}, \dots, p_{1m} \left(\sum_{i=1}^m p_{1i} = 1 \right), \\ p_{21}, p_{22}, \dots, p_{2m} \left(\sum_{i=1}^m p_{2i} = 1 \right). \end{aligned} \quad (\text{III. 2. 3})$$

Фактически нам понадобится случай, когда p_{ij} — переходные вероятности, т.е. (III.2.3) представляют собой две строки в матрице переходных вероятностей. В то же время все сказанное применимо к любым двум дискретным распределениям вероятностей с одинаковым числом классов.

Существует предположение, что некоторые пары вероятностей p_{1i}, p_{2i} пропорциональны с коэффициентом λ . Так, например, пусть

$$\frac{p_{11}}{p_{21}} = \frac{p_{13}}{p_{23}} = \dots = \frac{p_{1k}}{p_{2k}} = \lambda \quad (k < m),$$

где коэффициент пропорциональности λ нам неизвестен. Другие пары не предполагаются пропорциональными. Условимся нумеровать вероятности так, чтобы сначала были занумерованы вероятности пропорциональных пар. Рассмотрим сначала случай, когда m в (III.2.3) равно трем. Тогда мы будем иметь две строчки вероятностей

$$\begin{array}{ccc} p_{11} & p_{12} & p_{13}, \\ p_{21} & p_{22} & p_{23}, \end{array}$$

причем

$$\frac{p_{11}}{p_{21}} = \frac{p_{12}}{p_{22}} = \lambda.$$

Необходимо найти оценки максимального правдоподобия

$$\hat{p}_{11}, \hat{p}_{12}, \hat{p}_{13}, \hat{p}_{21}, \hat{p}_{22}, \hat{p}_{23}.$$

Для получения оценок воспользуемся количеством наблюдений, попадающих в соответствующие клетки (частоты):

$$n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23};$$

объемы выборок,

$$N_1 = n_{11} + n_{12} + n_{13},$$

$$N_2 = n_{21} + n_{22} + n_{23},$$

предполагаются фиксированными. Это дополнительное предположение сделано для упрощения выкладок. Допущение о постоянном объеме выборок эквивалентно тому, что безусловные вероятности попадания в первую и вторую строчки заданы:

$$p_1 = \frac{N_1}{N}, \quad p_2 = \frac{N_2}{N} \quad (N = N_1 + N_2).$$

Так как нас интересует только соотношение пропорциональности между переходными вероятностями, наличие которой не зависит от выбора безусловных вероятностей, то обе выборки N_1 и N_2 можно предположить фиксированными. Таким образом, n_{11}, n_{12}, n_{13} — случайные величины, а $N_1 = n_{11} + n_{12} + n_{13}$ — случайная величина. Это справедливо и для второй строчки.

Оценки максимального правдоподобия рассчитываются так.

1. Составляем функцию правдоподобия, для чего перемножаем выражения для вероятностей каждого наблюдения $p_i(x_i, \theta)$ по всем наблюдениям вместе:

$$L(\theta; x_1, \dots, x_n) = p_1(x_1; \theta) p_2(x_2; \theta), \dots, p_n(x_n; \theta).$$

В нашем случае наблюдения размещены по таблице 2×3 , чему соответствует полиномиальное распределение по шести клеткам с вероятностями π_{ij} ($i=1, 2; j=1, 2, 3$) попадания наблюдения в ij -тую клетку. Как известно, полиномиальное распределение

$$p(n_{11}, \dots, n_{23}; \pi_{11}, \pi_{12}, \dots, \pi_{23}) = \frac{N!}{n_{11}! n_{12}! \dots n_{23}!} \pi_{11}^{n_{11}} \dots \pi_{23}^{n_{23}}. \quad (\text{III. 2. 4})$$

Запишем (III. 2. 4) в форме, включающей оцениваемые параметры. Вероятность попадания в строчку, вероятность попадания в клетку и переходная вероятность связаны соотношением

$$\pi_{ij} = p_i p_{ij} \quad (i=1, 2; j=1, 2, 3). \quad (\text{III. 2. 5})$$

Так как вероятности p_i фиксированы,

$$p_1 = \frac{N_1}{N} \quad \text{и} \quad p_2 = \frac{N_2}{N},$$

то из (III. 2. 5) имеем

$$\pi_{ij} = \frac{N_i}{N} p_{ij}.$$

Подставляя найденные значения π_{ij} в (III. 2. 4), получаем

$$p(n_{11}, \dots, n_{23}; p_{11}, \dots, p_{23}) = \frac{N!}{n_{11}! \dots n_{23}!} \left(\frac{N_1}{N}\right)^{N_1} \left(\frac{N_2}{N}\right)^{N_2} p_{11}^{n_{11}} \dots p_{23}^{n_{23}}.$$

Используем теперь специфический вид совместного распределения. Вместо p_{21} мы должны подставить λp_{11} , а вместо p_{22} — величину λp_{12} . Тогда вместо p_{23} следует подставить $1 - \lambda(p_{11} + p_{12})$. Точно так же, чтобы избавиться от зависимого параметра p_{13} , подставляем вместо него $1 - p_{11} - p_{12}$. Тогда получаем

$$p(n_{11}, \dots, n_{23}; p_{11}, p_{12}) = \frac{N!}{n_{11}! \dots n_{23}!} \left(\frac{N_1}{N}\right)^{N_1} \left(\frac{N_2}{N}\right)^{N_2} p_{11}^{n_{11} + n_{21}} p_{12}^{n_{12} + n_{22}} \lambda^{n_{21} + n_{22}} \times \\ \times (1 - p_{11} - p_{12})^{n_{23}} [1 - \lambda(p_{11} + p_{12})]^{n_{23}}; \quad (\text{III. 2. 6})$$

$p(n_{11}, \dots, n_{23}; p_{11}, p_{12})$ в (III. 2. 6) представляет собой функцию правдоподобия.

2. Постоянные множители не влияют на положение экстремальных точек, поэтому мы их отбрасываем. Тогда из правой части (III. 2. 6) получаем

$$p_{11}^{n_{11} + n_{21}} p_{12}^{n_{12} + n_{22}} \lambda^{n_{21} + n_{22}} (1 - p_{11} - p_{12})^{n_{23}} [1 - \lambda(p_{11} + p_{12})]^{n_{23}}. \quad (\text{III. 2. 7})$$

Часто полезно логарифмировать полученное выражение, так как, будучи монотонным преобразованием, логарифмирование не изменяет положения экстремумов и может упрощать полученное выражение.

Логарифмируя (III.2.7), получаем

$$\mathcal{L} = (n_{11} + n_{21}) \ln p_{11} + (n_{12} + n_{22}) \ln p_{12} + (n_{21} + n_{22}) \ln \lambda + \\ + n_{13} \ln (1 - p_{11} - p_{12}) + n_{23} \ln [1 - \lambda (p_{11} + p_{12})]. \quad (\text{III. 2. 8})$$

3. Находим значения p_{11} , p_{12} и λ , максимизирующие (III.2.8). Для этого берем частные производные по p_{11} , p_{12} и λ и находим точки, где система этих производных обращается в нуль:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = (n_{21} + n_{22}) \frac{1}{\lambda} - n_{23} \frac{p_{11} + p_{12}}{1 - \lambda (p_{11} + p_{12})}, \\ \frac{\partial \mathcal{L}}{\partial p_{11}} = (n_{11} + n_{21}) \frac{1}{p_{11}} - \frac{n_{13}}{1 - p_{11} - p_{12}} - \frac{\lambda n_{23}}{1 - \lambda (p_{11} + p_{12})}, \\ \frac{\partial \mathcal{L}}{\partial p_{12}} = (n_{12} + n_{22}) \frac{1}{p_{12}} - \frac{n_{13}}{1 - p_{11} - p_{12}} - \frac{\lambda n_{23}}{1 - \lambda (p_{11} + p_{12})}. \quad (\text{III. 2. 9})$$

4. Решая систему, получаем

$$\hat{\lambda} = \frac{n_{21} + n_{22}}{N_2 (p_{11} + p_{12})}, \quad \hat{p}_{11} = \frac{(n_{11} + n_{21}) (n_{12} + n_{11})}{N_1 (n_{11} + n_{12} + n_{21} + n_{22})}, \\ \hat{p}_{12} = \frac{(n_{11} + n_{12}) (n_{12} + n_{22})}{N_1 (n_{11} + n_{12} + n_{21} + n_{22})}, \quad \hat{p}_{13} = 1 - \hat{p}_{11} - \hat{p}_{12}, \\ \hat{p}_{21} = \frac{(n_{11} + n_{21}) (n_{21} + n_{22})}{N_2 (n_{11} + n_{12} + n_{21} + n_{22})}, \quad \hat{p}_{22} = \frac{(n_{12} + n_{22}) (n_{21} + n_{22})}{N_2 (n_{11} + n_{12} + n_{21} + n_{22})}, \quad (\text{III. 2. 10}) \\ \hat{p}_{23} = 1 - \hat{p}_{21} - \hat{p}_{22}.$$

Нетрудно убедиться, что оценки (III. 2. 10) действительно дают максимум функции правдоподобия.

Пусть теперь m произвольно. Мы условились непропорциональные разряды нумеровать последними. Сведем эти разряды в один, тогда получим

$$p_{11} p_{12} \dots p_{1t} \sum_{j=t+1}^m p_{1j} = p'_{1(t+1)}, \\ p_{22} p_{22} \dots p_{2t} \sum_{j=t+1}^m p_{2j} = p'_{2(t+1)},$$

причем

$$\frac{p_{11}}{p_{21}} = \frac{p_{12}}{p_{22}} = \dots = \frac{p_{1t}}{p_{2t}} = \lambda.$$

Введем следующие обозначения:

$$s_{i.} = \sum_{j=1}^t n_{ij} \quad (i=1, 2) \quad s_{.j} = \sum_{i=1}^2 n_{ij} \quad (j=1, \dots, t), \\ S = \sum_{\substack{i=1, 2 \\ j=1, \dots, t}} n_{ij}.$$

Та же техника приводит к следующим оценкам:

$$\hat{p}_{ij} = \frac{s_{i.} s_{.j}}{S N_i} \quad (i=1, 2; j=1, 2, \dots, t),$$

$$P'_{i(t+1)} = 1 - \sum_{j=1}^t P_{ij} \quad (i = 1, 2). \quad (\text{III. 2. 11})$$

Произведем практические расчеты. При исследовании гранита из Якутии (Романова, 1976) изучались последовательности зерен ортоклаза (Or), кварца (Q) и плагиоклаза (Ab). При этом нужно было найти оценки переходных вероятностей для марковских матриц вида

$$P_{I, J; K} (I, J, K \in \{Or, Q, Ab\})$$

в строчках

$$P(Or^{(h)}/Ab^{(h-1)}Ob^{(h-2)}), \quad P(Q^{(h)}/Ab^{(h-1)}Q^{(h-2)}), \quad P(Ab^{(h)}/Ab^{(h-1)}Or^{(h-2)}), \\ P(Or^{(h)}/Ab^{(h-1)}Q^{(h-2)}), \quad P(Q^{(h)}/Ab^{(h-1)}Q^{(h-2)}), \quad P(Ab^{(h)}/Ab^{(h-1)}Q^{(h-2)}),$$

когда одна из проверяемых моделей предполагала пропорциональность

$$\frac{P(Or^{(h)}/Ab^{(h-1)}Or^{(h-2)})}{P(Or^{(h)}/Ab^{(h-1)}Q^{(h-2)})} = \frac{P(Q^{(h)}/Ab^{(h-1)}Or^{(h-2)})}{P(Q^{(h)}/Ab^{(h-1)}Q^{(h-2)})} \neq \frac{P(Ab^{(h)}/Ab^{(h-1)}Or^{(h-2)})}{P(Ab^{(h)}/Ab^{(h-1)}Q^{(h-2)})},$$

где h , $h-1$ и $h-2$ указывают место состояний в последовательности. Подсчеты под микроскопом дали следующие частоты:

	Or ^(h)	Q ^(h)	Ab ^(h)
Or ^(h-2) Ab ^(h-1)	42	45	55
Q ^(h-2) Ab ^(h-1)	52	61	50

Применение формул (III. 2. 11) или, что то же в данном случае (III. 2. 10), дает

$$\hat{P}(Or^{(h)}/Ab^{(h-1)}Or^{(h-2)}) = 0.2880, \quad \hat{P}(Q^{(h)}/Ab^{(h-1)}Or^{(h-2)}) = 0.3247, \\ \hat{P}(Ab^{(h)}/Ab^{(h-1)}Or^{(h-2)}) = 0.3873, \quad \hat{P}(Or^{(h)}/Ab^{(h-1)}Q^{(h-2)}) = 0.3258, \\ \hat{P}(Q^{(h)}/Ab^{(h-1)}Q^{(h-2)}) = 0.3674, \quad \hat{P}(Ab^{(h)}/Ab^{(h-1)}Q^{(h-2)}) = 0.3068, \\ \hat{\lambda} \approx 1.131.$$

Пример III.2. О смещении информационных статистик

В практике геологов встречаются задачи, требующие привлечения различных коэффициентов, зависящих от выражений вида

$$H_k(x) = - \sum_{i=1}^k p_i \ln p_i, \\ H_{m,r}(x, y) = - \sum_{\substack{i=1, \dots, m \\ j=1, \dots, r}} p_{ij} \ln p_{ij}. \quad (\text{III. 2. 12})$$

$H_k(x)$ и $H_{m,r}(x, y)$ называются *энтропией Шеннона* соответственно для строки и двухвходовой таблицы. Здесь $p_i (i=1, 2, \dots, k)$ — вероятности попадания значения случайной величины x в i -тую клетку, т.е. параметр полиномиального распределения, к которому приводит схема независимых испытаний, а p_{ij} — ве-

роятность попадания в ij -тую клетку для некоторого полиномиального распределения с $m \times r$ клетками. Коэффициенты указанного вида используются при оценке связи в двух- и многоходовых таблицах сопряженности (Кульбак, 1967). При этом в двух- и трехходовых таблицах связь оценивается с помощью статистик

$$\left. \begin{aligned} \bar{I}_{m,r}(x, y) &= \sum_{\substack{i=1, 2, \dots, m \\ j=1, 2, \dots, r}} \bar{p}_{ij} \ln \frac{\bar{p}_{ij}}{\bar{p}_{i.} \bar{p}_{.j}} \\ \text{или} \\ \bar{I}_{m,r,s}(x, y, z) &= \sum_{\substack{i=1, 2, \dots, m \\ j=1, 2, \dots, r \\ k=1, 2, \dots, s}} \bar{p}_{ijk} \ln \frac{\bar{p}_{ijk}}{\bar{p}_{i.} \bar{p}_{.j} \bar{p}_{..k}}, \end{aligned} \right\} \quad (\text{III. 2. 13})$$

где \bar{p}_{ij} — оценка соответствующих вероятностей p_{ij} (здесь точка вместо нижнего индекса указывает на суммирование по всем значениям этого индекса).

Статистика (III.2.13) для двумерного случая дает *среднюю информацию*, содержащуюся в строках относительно столбцов. Смысл статистики в (III.2.13) для трехмерного случая аналогичен. Если умножить статистику (III.2.13) на n — общее число наблюдений, то получим статистику информационной силы связи в двухходовой (соответственно и в трехходовой) таблице сопряженности.

Поскольку в (III.2.13) $\bar{I}_{m,r}(x, y)$ можно представить в виде

$$\bar{I}_{m,r}(x, y) = \bar{H}_{m,r}(x, y) - \bar{H}_m(x) - \bar{H}_r(y), \quad (\text{III. 2. 14})$$

где волна указывает на то, что мы имеем дело с оценкой, отсюда следует, что информационная статистика силы сопряженности подсчитывается с помощью оценивания трех энтропий Шеннона.

Также нередко используют *информационный коэффициент корреляции*, оцениваемый по формуле

$$\bar{\rho}_{m,r}(x, y) = \sqrt{1 - e^{-2|\bar{I}_{m,r}(x, y)|}}, \quad (\text{III. 2. 15})$$

и, таким образом, $\bar{\rho}$ также оценивается через три энтропии (Кульбак, 1967).

Статистики $\bar{H}_k(x)$, $\bar{I}_{m,r}(x, y)$ и $\bar{\rho}_{m,r}(x, y)$ являются асимптотически несмещенными, но, как мы покажем в дальнейшем, при умеренном объеме выборки и значительном количестве классов смещение может оказаться очень значительным. Все эти статистики, как мы видели, выражаются через энтропию Шеннона, поэтому прежде всего необходимо изучать ее смещение. Относительно энтропии Шеннона известна следующая теорема Башарина (Башарин, 1959): $\bar{H}_k(x)$ *распределена асимптотически нормально с параметрами*

$$E[\bar{H}_k(x)] = - \sum_{i=1}^k p_i \ln p_i + o\left(\frac{1}{n}\right) = H_0 + o\left(\frac{1}{n}\right), \quad (\text{III. 2. 16})$$

$$D[\tilde{H}_k(x)] = \frac{1}{n} \left[\sum_{i=1}^k p_i (\ln p_i)^2 - H_0^2 \right] + o\left(\frac{1}{n^2}\right), \quad (\text{III. 2. 17})$$

при этом n означает объем выборки, а $H_0 = -\sum p_i \ln p_i$ — энтропию.

Формулы (III. 2. 16) и (III. 2. 17) свидетельствуют, что поправка на смещение не вводится.

Точно так же для статистик $\tilde{I}_{m,r}(x, y)$ и соответственно $\tilde{\rho}_{m,r}(x, y)$ приводятся формулы (Кульбак, 1967)

$$\begin{aligned} & E[\tilde{I}_{m,r}(x, y)] = I_0 + o\left(\frac{1}{n}\right) \\ \text{и} & E[\tilde{\rho}_{m,r}(x, y)] = \rho_0 + o\left(\frac{1}{n}\right), \end{aligned} \quad (\text{III. 2. 18})$$

где I_0 и ρ_0 — теоретические значения соответствующих коэффициентов (т. е. выражения (III. 2. 13) для I и (III. 2. 15) для ρ , куда вместо оценок подставлены значения соответствующих параметров).

Ниже будет показано, что во многих случаях, важных для практики, при расчете \tilde{H} необходимо вводить поправку на смещение. При этом расчет поправки разумно делать с точностью до $o(1/n^2)$. Точно так же при расчете дисперсии $D(\tilde{H})$ во многих случаях следует вводить поправку на смещение. Эту поправку рационально вводить с точностью до $o(1/n^3)$. Если бы мы не вводили поправок на смещение и пользовались бы нормальной аппроксимацией Башарина (III.2.16) и (III.2.17), то в случае равномерного распределения (частный случай равенства параметров полиномиального распределения)

$$p_1 = p_2 = \dots = p_k = \frac{1}{k} \quad (\text{III. 2. 19})$$

мы получили бы дисперсию, равную нулю, что лишено смысла. При полиномиальных распределениях, близких к (III.2.19), потребовалось бы чрезвычайно большое число наблюдений для того, чтобы обеспечить удовлетворительную точность. Такое уменьшение точности связано с тем, что в случае (III.2.19) уже нельзя пользоваться нормальной аппроксимацией для \tilde{H} (случайная величина $2n |H_0 - \tilde{H}|$, как можно показать, асимптотически распределена как центральное χ^2 с $k - 1$ степенью свободы). У ненормального распределения будет другое математическое ожидание случайной величины и ее дисперсии, чем у нормального, причем они сдвинуты по отношению к (III.2.16) и (III.2.17).

В общем случае величина смещения зависит от значений параметров, которые нам неизвестны. Если вместо параметров подставить их оценки, то такая операция сама вызовет некоторое смещение. Однако при введении поправки с точностью до $o(1/n^2)$ для \tilde{H} и с точностью до $o(1/n^3)$ — для $D(\tilde{H})$ такие поправки оказы-

ваются не зависящими от значений параметров (дальнейшие поправки более высоких порядков уже зависят от значений параметров p_i). Одновременно поправки указанного порядка обеспечивают достаточную точность и на практике.

Ниже дается подробный вывод поправки на смещение для \bar{H} , а для $D(\bar{H})$ приводится только конечный результат.

При расчете смещения в обоих случаях мы используем формулу Тейлора, так как обычно употребляемый метод Кенуэя для устранения смещения (Кендалл, Стьюарт, 1966, 1973) неудобен для приращения к величине $\bar{p}_i \ln \bar{p}_i$.

Найдем сначала знак смещения. Согласно *неравенству Йенсена* (Рао, 1968, с. 62), для выпуклых функций справедливо соотношение

$$E f(\chi) \geq f(E\chi), \quad (\text{III. 2. 20})$$

где χ — случайная величина, а $f(\chi)$ — выпуклая функция от χ . Так как для функции

$$f(\bar{p}_i) = \bar{p}_i \ln \bar{p}_i$$

математическое ожидание существует и вторая производная отрицательна (она равна $1/\bar{p}_i$, если $\bar{p}_i \neq 0$), то это выпуклая функция, и мы можем применить к ней (III. 2. 20), откуда

$$E(\bar{p}_i \ln \bar{p}_i) \geq E\bar{p}_i \ln(E\bar{p}_i).$$

Учитывая, что частость есть несмещенная оценка вероятности

$$E\bar{p}_i = p_i,$$

получаем

$$E(\bar{p}_i \ln \bar{p}_i) \geq p_i \ln p_i.$$

Отсюда

$$E\left(-\sum_{i=1}^k \bar{p}_i \ln \bar{p}_i\right) \leq -\sum_{i=1}^k p_i \ln p_i$$

или

$$E\bar{H}_k(x) \leq H_k(x).$$

Иными словами, выборочная оценка энтропии всегда занижена в среднем, поэтому для исправления смещения к оценке должна быть прибавлена некоторая положительная поправка.

Далее для упрощения мы принимаем, что $0 < \bar{p}_i < 1$ при $i=1, 2, \dots, k$. Таким образом, мы не допускаем в нашей классификационной таблице пустых клеток.

Докажем теперь, что для любого $0 \leq \bar{p}_i \leq 1$ справедливо следующее соотношение:

$$\begin{aligned} \bar{p}_i \ln \bar{p}_i = \bar{p}_i \ln p_i + \frac{\bar{p}_i (\bar{p}_i - p_i)}{p_i} - \frac{\bar{p}_i (\bar{p}_i - p_i)^2}{2p_i^2} + \\ + \frac{\bar{p}_i (\bar{p}_i - p_i)^3}{3p_i^3} + \bar{v}_i, \end{aligned} \quad (\text{III. 2. 21})$$

где

$$|\bar{v}_i| < \frac{(\bar{p}_i - p_i)^4}{p_i^4},$$

\bar{p}_i и \bar{v}_i — случайные величины, p_i — параметр.

Пусть $\bar{p}_i = 0$.

Тогда (III. 2. 21) преобразуется к виду

$$0 = \bar{v}_i, \quad |\bar{v}_i| < 1. \quad (\text{III. 2. 22})$$

Примем теперь, что $0 < \bar{p}_i \leq 1$, и введем случайную величину

$$\bar{\delta}_i = \frac{\bar{p}_i - p_i}{p_i},$$

которая во всех случаях больше, чем -1 .

Ряд Тейлора для функции $\ln(1 + \bar{\delta}_i)$ сходится только в том случае, когда $\bar{\delta}_i \in (-1; 1)$, но формула Тейлора верна для $\bar{\delta}_i \in (-1, h]$, где $h > 0$.

Перепишем теперь выражение

$$\ln \left[1 + \frac{(\bar{p}_i - p_i)}{p_i} \right] = \ln(1 + \bar{\delta}_i),$$

используя формулу Тейлора с остаточным членом в форме Коши.

Тогда

$$\begin{aligned} \ln \left[1 + \frac{(\bar{p}_i - p_i)}{p_i} \right] &= \frac{\bar{p}_i - p_i}{p_i} - \frac{(\bar{p}_i - p_i)^2}{2p_i^2} + \frac{(\bar{p}_i - p_i)^3}{3p_i^3} - \\ &- \frac{(\bar{p}_i - p_i)^4 [1 - \bar{\theta}(p_i, \bar{p}_i)]^3}{p_i^4 \left[1 + \bar{\theta}(p_i, \bar{p}_i) \frac{(\bar{p}_i - p_i)}{p_i} \right]^4}, \end{aligned} \quad (\text{III. 2. 23})$$

где для случайной величины $\bar{\theta}(p_i, \bar{p}_i)$ справедливо неравенство

$$0 < \bar{\theta}(p_i, \bar{p}_i) < 1.$$

Умножая правую и левую части (III. 2. 23) на \bar{p}_i , получаем

$$\begin{aligned} \bar{p}_i \ln \bar{p}_i &= \bar{p}_i \ln p_i + \frac{\bar{p}_i (\bar{p}_i - p_i)}{p_i} - \frac{\bar{p}_i (\bar{p}_i - p_i)^2}{2p_i^2} + \frac{\bar{p}_i (\bar{p}_i - p_i)^3}{3p_i^3} - \\ &- \frac{\bar{p}_i (\bar{p}_i - p_i)^4 [1 - \bar{\theta}(p_i, \bar{p}_i)]^3}{p_i^4 \left[1 + \bar{\theta}(p_i, \bar{p}_i) \frac{(\bar{p}_i - p_i)}{p_i} \right]^4} \left[1 + \bar{\theta}(p_i, \bar{p}_i) \frac{(\bar{p}_i - p_i)}{p_i} \right]^3. \end{aligned} \quad (\text{III. 2. 24})$$

Обозначим последний член в (III. 2. 24) через $\bar{v}_i(\bar{p}_i, p_i, \bar{\theta})$.

Докажем теперь, что

$$|\bar{v}_i(\bar{p}_i, p_i; \bar{\theta})| < \left(\frac{\bar{p}_i - p_i}{p_i} \right)^4 \quad (\text{III. 2. 25})$$

при условиях

$$0 < p_i < 1, \quad \bar{p}_i > 0 \quad \text{и} \quad 0 < \bar{\theta} < 1.$$

Из очевидного неравенства

$$0 < \frac{1 - \bar{\theta}}{1 + \bar{\theta} \frac{(\bar{p}_i - p_i)}{p_i}} < 1$$

следует, что

$$|v_i| < \frac{\bar{p}_i (\bar{p}_i - p_i)^4}{p_i^3 [p_i + \bar{\theta} (\bar{p}_i - p_i)]} = \frac{\bar{p}_i (\bar{p}_i - p_i)^4}{p_i^3 [(1 - \bar{\theta}) p_i + \bar{\theta} \bar{p}_i]}. \quad (\text{III. 2. 26})$$

Так как $(1 - \bar{\theta}) p_i + \bar{\theta} \bar{p}_i$ положительна и не меньше $\min(p, \bar{p}_i)$, то

$$|\bar{v}_i| \leq \frac{\bar{p}_i (\bar{p}_i - p_i)^4}{p_i^3 \min(p_i, \bar{p}_i)} = r(\bar{p}_i; p_i) = \begin{cases} \frac{(\bar{p}_i - p_i)^4}{p_i^3}, & \text{если } \bar{p}_i < p_i, \\ \frac{\bar{p}_i (\bar{p}_i - p_i)^4}{p_i^4}, & \text{если } \bar{p}_i \geq p_i. \end{cases}$$

Введем функцию

$$r^*(\bar{p}_i; p_i) = \frac{(\bar{p}_i - p_i)^4}{p_i^4}.$$

Очевидно,

$$r^*(\bar{p}_i; p_i) \geq r(\bar{p}_i; p_i)$$

для значений $\bar{p}_i > 0$ и $0 < p_i < 1$.

Итак,

$$|\bar{v}_i| < \frac{(\bar{p}_i - p_i)^4}{p_i^4};$$

одновременно с (III. 2. 25) доказано также (III. 2. 21).

Используем теперь неравенство

$$|E\bar{v}_i| \leq E|\bar{v}_i| < Er^*(\bar{p}_i; p_i).$$

Тогда

$$E(\bar{p}_i \ln \bar{p}_i) = \ln p_i E\bar{p}_i + \frac{1}{p_i} E\bar{p}_i^2 - E\bar{p}_i - \frac{1}{2p_i^2} E\bar{p}_i^3 - \frac{1}{2} E\bar{p}_i + \frac{1}{p_i} E\bar{p}_i^2 + \\ + \frac{1}{3p_i^3} E\bar{p}_i^4 - \frac{1}{3} E\bar{p}_i - \frac{1}{p_i^2} E\bar{p}_i^3 + \frac{1}{p_i} E\bar{p}_i^2 + E\bar{v}_i(\bar{p}_i, p_i, \theta), \quad (\text{III. 2. 27})$$

где

$$E\bar{v}_i(\bar{p}_i, p_i, \theta) \leq Er^*(\bar{p}_i, p_i) \leq \frac{1}{p_i^4} E\bar{p}_i^4 - \frac{4}{p_i^3} E\bar{p}_i^3 + \frac{6}{p_i^2} E\bar{p}_i^2 - \frac{4}{p_i} E\bar{p}_i + 1.$$

После приведения подобных членов и сокращения

$$E(\bar{p}_i \ln \bar{p}_i) = \ln p_i E\bar{p}_i - \frac{11}{6} E\bar{p}_i + \frac{3}{p_i} E\bar{p}_i^2 - \frac{3}{2p_i^2} E\bar{p}_i^3 + \\ + \frac{1}{3p_i^3} E\bar{p}_i^4 + E\bar{v}_i(\bar{p}_i, p_i, \theta). \quad (\text{III. 2. 28})$$

Найдем теперь моменты $E\bar{p}_i$, $E\bar{p}_i^2$, $E\bar{p}_i^3$ и $E\bar{p}_i^4$, подставим их в (III.2.27) и (III.2.28) и проведем суммирование по i от 1 до k . Напомним, что для вычисления k -того момента случайной величины χ удобно взять характеристическую функцию $\varphi_\chi(t)$ этой величины, получить ее k -тую производную $\varphi_\chi^{(k)}(t)$ по t , вычислить

величину производной при $t=0$ и полученный результат разделить на i^k .

Изучаемая случайная величина \bar{p}_i совпадает с величиной \tilde{n}_i/n , где \tilde{n}_i — число наблюдений в i -той клетке, а n — общее число наблюдений; \tilde{n}_i имеет биномиальное распределение, поскольку попадание в клетку может рассматриваться как успех испытания в биномиальной схеме с вероятностью успеха p_i . Непопадание в клетку рассматривается как неуспех испытания, который имеет вероятность $q_j = 1 - p_j$. В этом случае характеристическая функция для \bar{p}_j

$$\varphi_{\bar{p}_j}(t) = \left(p_j e^{i \frac{t}{n}} + q_j \right)^n.$$

Прямое применение изложенной процедуры для нахождения моментов дает

$$\begin{aligned} E\bar{p}_i &= p_i, \\ E\bar{p}_i^2 &= \frac{p_i}{n} + \frac{(n-1)p_i^2}{n}, \\ E\bar{p}_i^3 &= \frac{p_i}{n^2} + \frac{p_i^2(3n-3)}{n^2} + \frac{(n-1)(n-2)p_i^3}{n^2}, \\ E\bar{p}_i^4 &= \frac{p_i}{n^3} + \frac{7(n-1)p_i^2}{n^3} + \frac{6(n-1)(n-2)p_i^3}{n^3} + \frac{(n-1)(n-2)(n-3)p_i^4}{n^3}. \end{aligned}$$

Подставляя полученные моменты в (III. 2. 27) и (III. 2. 28), получаем

$$\begin{aligned} E(\bar{p}_i \ln \bar{p}_i) &= p_i \ln p_i - \frac{11}{6} p_i + \frac{3}{p_i} \left[\frac{p_i}{n} + \frac{(n-1)p_i^2}{n} \right] - \\ &- \frac{3}{2p_i^2} \left[\frac{p_i}{n^2} + \frac{p_i^2(3n-3)}{n^2} + \frac{(n-1)(n-2)p_i^3}{n^2} \right] + \\ &+ \frac{1}{3p_i^3} \left[\frac{p_i}{n^3} + \frac{7(n-1)p_i^2}{n^3} + \frac{6(n-1)(n-2)p_i^3}{n^3} + \right. \\ &\left. + \frac{(n-1)(n-2)(n-3)p_i^4}{n^3} \right] + t_i, \end{aligned}$$

где

$$\begin{aligned} t_i &< \frac{1}{p_i^4} \left[\frac{p_i}{n^3} + \frac{7(n-1)p_i^2}{n^3} + \frac{6(n-1)(n-2)p_i^3}{n^3} + \frac{(n-1)(n-2)(n-3)p_i^4}{n^3} \right] - \\ &- \frac{4}{p_i^3} \left[\frac{p_i}{n^2} + \frac{p_i^2(3n-3)}{n^2} + \frac{(n-1)(n-2)p_i^3}{n^2} \right] + \frac{6}{p_i^2} \left[\frac{p_i}{n} + \frac{(n-1)p_i^2}{n} \right] - 3. \end{aligned}$$

Приводя подобные члены и суммируя по i от 1 до k , получаем

$$\begin{aligned} E \sum_{i=1}^k \bar{p}_i \ln \bar{p}_i &= \sum_{i=1}^k p_i \ln p_i + \frac{k-1}{2n} + \frac{1}{n^2} \left(\frac{5}{6} \sum_{i=1}^k \frac{1}{p_i} + \frac{2}{3} - \frac{3k}{2} \right) + \\ &+ \frac{1}{n^3} \left(\frac{1}{3} \sum_{i=1}^k \frac{1}{p_i^2} - \frac{7}{3} \sum_{i=1}^k \frac{1}{p_i} + 4k - 2 \right) + \sum_{i=1}^k t_i, \quad (\text{III. 2. 29}) \end{aligned}$$

где

$$\begin{aligned} \sum_{i=1}^k t_i < \frac{1}{n^2} \left(3 \sum_{i=1}^k \frac{1}{p_i^2} - 6 \sum_{i=1}^k \frac{1}{p_i} + 3k \right) + \\ + \frac{1}{n^3} \left(\sum_{i=1}^k \frac{1}{p_i^3} - 7 \sum_{i=1}^k \frac{1}{p_i^2} + 12 \sum_{i=1}^k \frac{1}{p_i} - 6k \right). \end{aligned} \quad (\text{III. 2. 30})$$

Таким образом,

$$E \sum_{i=1}^k \tilde{p}_i \ln \tilde{p}_i = \sum_{i=1}^k p_i \ln p_i + \frac{k-1}{2n} + w,$$

причем, согласно (III. 2. 29) и (III. 2. 30),

$$w = o\left(\frac{1}{n^2}\right).$$

Таким образом,

$$E(\tilde{H}_k(x)) = H_0 - \frac{k-1}{2n} + o\left(\frac{1}{n^2}\right).$$

Иными словами, к оценке $\tilde{H}_k(x)$ следует добавлять поправку на смещение, равную $\frac{k-1}{2n}$, после чего остающееся смещение имеет порядок $1/n^2$.

Из тех же формул (III. 2. 29) и (III. 2. 30) видно, что члены порядка $1/n^2$ уже зависят от параметров p_i .

Уточним теперь выборочную оценку дисперсии статистики $\tilde{H}_k(x)$.

Мы уже использовали обозначение $-\sum_{i=1}^k p_i \ln p_i = H_0$.

Введем аналогично построенное обозначение

$$\frac{1}{n} \left[\sum_{i=1}^k p_i (\ln p_i)^2 - H_0^2 \right] = D_0.$$

Тогда, пользуясь методикой, аналогичной только что использованной, получим

$$D(\tilde{H}_k(x)) = D_0 + \frac{k-1}{2n^2} + o\left(\frac{1}{n^3}\right). \quad (\text{III. 2. 31})$$

Итак, при оценивании дисперсии выборочной энтропии Шеннона из полученного значения дисперсии $\tilde{H}_k(x)$ нужно вычесть $\frac{k-1}{2n^2}$, после чего остающееся смещение будет иметь порядок $1/n^3$, причем дальнейшая поправка уже будет зависеть от параметров p_i .

Используя полученные результаты, внесем теперь поправку на смещение в статистику количества информации в двухвходовой таблице сопряженности $\tilde{I}_{m,r}(x,y)$.

Поправка на смещение в (III.2.14) складывается из поправок для каждой из трех энтропий.

В итоге получаем

$$\begin{aligned} \Delta \tilde{I}_{m,r}(x,y) &= -\frac{mr-1}{2n} + o\left(\frac{1}{n^2}\right) + \frac{m-1}{2n} + o\left(\frac{1}{n^2}\right) + \frac{r-1}{2n} + o\left(\frac{1}{n^2}\right) = \\ &= -\frac{(m-1)(r-1)}{2n} + o\left(\frac{1}{n^2}\right). \end{aligned} \quad (\text{III. 2. 32})$$

Таким образом, выборочная информация в среднем завышена и требует вычитания поправки в соответствии с (III.2.32). Из (III.2.32) для статистики информационной силы связи, т.е. для $\tilde{I}_{m,r} = n\tilde{I}_{m,r}$, получаем поправку на смещение, равную

$$-\frac{(m-1)(r-1)}{2} + o\left(\frac{1}{n}\right).$$

Расчеты для трехвходовой таблицы сопряженности дают поправку

$$\Delta \tilde{I}_{m,r,s}(x,y,z) = -\frac{1}{2n}(mrs - m - r - s + 2) + o\left(\frac{1}{n^2}\right). \quad (\text{III. 2. 33})$$

Эта поправка должна быть вычтена из величины, полученной по соответствующей формуле в (III.2.13).

Рассмотрим теперь поправку для статистики информационного коэффициента корреляции

$$\tilde{\rho}_{m,r}(x,y) = \sqrt{1 - e^{-2|\tilde{I}_{m,r}(x,y)|}}. \quad (\text{III. 2. 34})$$

Если в (III.2.34) не вводить поправку на смещение для $\tilde{I}_{m,r}(x,y)$, то, очевидно, среднее значение $\tilde{\rho}_{m,r}(x,y)$ окажется завышенным, как это видно из (III.2.32).

Если ввести поправку $-\frac{(m-1)(r-1)}{2n}$ в значение $\tilde{I}_{m,r}(x,y)$, то величина $\tilde{\rho}_{m,r}(x,y)$ будет в среднем занижена. Однако если мы введем поправку, соответствующую (III.2.32), в $\tilde{I}_{m,r}(x,y)$ и используем исправленную величину $\tilde{I}_{m,r}(x,y)$, равную $\hat{I}_{m,r}(x,y)$, для вычисления

$$\hat{\rho}_{m,r}(x,y) = \sqrt{1 - e^{-2|\hat{I}_{m,r}(x,y)|}},$$

то получим статистику $\hat{\rho}_{m,r}(x,y)$, смещение которой очень мало и имеет отрицательный знак.

Действительно, $\hat{\rho}_{m,r}(x,y)$ есть вогнутая функция $\hat{I}_{m,r}(x,y)$. Поэтому из неравенства Йенсена следует, что

$$E\hat{\rho}_{m,r}(x,y) = E\sqrt{1 - e^{-2|\hat{I}_{m,r}(x,y)|}} \leq \sqrt{1 - e^{-2|E\hat{I}_{m,r}(x,y)|}}.$$

Далее, как было показано, $I_{m,r}(x, y)$ практически не имеет смещения, т. е. $E\hat{I}_{m,r}(x, y) \approx I_{m,r}(x, y)$. Подстановка $E\hat{I}_{m,r}(x, y) \approx I_{m,r}(x, y)$ в правую часть неравенства для $E\hat{\rho}_{m,r}(x, y)$ дает окончательно

$$E\hat{\rho}_{m,r}(x, y) \leq \sqrt{1 - e^{-2|I_{m,r}(x, y)|}} = \rho_{m,r}(x, y).$$

Таким образом, параметрическая точка $\rho_{m,r}(x, y)$ несколько больше, чем математическое ожидание статистики $\hat{\rho}_{m,r}(x, y)$.

Знание смещения позволяет исправлять статистики, определять, сколько классов должно быть использовано при фиксированном числе наблюдений, каково должно быть разумное число наблюдений, если мы фиксируем размерность таблицы сопряженности (это особенно важно для таблиц со многими входами). Формулы, выведенные в этом разделе, дают результаты, точность которых зависит от числа наблюдений (n) и числа классов (k). Точность повышается с ростом n и уменьшением k . Приемлемые соотношения между n и k могут быть получены путем соответствующих подсчетов. Как правило, отношение k/n должно быть порядка 0.1 или менее. В этом случае наши рекомендации дают приемлемые поправки.

Рассмотрим некоторые примеры.

В табл. III.1 приведены частоты ориентировки острых концов зерен песка в свите надкирмакинских песчаников продуктивной толщи Азербайджана в Кирмакинской долине. Цифры показывают, что величина энтропии порядка 2.17, т. е. в нашем случае поправка несущественна.

Следующий пример (основанный на результатах вычислений $\hat{I}_{m,r}(x, y)$ и $\hat{\rho}_{m,r}(x, y)$) посвящен изучению связи между TiO_2 и P_2O_5 в гранитоидах (с содержанием $SiO_2 \geq 60\%$) Швейцарии (без Гельветского покрова). Не учитывая величины поправок и зависимости их от величин m, r и n , для вычисления $\hat{\rho}$ мы построили табл. III.2 (Вистелиус, 1964). Разработав только что изложенную методику исправления смещения, мы убедились, что при имеющемся числе наблюдений разумно работать лишь с таблицами 3×3 или 2×2 . Таблица 3×3 дана ниже под номером III.3. Для иллюстрации влияния соотношения между k и n на величину поправки мы произвели подсчеты также по таблице 5×5 , которую не приводим.

Результаты подсчетов таковы:

$$\begin{aligned} I_{15,14}(TiO_2, P_2O_5) &= 0.5915, & \hat{\rho}_{15,14}(TiO_2, P_2O_5) &= 0.83, \\ I_{5,5}(TiO_2, P_2O_5) &= 0.0601, & \hat{\rho}_{5,5}(TiO_2, P_2O_5) &= 0.34, \\ I_{3,3}(TiO_2, P_2O_5) &= 0.0315, & \hat{\rho}_{3,3}(TiO_2, P_2O_5) &= 0.25. \end{aligned}$$

Рассчитаем теперь поправки на смещение для таблиц 5×5 и 3×3 . Результаты после введения поправки следующие:

$$\begin{aligned} \hat{I}_{5,5}(TiO_2, P_2O_5) &= 0.0071, & \hat{\rho}_{5,5}(TiO_2, P_2O_5) &= 0.12, \\ \hat{I}_{3,3}(TiO_2, P_2O_5) &= 0.0180, & \hat{\rho}_{3,3}(TiO_2, P_2O_5) &= 0.19. \end{aligned}$$

Распределение частот по разрядам в 20° от 270 до 90° ($n = 227$) (Грязнова, 1953)

Разряды	270 < 290	290 < 310	310 < 330	330 < 350	350 < 10	10 < 30	30 < 50	50 < 70	70 < 90
\bar{n}_i	26	31	33	17	18	32	32	23	15
\bar{p}_i	0.1145	0.1366	0.1454	0.0749	0.0793	0.1410	0.1410	0.1013	0.0661
$ \ln p_i $	2.16718	1.99070	1.92827	2.59160	2.53452	1.95900	1.95900	2.28967	2.71659

Примечание. $\bar{H}_9 = 2.15938$, $\Delta \bar{H}_9 = 0.01762$, $\bar{H}'_9 = 2.1770$. Последующие члены поправки, соответствующие $o\left(\frac{1}{n^2}\right)$, вместе взятые, не превосходят 0.003, они вычисляются по (III. 2.29) и (III. 2.30), для этого в указанные формулы вместо параметров p_i подставляются их оценки \bar{p}_i .

Таблица сопряженности между P_2O_5 и TiO_2 (в %) для гранитоидов Швейцарии (Вистелиус, 1964)

TiO_2	P_2O_5														n
	< 0.03	$0.03 < 0.06$	$0.06 < 0.09$	$0.09 < 0.12$	$0.12 < 0.15$	$0.15 < 0.18$	$0.18 < 0.21$	$0.21 < 0.24$	$0.24 < 0.27$	$0.27 < 0.30$	$0.30 < 0.33$	$0.33 < 0.36$	$0.36 < 0.39$	≥ 0.39	
< 0.1	3	3	1	3	1	2		4	1	2	1	1		5	27
$0.1 < 0.2$	1		4	1		3		1							10
$0.2 < 0.3$	1	1	1	2		2	3	1		1	1			1	15
$0.3 < 0.4$		1	4		3		2	4	4	1		2	1	2	20
$0.4 < 0.5$			4	3	1	3	4	2	1	2		1		2	24
$0.5 < 0.6$	1			2	2	3	1	2	3		1	1		2	18
$0.6 < 0.7$		1			1	2		2			2		1	5	13
$0.7 < 0.8$				2	1		2	2						1	8
$0.8 < 0.9$				2					1		1		1	2	7
$0.9 < 1.0$											1				1
$1.0 < 1.1$								1							1
$1.1 < 1.2$					1										1
$1.2 < 1.3$										1	1				2
$1.3 < 1.4$						1							1		2
≥ 1.4	6	6	14	15	10	16	12	17	10	7	9	4	5	20	151

Таблица сопряженности между TiO_2 и P_2O_5 в гранитоидах Швейцарии (Вистеллус, 1964)

TiO_2	P_2O_5			
	<0.15	0.15<0.30	≥ 0.30	Σ
0.30 < 0.30	22	20	40	52
< 0.60	21	29	12	62
≥ 0.60	8	13	16	37
Σ	51	62	38	151

Расчеты показывают, что оценка стабилизировалась, а ее незначительные расхождения скорее всего обязаны эффекту группировки.

Итак, параметрическое значение $\rho(TiO_2, P_2O_5)$ скорее всего около 0.15. Судя по тому что $0 < \rho < 1$, можно полагать, что при $\bar{\rho}=0.15$ и 151-ом наблюдении гипотеза о наличии связи между TiO_2 и P_2O_5 скорее всего не поддерживается наблюдениями. Если бы мы не знали о величине поправки и не использовали ее, то мы приняли бы $\rho_{15,14}(TiO_2, P_2O_5)=0.83$, что можно было бы рассматривать как сильный аргумент в пользу гипотезы о наличии связи.

Пример III.3. О достаточности максимального наблюдения при оценивании правой точки усечения

Если наибольшее значение, которое может принимать случайная величина χ , есть θ , то распределение этой случайной величины называется ограниченным справа, а параметр θ — верхней границей. Аналогично вводится распределение, ограниченное слева. Во многих задачах геологии представляют интерес неотрицательные случайные величины, ограниченные справа и слева. Такие случайные величины имеют две границы, но левая граница, равная нулю, известна. Таким образом, остается задача оценивания точки θ . Существует ряд ситуаций, когда естественная оценка x_{\max} (максимальное наблюдение в выборке) является достаточной для правой точки θ , что кажется интуитивно ясным. Цель настоящего примера — выяснение и разграничение ситуаций, в которых x_{\max} является достаточной статистикой, и тех, в которых оно может ею не быть.

Выяснение этого вопроса практически важно потому, что в ряде случаев — при ограниченном времени исследования, трудной достижимости объекта, большой стоимости наблюдений и т. д. — необходимо по возможности уменьшить число измерений. В част-

ности, было бы весьма желательно знать заранее, нужно нам измерять только максимальное значение исследуемого признака или необходимо также фиксировать значения других наблюдений.

При рассмотрении поставленной задачи мы примем упрощающее предположение, что у наблюдаемой случайной величины существует плотность, а наблюдения собраны с помощью простого случайного выбора.

Разделим все ограниченные плотности на два класса. В первый войдут те плотности, которые описываются следующей стохастической схемой: исходная плотность $f(x; \theta)$, отражающая природное явление, не имеет правой граничной точки, т. е. случайная величина хотя и с чрезвычайно малыми вероятностями может принимать сколь угодно большие значения. Далее, наблюдения, большие, чем θ , отбрасываются из каких-то предметных соображений и плотность оставшейся части наблюдений умножается на компенсирующий постоянный множитель, т. е. вводится плотность усеченного распределения

$$f^*(x; \theta) = \frac{f(x)}{\int_0^{\theta} f(x) dx}. \quad (\text{III. 2. 35})$$

Как будет показано в дальнейшем, для всех плотностей $f^*(x, \theta)$ значение x_{\max} является достаточной статистикой.

Существуют, однако, стохастические схемы, согласно которым усечение возникает как следствие соответствующего механизма, определяющего явление, а не в процессе урезания и перенормирования по (III.2.35). Для этого класса плотностей x_{\max} не обязано быть достаточной статистикой, что будет показано ниже. Для выяснения, когда в последнем случае x_{\max} оказывается достаточной статистикой, существует следующая теорема.

Т е о р е м а III.1. Пусть имеется плотность вероятности $f(x, \theta)$, сосредоточенная на сегменте $[0; \theta]$, и опробование производится с помощью простого случайного выбора. Статистика x_{\max} достаточна для параметра θ в том и только в том случае, когда плотность можно представить в виде

$$f(x; \theta) = t(\theta) r(x) \quad (x < \theta). \quad (\text{III. 2. 36})$$

При этом первый множитель в правой части (III.2.36) не зависит от наблюдений, но кроме θ может зависеть от каких-либо других параметров, входящих в плотность, а второй множитель в правой части не должен зависеть от параметра θ .

Доказательство этой теоремы не приводится, так как оно может быть получено с помощью незначительного изменения результатов в разделах 17.40 и 17.41 в курсе Кендалла и Стьюарта (1973).

Рассмотрим теперь $f^*(x; \theta)$ в (III.2.35) с точки зрения определения (III.2.36). Положим в (III.2.35)

$$\frac{1}{\int_0^\theta f(x) dx} = t(\theta) \text{ и } f(x) = r(x).$$

Тогда $f^*(x; \theta) = t(\theta)r(x)$, откуда следует, что x_{\max} является достаточной статистикой для $f^*(x; \theta)$.

Однако приводимый ниже пример показывает, что имеются случаи, когда x_{\max} не является достаточной статистикой. Рассмотрим для этого плотность, сосредоточенную на $[0; \theta]$:

$$f(x; \theta) = \begin{cases} \frac{\pi}{2\theta} \sin \frac{\pi x}{2\theta}, & \text{если } x \in [0; \theta], \\ 0, & \text{если } x \notin [0; \theta], \end{cases} \quad (\text{III. 2. 37})$$

где угловой аргумент $\frac{\pi x}{2\theta}$ относится к первому квадранту и $f(x; \theta)$ неотрицательна.

Поскольку

$$\frac{\pi}{2\theta} \int_0^\theta \sin \frac{\pi x}{2\theta} dx = 1,$$

функция (III. 2. 37) является плотностью. Очевидно, что плотность (III. 2. 37) не может быть разделена на два множителя, из которых один содержит x , а другой только θ . В этом случае плотность $f(x, \theta)$ нельзя представить в виде (III. 2. 26), и поэтому x_{\max} не является достаточной статистикой для θ .

Если мы сравним эти два случая с точки зрения существа дела, то сможем обнаружить следующие соотношения. В том случае, когда x_{\max} является достаточной статистикой, кривая $f^*(x; \theta)$ подобна $f(x)$, ее конфигурация определена $f(x)$, а большая высота по сравнению с $f(x)$ целиком зависит от положения точки усечения θ . Наблюдения над точками, принадлежащими кривой слева от θ , не содержат никакой новой информации, если мы знаем аналитическую форму $f(x)$.

В случае плотности (III. 2. 37) изменение значений θ приводит к неподобному преобразованию конфигурации кривой. Когда мы приводим в соответствие новому значению θ возникающую при этом конфигурацию кривой, то мы тем самым связываем значение каждой точки на кривой со значением θ . Таким образом, требуются все наблюдения, расположенные слева от θ в (III. 2. 37), для того чтобы по выборке оценить θ . Каждое из этих наблюдений содержит информацию о форме $f(x; \theta)$, а следовательно, и о позиции θ . Это объясняет, почему одно x_{\max} недостаточно для оценки положения θ по выборке.

На рис. III. 1 приведено усеченное распределение зерен граната в песке с оз. Мичиган. Представим себе, что песок просеян через некоторое сито с неизвестным диаметром отверстий, и мы

хотим оценить этот диаметр по размеру прошедших через сито песчинок. В этом случае усечение введено искусственно. При этом получается распределение с неизвестным параметром — правой точкой усечения. Согласно теореме III.1, x_{\max} не зависит от вида функции распределения и является достаточной статистикой для параметра точки усечения.

Рассмотрим другой пример. В миоценовых слоях, образующих обрывы в хребте Иланлы в Туркмении (на юго-восток от сел. Доната), имеются прослои ангидритов. Максимальная неизвестная нам мощность слоя ангидрита не превышает d . Обломки ангидрита об-

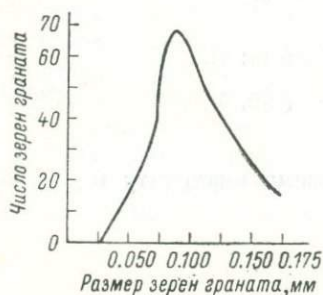


Рис. III.1. Усеченное распределение зерен граната в песке с оз. Мичиган у Эванстоуна (материалы А. Б. Вистелиуса и Е. Дашпlesa, 1975 г.).

валиваются, попадают во впадины близ подножия обрыва и здесь окатываются, образуя валуны. Очевидно, что в этом случае мы имеем плотность $f(x; d)$, где x — диаметр валуна, а d — оцениваемая мощность слоя ангидрита. Эта плотность оказалась усеченной благодаря специфическому природному условию, определяющему размер валуна ангидрита. Для того чтобы выяснить, является ли статистика x_{\max} достаточной, необходимо проведение специальных исследований. Если она окажется достаточной, то геологу нет необходимости для определения θ измерять множество валунов. Нужно обеспечить сбор наблюдений по схеме простого выбора, подсчитать общее число валунов в пробе и измерить один из них — самый большой.

III.3. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Ниже показан круг основных идей и методов, с которыми приходится иметь дело при проверке статистических гипотез. Для достижения большей ясности рассматриваются только важнейшие понятия. В начале дается общая характеристика предмета, затем введенные понятия иллюстрируются и развиваются на материале равномерного распределения, позволяющего показать эти понятия достаточно просто.³ В заключение дается пример — показывается

³ Нужно иметь в виду, что равномерная плотность на сегменте $[a, b]$ нерегулярна, так как у нее отсутствуют производные в точках a и b . В связи с этим некоторые методы точечного оценивания неудобно иллюстрировать на материале этой плотности. Что же касается попытки проверки теории статистических гипотез, то равномерная плотность позволяет их рассмотреть с достаточной полнотой.

техника работы с конкретным материалом, часто встречающимся в геохимических исследованиях (проблема проверки предполагаемого распределения концентраций химического элемента). В качестве исходных данных рассматриваются ранее изученные нами сведения о концентрации Na_2O в базальтах.

III.3.1. Основные идеи и важнейшие определения

Пусть мы имеем некоторую совокупность и хотим проверить предположение относительно ее вероятностных свойств с помощью наблюдений. Такое предположение, заключающееся в гипотезе относительно распределения вероятностей наблюдаемых точек в выборочном пространстве W , называется статистической гипотезой и обычно обозначается через H_i , где i означает конкретный тип гипотезы.

Необходимость построения статистической гипотезы в работе геолога определяется следующим. В процессе исследования накапливаются в основном интуитивные представления о формировании исследуемого объекта (цель работы — узнать, как формировался объект). Возникающие при этом представления в дальнейшем называются геологической аксиоматикой. Такая аксиоматика недостаточно определена для того, чтобы поддаваться точной проверке. Таким образом, возникает необходимость в дальнейшей схематизации. Эта схематизация дает формальное математическое представление о том, каковы должны быть соотношения между наблюдениями, если аксиоматика правильно отражает природную ситуацию. Математическое выражение аксиоматики называется моделью. Модель беднее всей геологической аксиоматики, взятой в целом, но при удачной работе отражает основные черты аксиоматики. Проверить непосредственно непротиворечивость модели наблюдениям часто трудно. Практически можно взять у модели некоторые черты, поддающиеся сопоставлению с наблюдениями, и выяснить, не противоречат ли наблюдения этим чертам. Для стохастической модели эти черты являются статистической гипотезой. Предполагается, естественно, что в гипотезу включены характернейшие особенности модели. Проверять гипотезу по наблюдениям, мы принимаем или отвергаем ее. Если мы отвергли гипотезу, то мы отвергли модель и лежащую в ее основе геологическую аксиоматику. Если мы приняли гипотезу, то это значит, что модель и геологическая аксиоматика в тех чертах, которые охвачены гипотезой, не противоречат наблюдениям.

Заключение о правильности модели и аксиоматики, принятое на основе того, что статистическая гипотеза не забракована, лежит уже вне математики. Все зависит от того, насколько статистическая гипотеза отражает важнейшие черты модели.

Как мы отмечали, изучение выборки дает возможность проверить статистическую гипотезу. Эта проверка может забраковать гипотезу, может привести к ее приемке или может показать, что

данных для того, чтобы принять решение, недостаточно. В процессе проверки гипотезы могут быть сделаны два типа ошибок. *Гипотеза может быть отвергнута, когда она верна. Такая ошибка называется ошибкой первого рода.* Возможен, естественно, и случай приемки ошибочной гипотезы. *Ошибка, заключающаяся в приеме ошибочной гипотезы, носит название ошибки второго рода.*

Пусть мы имеем две статистические гипотезы. При этом одна из них охватывает определенные допущения относительно свойств совокупности, из которой сделана выборка. Обозначим эту гипотезу H_0 и назовем ее проверяемой гипотезой. Пусть вторая гипотеза допускает существование у совокупности иных вероятностных свойств, чем охваченные H_0 . Эта вторая гипотеза обозначается H_1 (Кендалл, Стьюарт, 1973, с. 223). Если H_1 является отрицанием H_0 , то H_1 обычно называют *альтернативной гипотезой*.

При исследовании соответствия наблюдениям H_0 и H_1 возможно, что H_0 и H_1 , вместе взятые, не исчерпывают все ситуации, т. е. H_1 не обязана являться отрицанием H_0 . Так, например, в дальнейшем мы столкнемся с двумя ситуациями при изучении распределения Na_2O в базальтах.

При одной из них H_0 будет утверждать, что распределение изучаемой величины равномерно на $(0; \sqrt{3}]$, в качестве альтернативы рассматриваются все остальные возможности (равномерное распределение на других множествах, либо вообще неравномерные распределения), здесь H_1 является отрицанием H_0 . При другом подходе мы принимаем H_0 такое же, как в первом случае, но в качестве H_1 рассматриваем возможности только равномерных распределений на $(0; \theta]$, где $\theta \neq \sqrt{3}$. В этом случае H_0 и H_1 , вместе взятые, в реальной задаче не исчерпывают всех возможностей, так как нельзя а priori отбросить реальность встречи вообще неравномерных распределений.

После выбора H_0 и H_1 рассматриваются только эти возможности. Так повышается эффективность методов проверки гипотез, когда заведомо верна H_0 или H_1 . Эта методика может оказаться опасной, если ни H_0 , ни H_1 не соответствуют изучаемому явлению.

Во многих случаях распределения вероятностей для H_0 и $\overline{H_1}$ можно охарактеризовать, указав два множества ω_0 и ω_1 значений параметра θ . Параметр принадлежит ω_0 , если мы имеем дело с H_0 , и он принадлежит ω_1 , если имеет место H_1 . В этом случае говорят, что *задача поставлена в параметрическом виде*. Например, если H_0 и H_1 отвечают равномерному распределению на $(0; \theta]$, причем, согласно H_0 , $\theta \in [\theta_1; \theta_2] \equiv \omega_0$, а, согласно H_1 , $\theta \in [\theta_3; \theta_4] \equiv \omega_1$, то вопрос поставлен как параметрическая задача. Если хотя бы одна из гипотез не допускает представления путем задания ω_0 или ω_1 для значений параметра, то *задача называется непараметрической*. Так, если мы проверяем H_0 на соответствие равномерному распределению на $(0; \sqrt{3}]$ против H_1 любого другого распределения, то мы имеем дело с непараметрической задачей. Если же H_0

проверяется на соответствие равномерному распределению на $(0; \sqrt{3}]$ против равномерного же распределения на $(0; \theta]$ с параметром $\theta \neq \sqrt{3}$, то мы имеем дело с параметрическим случаем с двумя множествами параметров — ω_0 , включающим одну параметрическую точку, совпадающую с $\sqrt{3}$, и ω_1 , которое включает все положительные точки, за исключением $\sqrt{3}$.

В параметрической задаче *проверяемая гипотеза H_0 называется простой*, если ω_0 содержит только одну точку (в примере этой точкой был $\sqrt{3}$). Если в ω_0 более одной точки, то *гипотезу называют сложной*. Та же терминология принята для H_1 .

Рассмотрим теперь в общих чертах те операции, которые используются при проверке статистических гипотез.

Разделим выборочное пространство W на две части. Одну его часть w будем называть *критической областью*, а другую определим как $W-w$. Решение о приемке гипотезы производится так. Если выборка $x \in w$, то H_0 отбрасывается. Правило для приемки или браковки H_0 при сравнении ее с H_1 называется *критерием проверки гипотезы H_0 против H_1* . Мы будем рассматривать критерии лишь определенного типа.⁴ Эти критерии заключаются в том, что строят критическую область w , а затем определяют, попала ли туда выборочная точка x . H_0 бракуется в том и только в том случае, когда $x \in w$.

Определение различных критических областей дает возможность использовать критерии, имеющие различные качества. Так, например, для того чтобы критерий был «хорош», нужно, чтобы вероятности

$$p(x \in w | H_0)$$

(вероятность ошибки первого рода)

и

$$p(x \notin w | H_1)$$

(вероятность ошибки второго рода)

были по возможности малы. Добиться того, чтобы вероятности ошибок обоих типов были одновременно сколь угодно малы, невозможно, так как, вообще говоря, уменьшение вероятности одной ошибки обычно сопровождается ростом вероятности другой. Вопрос о том, как выбрать оптимальные значения для обеих ошибок, составляет нетривиальную математическую задачу, имеющую различные решения. В этом очерке мы будем пользоваться в основном *теорией Неймана—Пирсона*. Эта теория в простейшем случае определяет процесс оптимизации, при котором вероятность ошибки какого-либо определенного типа удерживается постоянной, в то

⁴ Ограничимся описанием лишь одного класса критериев, называемых *нерандомизованными*. В нерандомизованных критериях решение принимается исключительно по заранее построенной критической области и выборке. Сведения о других типах критериев можно найти в книге Кедалла и Стьюарта (1973, с. 226).

время как вероятность ошибки другого типа уменьшается настолько, насколько это возможно. Обычно на фиксированном уровне удерживается ошибка первого рода, так как, как правило, H_0 определяет распределение вероятностей более конкретно, чем H_1 , и тем самым удобнее для проверки.

Допустим, что критическая область w построена тем или иным путем. Если проверяется простая гипотеза, то вероятность ошибки первого рода равна некоторому числу α . Если же проверяется сложная гипотеза, то вероятность ошибки первого рода есть функция $\alpha(\theta)$ от параметра θ . В обоих случаях эта вероятность называется *уровнем значимости, основанным на критической области w (уровня α)*.

Выбор численного значения α не является математической задачей и целиком определяется опытом. Практика показывает, что в зависимости от содержания задачи за α обычно принимается одна из точек в сегменте $0.001 \leq \alpha \leq 0.05$. В геологии широко пользуются значениями $\alpha = 0.05$, хотя оно никогда не обосновывалось специальными исследованиями.

Во многих случаях вместо вероятности ошибки второго рода используется дополнение этой вероятности до единицы. Это дополнение называется *функцией мощности* (кратко ее называют просто *мощностью*). При простой альтернативе мощность дается числом β . При сложной альтернативе мощность является функцией от параметра и обозначается $\beta(\theta)$.

При проверке статистической гипотезы наиболее простым является случай, когда проверяется простая гипотеза, состоящая в том, что выборочный вектор x имеет распределение с плотностью p_0 (со значением параметра $\theta = \theta_0$) против простой альтернативы. Наилучшая критическая область w_0 в этом случае может быть построена следующим образом. Прежде всего, в w_0 включаются все точки, в которых $p_0(x) = 0$. Далее в область последовательно включаются точки, для которых $p_1(x)/p_0(x)$ наиболее велика. Этот процесс продолжается до тех пор, пока $\int_{w_0} p_0(x) dx$ не окажется равным α .

Полученная область является наилучшей в следующем смысле. Если мы построим какую-либо другую область w_1 того же уровня α , так что

$$\int_{w_1} p_0(x) dx = \alpha,$$

то тогда мощность нового критерия будет не больше, чем мощность оптимального критерия, т. е. всегда будет удовлетворяться неравенство

$$\int_{w_0} p_1(x) dx \geq \int_{w_1} p_1(x) dx.$$

Если альтернатива или гипотеза или обе они не являются простыми или если не существует непрерывных плотностей, то опти-

мизация этой простой процедурой не гарантируется, и вопрос о выделении наилучшей критической области осложняется. Еще реже существует оптимальное решение при непараметрической постановке задачи. В этом случае обычно приходится прибегать к критериям согласия, которые широко известны в геологии и на которых мы здесь останавливаться не будем.

В распространенном случае проверки простой гипотезы против сложной альтернативы возникает вопрос о выборе наилучшей критической области для этой альтернативы. При этом возможно, что для каждого значения параметра $\theta \in \omega_1$ существует одна и та же наилучшая критическая область w , не зависящая от конкретного значения θ . Такую область называют *равномерно наилучшей критической областью*, а связанный с ней критерий — *равномерно наиболее мощным критерием*. Слово «равномерно» означает, что наибольшая мощность для указанного критерия достигается при каждом $\theta \in \omega_1$. Равномерно наиболее мощные критерии встречаются достаточно редко. В тех случаях, когда получить равномерно наиболее мощный критерий невозможно, прибегают к оптимальным процедурам, которые строятся на иных критериях. Эти критерии для каждого типа процедуры имеют свои названия. Некоторые из этих критериев мы рассмотрим относительно детально, изучая вопрос проверки гипотезы о равномерном распределении.

Следует еще раз подчеркнуть, что в излагаемой теории Неймана—Пирсона неизвестный параметр рассматривается как неизвестная, но постоянная величина.

Положения о приемке и браковке статистических гипотез кажутся весьма абстрактными до тех пор, пока они не разобраны на простом и конкретном материале. После такого разбора становится очевидно, что это важнейшие понятия, которые необходимо использовать в каждодневной работе. В связи с этим ниже дается разбор изложенных понятий при решении вопроса о приемке или браковке гипотезы равномерного распределения. Равномерное распределение имеет большое значение во многих геологических задачах. В то же время анализ ситуации, в которой принимается или бракуется гипотеза равномерного распределения, дает возможность особенно просто продемонстрировать важнейшие статистические понятия.

III.3.2. Построение наилучших критических областей при проверке гипотез о параметре θ на примере равномерного распределения

Пусть имеется равномерно распределенная на $(0; \theta]$ случайная величина. При этом значение θ неизвестно. Относительно значения θ мы будем строить различные гипотезы. Пусть проверяемая гипотеза H_0 ради простоты будет одной и той же, а именно $\theta = \theta_0 = 10$. Альтернативы H_1 будем выбирать разные. Для наглядности

примем, что мы работаем с выборками, содержащими всего два наблюдения, — x_1 и x_2 . При этом для всех рассматриваемых задач мы принимаем уровень значимости $\alpha = 0.25$.

Таким образом, выборочным пространством W в наших примерах всегда является положительный квадрант плоскости x_1, x_2 , а параметрическим пространством Ω — полупрямая $(0; \infty)$. На рис. III.2 изображена часть выборочного пространства x_1, x_2 и

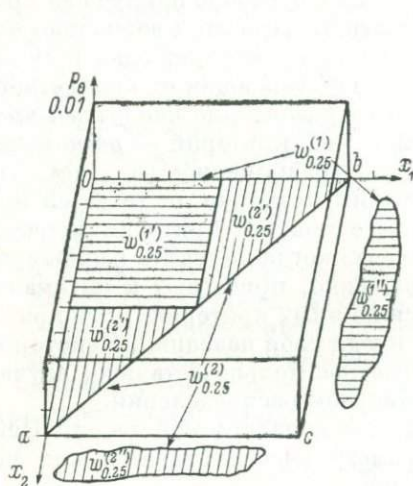


Рис. III.2. Критические области 25%-го уровня для проверки гипотезы $H_0: \theta = 10$.

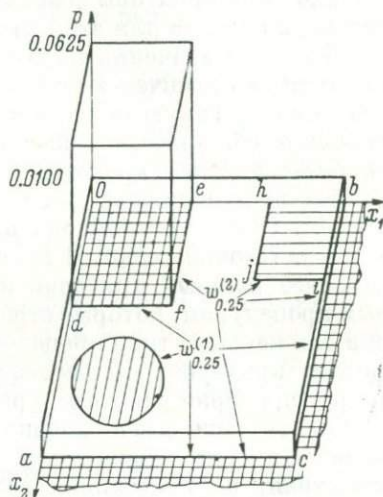


Рис. III.3. Наилучшие критические области 25%-го уровня для проверки гипотезы $H_0: \theta = 10$ против простой альтернативы $H_1: \theta = 4$.

плотность вероятности выборочной точки $x = (x_1, x_2)$ для H_0 . Эта плотность равна

$$p_0(x_1; x_2; \theta = 10) = \begin{cases} \frac{1}{100}, & \text{если } 0 \leq x_1, x_2 \leq 10, \\ 0, & \text{если хотя бы одно из чисел } x_1 \\ & \text{или } x_2 > 10 \text{ либо } < 0, \end{cases}$$

и показана на рис. III.2.

Рассмотрим теперь множество всех возможных критических областей $w_{0.25}$ уровня 0.25 для проверки H_0 . Согласно определению, каждая такая область должна содержать $1/4$ общей площади квадрата $Oacb$. Кроме того каждая такая область может содержать произвольную площадь на выборочном пространстве W вне квадрата $Oacb$. Любой из рассматриваемых участков критической области $w_{0.25}$ может состоять из отдельных частей и даже из разрозненных точек.

Пусть $w_{0.25}^{(1)}$ — первая критическая область, которая состоит из двух частей. Первая часть $w_{0.25}^{(1')}$ находится внутри $Oacb$. Другая

часть $w_{0.25}^{(1)}$, которую мы обозначаем $w_{0.25}^{(1'')}$, находится вне квадрата. Вся область $w_{0.25}^{(1)}$ на рис. III.2 покрыта горизонтальной штриховкой.

Тогда

$$\int_{w_{0.25}^{(1'')}} p_0(x | \theta = 10) dx = \frac{1}{4}$$

и

$$\int_{w_{0.25}^{(1'')}} p_0(x | \theta = 10) dx = 0.$$

Аналогично изображены области $w_{0.25}^{(2)}$, покрытые на рис. III.2 вертикальной штриховкой. Вероятность попадания в критическую область равна сумме приведенных интегралов и равна $1/4$, как это требуется по определению критической области уровня 0.25.

На рис. III.3 показаны две (из возможного бесконечного числа) критические области уровня 0.25. Одна из них покрыта горизонтальной штриховкой и состоит из квадратов $Odfe$ и $hjib$, а также части вне квадрата $Oacb$, указанной стрелками, исходящими от $w_{0.25}^{(2)}$. Вторая критическая область состоит из того же квадрата $Odfe$, круга, равновеликого квадрату $hjib$, и той же внешней области за пределами $Oacb$. Она покрыта вертикальной штриховкой и указана стрелками, исходящими от $w_{0.25}^{(1)}$. Очевидно, что таких областей имеется бесконечное количество. Обозначим их множество как S ($\alpha=0.25$; $\theta_0=10$). Ни одна из этих критических областей не представляет каких-либо преимуществ по сравнению с другой, пока не выбрана альтернатива.

Выберем теперь определенную альтернативу, скажем, $\theta = \theta_1 = 4$. Обе плотности

$$p_1(x | \theta_1 = 4) \text{ и } p_0(x | \theta_0 = 10)$$

показаны на рис. III.3 в виде призм соответственно для p_0 с высотой 0.01, а для p_1 — соответственно с высотой 0.0625.

Выделим теперь из множества критических областей S ($\alpha=0.25$; $\theta_0=10$) те, которые являются наиболее благоприятными при альтернативе $\theta_1=4$. Как отмечалось ранее, оптимальная область для проверки простой гипотезы против простой альтернативы должна содержать такие точки, в которых отношение

$$R = \frac{p_1(x | \theta_1 = 4)}{p_0(x | \theta_0 = 10)}$$

было бы возможно больше. Кроме того, она должна содержать такие точки, в которых $p_0(x | \theta_0 = 10) = 0$, т. е. в нашем случае — все точки за пределами квадрата $Oacb$ на рис. III.3. Наконец, область должна быть критической областью уровня 0.25 при H_0 . Иными словами, эта область должна содержаться в нашем множестве S ($\alpha=0.25$; $\theta_0=10$).

Посмотрим теперь, какое множество точек удовлетворяет отмеченным трем условиям. Из рис. III.3 ясно, что наибольшее отношение R достигается для точек плоскости (X_1, X_2) , лежащих в квадрате $Odfe$. Отберем все эти точки в критическую область. Однако этих точек слишком мало, чтобы уровень области был $1/4$.

Действительно,

$$\int_{\{Odfe\}} p_0(x | \theta_0 = 10) dx = \int_{\{Odfe\}} \frac{1}{100} dx_1 dx_2 = 0.16,$$

тогда как нужно получить 0.25.

Следовательно, мы должны добавить еще интеграл, равный 0.09. Для этого нужно взять дополнительную площадь, равновеликую с 0.3×0.3 . Точки для этого нельзя взять вне квадрата $Oacb$, так как вне его $p_0(x/\theta_0=10)=0$. Таким образом, нужно взять площадь, равную 0.09, где-то внутри $Oacb$, но вне $Odfe$. Положение этой площади может быть произвольное.

Таким образом, наилучшая критическая область для проверки принятых H_0 и H_1 состоит из трех участков — квадрата $Odfe$, области площадью 0.09, располагающейся между контурами dfe и acb , и всех точек вне контура $Oacb$. Для нашей задачи имеется бесконечное число таких наилучших критических областей, каждая из которых так же оптимальна, как любая другая.

Обозначим множество таких оптимальных критических областей через S ($\alpha=0.25$; $\theta_0=10$; $\theta_1=4$). Очевидно, что S ($\alpha=0.25$; $\theta_0=10$) $\supset S$ ($\alpha=0.25$; $\theta_0=10$; $\theta_1=4$). Две из бесконечного числа лучших критических областей показаны на рис. III.3. В каждую из этих областей входят два общих участка, покрытых пересекающейся штриховкой. Первый из этих участков — квадрат $Odfe$, второй — вся площадь вне квадрата $Oacb$. Участки, принадлежащие только к $w_{0.25}^{(1)}$ или только к $w_{0.25}^{(2)}$, отмечены горизонтальной или вертикальной штриховкой. Подчеркнем еще раз, что эти участки должны иметь фиксированную площадь, но их форма и положение внутри контура $adfeb$ произвольны. Из сказанного ясно, что для $\alpha=0.25$, $\theta_0=10$ и $\theta_1=4$ не существует единственного наилучшего критерия — их бесконечно много и каждый из них так же хорош, как и все остальные.

Обратимся теперь к альтернативе $\theta_1=8$. На рис. III.4 показаны плотности

$$p_0(x | \theta_0 = 10) \text{ и } p_1(x | \theta_1 = 8)$$

в виде призм с высотами соответственно 0.01 для p_0 и 0.0156 — для p_1 .

Рассмотрим теперь множество лучших критических областей S ($\alpha=0.25$; $\theta_0=10$; $\theta_1=8$). Ясно, что точки, при которых R получает наибольшее значение, находятся внутри квадрата $Ojih$. Однако выделить всех их в критическую область мы не можем, так как их

слишком много. Действительно, если бы мы выделили их все, то получили бы

$$\int_{\{0gih\}} p_0(x | \theta_0 = 10) dx = \int_{\{0gih\}} \frac{1}{100} dx_1 dx_2 = 0.64,$$

а нам требуется только 0.25.

Таким образом, мы можем взять произвольную область площадью 0.25 в произвольном положении и произвольной формы внутри квадрата $Ogih$. Кроме того критическая область должна

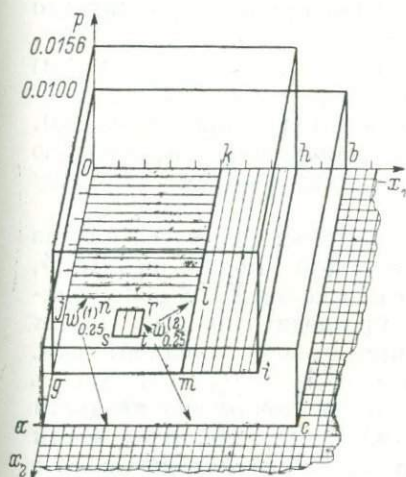


Рис. III.4. Наилучшие критические области 25%-го уровня для проверки гипотезы $H_0: \theta=10$ против простой альтернативы $H_1: \theta=8$.

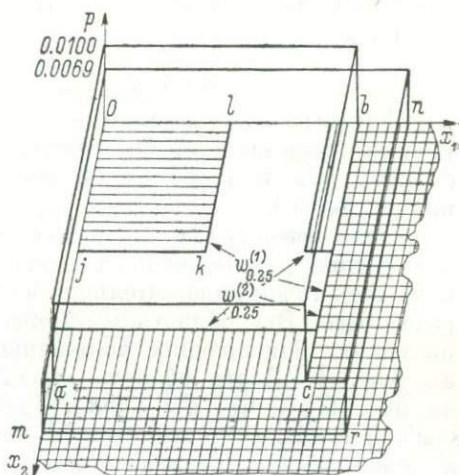


Рис. III.5. Наилучшие критические области 25%-го уровня для проверки гипотезы $H_0: \theta=10$ против сложной альтернативы $H_1: \theta > 10$.

включать все точки вне квадрата $Oacb$. Так как нужных нам точек внутри $Ogih$ и так больше, чем требуется для построения оптимальной области, то площадь между $Ogih$ и $Oacb$ не может включаться в наилучшую область. На рис. III.4 разными штриховками показаны две из бесчисленного количества возможных наилучших критических областей из множества S ($\alpha=0.25$; $\theta_0=10$; $\theta_1=8$). Одна из них $w_{0.25}^{(1)}$, покрытая горизонтальной штриховкой, состоит из квадрата $Ojlk$ и области вне квадрата $Oacb$. Другая $w_{0.25}^{(2)}$, покрытая вертикальной штриховкой, образована прямоугольником $kmih$, квадратом $nstr$ и точками вне квадрата $Oacb$.

Очевидно, что, действуя подобным образом, для каждой простой гипотезы $\theta=\theta_0$ и каждой простой альтернативы $\theta=\theta_1$ мы можем построить бесконечное множество лучших критических областей S ($\alpha=0.25$; $\theta_0=10$; θ_1).

Рассмотрим теперь сложную левостороннюю альтернативу в целом. До сих пор мы выдвигали в качестве альтернатив отдельные, конкретные значения θ (например, $\theta_1=4$, $\theta_1=8$). Теперь мы рассмотрим всю совокупность левых альтернатив, т. е. неопределенное значение θ_1 , меняющееся в интервале $(0; 10)$. Для каждого отдельного значения θ_1 из этого интервала мы имеем, как отмечалось, бесконечное множество наилучших критических областей S ($\alpha=0.25$; $\theta_0=10$; θ_1). Зададимся теперь вопросом: нет ли среди всех этих множеств областей такой, которая содержалась бы в каждом из этих множеств (т. е. имеется ли непустое пересечение всех этих множеств)? В теоретико-множественных терминах это обозначается как

$$\bigcap_{\theta_1 \in (0; 10]} S(\alpha=0.25; \theta_0=10; \theta_1). \quad (\text{III. 3. 1})$$

Если бы нашлась такая общая область (одна или несколько), то она представляла бы равномерно наилучшую критическую область для проверки всех левосторонних альтернатив сразу против $\theta_0=10$.

Легко заметить, что такая область существует и состоит из всех точек, расположенных внутри квадрата $Ojlk$ на рис. III.4, и из всех точек положительной части плоскости X_1, X_2 вне квадрата $Oacb$. Эту область мы можем проследить на всех наших рисунках, поясняющих положение критических областей. Так, на рис. III.2 эта область обозначена как $w_{0.25}^{(1)}$, если учесть на рис. III.2, что мы можем брать любые точки вне квадрата $Oacb$. На рис. III.3 эта область получилась бы, если бы мы взяли выделенный нами квадрат $Odfe$ и добавили бы к нему недостающую площадь 0.3×0.3 , наращивая каждую из сторон квадрата на одно деление. Легко понять, что указанная область фигурировала бы также на любом другом рисунке, изображающем соотношения при простой альтернативе для $\theta_1 < 10$. Нетрудно также убедиться, что имеется только одна область, входящая в пересечение (III.3.1).

Таким образом, для сложной левосторонней альтернативы существует равномерно наиболее мощный критерий (надо помнить, что для простоты и наглядности мы выбрали $\alpha=0.25$ при выборке с $n=2$). Гипотеза $\theta=10$ бракуется в том и только в том случае, когда выборочная точка (x_1, x_2) попадает в квадрат $Ojlk$ или за пределы квадрата $Oacb$.

Рассмотрим теперь вопрос о попадании точки в область вне квадрата $Oacb$. Точки вне этого квадрата включены в критическую область на том основании, что вероятность попадания выборочной точки в эту область при H_0 равна нулю.

Таким образом, сколько бы точек из этой «внешней» области мы ни включили бы в критическую область, это не скажется на уровне критической области. Иными словами, включение этих точек в критическую область не увеличивает ошибки первого рода. Если бы в этих точках была ненулевая плотность при H_1 , то их включение в критическую область увеличивало бы мощность. Однако в данном

случае, при левосторонних альтернативах, плотность здесь будет нулевой и при H_1 .

Следовательно, включение этих точек не меняет ни ошибки первого, ни ошибки второго рода. Включаются же они для того, чтобы придать критической области наиболее общий вид. Преимущество этого подхода выяснится тогда, когда мы включим в рассмотрение правосторонние альтернативы и будем рассматривать задачи, охватывающие также положение неизвестного параметра θ вправо от точки 10.

Рассмотрим теперь какую-нибудь простую правостороннюю альтернативу, например $\theta_1=12$. На рис. III.5 соответствующими призмами показаны плотности

$$p_0(x | \theta_0 = 10) \text{ и } p_1(x | \theta_1 = 12).$$

Наилучшая критическая область, как это неоднократно отмечалось, должна включать точки, для которых $p_0(x | \theta_0 = 10) = 0$, т. е. те же точки вне квадрата $Oacb$, о которых мы уже говорили. Однако теперь для части этих точек, находящихся между квадратами $Oacb$ и $Omrn$, $p_1(x | \theta_1 = 12) > 0$. Таким образом, включив эти точки, мы повышаем мощность, не изменяя уровня значимости.

Рассмотрим теперь точки внутри квадрата $Oacb$. Для всех этих точек отношение

$$R = \frac{p_1(x | \theta_1 = 12)}{p_0(x | \theta_0 = 10)}$$

постоянно (оно равно $36/25$). Из сказанного видно, что безразлично, каким способом мы отбираем точки из квадрата $Oacb$ в критическую область, нужно только, чтобы площадь, охваченная этими точками, была равна 0.25. Две наилучшие (из бесконечного количества возможных) критические области $w_{0.25}^{(1)}$ и $w_{0.25}^{(2)}$ показаны на рис. III.5. Первая область $w_{0.25}^{(1)}$, покрытая горизонтальной штриховкой, образована квадратом $Ojkl$ и областью вне квадрата $Oabc$. Вторая критическая область $w_{0.25}^{(2)}$ показана вертикальной штриховкой, она включает два участка внутри квадрата $Oacb$ и внешние точки по отношению к этому квадрату.

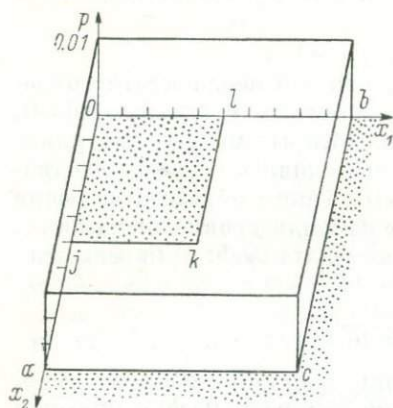
Множество всех наилучших критических областей, очевидно, бесконечно. Мы его обозначим S ($\alpha=0.25$; $\theta_0=10$; $\theta_1=12$). Заметим теперь, что при построении областей, содержащихся во множестве S ($\alpha=0.25$; $\theta_0=10$; $\theta_1=12$), мы не использовали специфику, возникающую от того, что $\theta_1=12$. Действительно, мы берем все точки, внешние по отношению к квадрату $Oacb$, и прибавляем к ним произвольно расположенную область внутри этого квадрата площадью 0.25. Если бы мы взяли вместо $\theta_1=12$ значение $\theta_1=14$, то это ничего бы не изменило. Таким образом,

$$S(\alpha=0.25; \theta_0=10; \theta_1=12) = S(\alpha=0.25; \theta_0=10; \theta_1=14) = S(\alpha=0.25; \theta_0=10; \theta_1)$$

для любого $\theta_1 > 10$.

Следовательно, множество S ($\alpha=0.25$; $\theta_0=10$; θ_1) относится вообще ко всем правосторонним альтернативам. Из этого следует,

что для правосторонних альтернатив существует бесконечное множество наилучших критических областей и бесконечное множество равномерно наиболее мощных критериев. Так, например, мы можем использовать оптимальный критерий, заключающийся в том, что $H_0 : \theta_0 = 10$ бракуется тогда и только тогда, когда выборочная точка попала в область, покрытую на рис. III.5 вертикальной штриховкой. Столь же оптимальным в теории Неймана—Пирсона будет критерий, построенный на области, указанной на рис. III.5 горизонтальной штриховкой. В этом отличие правосторонней сложной альтернативы от левосторонней, при которой имелось единственное оптимальное решение.



Рассмотрим теперь задачу с двусторонней сложной альтернативой, т. е.

$$H_0 : \theta = \theta_0 = 10 \text{ против } H_1 : \theta = \theta_1 \neq 10 \quad (\theta_1 > 0).$$

Рис. III.6. Критическая область 25%-го уровня для равномерного наиболее мощного критерия проверки гипотезы $H_0 : \theta = 10$ против двусторонних альтернатив $H_1 : \theta \neq 10$.

Существование в этом случае единственного равномерно наиболее мощного критерия доказывается весьма просто. Действительно, область $w_{0.25}^{(1)}$ на рис. III.5 была равномерно наилучшей критической областью для левосторонней сложной альтернативы. Так как эта область содержится среди лучших областей для правосторонней сложной альтернативы, то на ней основан равномерно наиболее мощный критерий и для двусторонних альтернатив.

Итак, для задачи

$$H_0 : \theta = \theta_0 = 10 \text{ против } H_1 : \theta = \theta_1 \neq 10 \quad (\theta > 0)$$

существует единственный равномерно наиболее мощный критерий уровня $\alpha = 0.25$ при выборке объема $n = 2$ (два наблюдения), отвергающий H_0 тогда и только тогда, когда либо оба наших наблюдения попадают в квадрат со сторонами $[0; 5]$, либо когда хотя бы одно из наблюдений попадает в область вне квадрата со сторонами $[0; 10]$ (рис. III.6).

Выбранные нами из соображений удобства $\alpha = 0.25$ и $n = 2$ не имеют никакой специфики. Изложенные в этом примере результаты обобщаются на любые $0 < \alpha < 1$, любое целое n и любое $\theta_0 > 0$.

Итак, для задачи

$$H_0 : \theta = \theta_0 \text{ против } H_1 : \theta \neq \theta_0 \quad (\theta > 0)$$

существует единственный равномерно наиболее мощный критерий уровня α для объема выборки n , отвергающий H_0 в том и только том случае, когда либо точка (x_1, x_2, \dots, x_n) попадает в область $\{0 \leq x_i \leq \alpha^{1/n} \theta_0\}$ ($i=1, 2, \dots, n$), либо когда хотя бы одно из наблюдений x_i окажется $> \theta_0$.

III.3.3. О некоторых важных характеристиках статистического критерия (продолжение исследования гипотезы равномерного распределения)

Напомним, что при точечном оценивании параметра мы рассмотрели такие характеристики, как достаточность, состоятельность, несмещенность и асимптотическая несмещенность. Характеристики с теми же названиями имеются и для критериев. Так, говорят о достаточных статистиках критерия, состоятельных критериях, несмещенных и асимптотически несмещенных критериях. Подобные понятия, связанные с критериями, близки к соответствующим понятиям, отнесенным к статистикам при точечном оценивании, но не тождественны им. Ниже мы рассмотрим эти понятия, иллюстрируя их на примере введенного ранее равномерного распределения.

Рассмотрим полученный нами для равномерного распределения равномерно наиболее мощный критерий, охарактеризованный на с. 190, 191.

Как мы видели, H_0 бракуется в одном из двух случаев:

$$\left. \begin{array}{l} \text{если для всех } x_i \text{ из } (x_1, x_2, \dots, x_n) \text{ выполняется} \\ 0 \leq x_i \leq \alpha^{1/n} \theta_0 \\ \text{или} \\ \text{если хотя бы одно из } x_i \text{ удовлетворяет неравен-} \\ \text{ству } x_i > \theta_0 \text{ (} i=1, 2, \dots, n \text{).} \end{array} \right\} \quad (\text{III. 3. 2})$$

Первый из рассмотренных случаев можно записать в эквивалентной формулировке как

$$x_{\max} \in [0; \alpha^{1/n} \theta_0],$$

второй как

$$x_{\max} > \theta_0.$$

Таким образом, решение о приемке или браковке H_0 принимается исключительно на основании статистики x_{\max} и нет необходимости знать всю выборку x_1, x_2, \dots, x_n .

Статистика размерности меньше чем n (объем выборки) называется нетривиальной достаточной статистикой критерия, если решение относительно H_0 может быть принято исключительно на основании выборочного значения этой статистики. Вся выборка целиком, очевидно, является достаточной статистикой размерности n и называется тривиальной достаточной статистикой критерия.

Как видим, в случае изученного равномерно наиболее мощного критерия статистика x_{\max} является одномерной достаточной статистикой критерия. Сравнивая приведенные выше определения достаточной статистики и достаточной статистики критерия, можно убедиться в большом сходстве этих понятий. Так, на с. 175—177 мы показали, что x_{\max} является достаточной статистикой параметра θ в равномерном распределении. В то же время x_{\max} оказалось достаточной статистикой критерия для гипотезы о том же параметре. Этот случай очень типичен. Примеры, когда достаточная статистика параметра не является достаточной статистикой критерия для гипотезы о том же параметре или когда достаточная статистика критерия не является достаточной статистикой параметра, могут быть найдены, например, в курсе Кендалла и Стьюарта (1973, § 22.20—22.23).

Понятия о состоятельности и несмещенности связаны с понятием о мощности критерия. В связи с этим рассмотрим несколько подробнее понятие «мощность».

На с. 182 *мощностью критерия мы назвали дополнение до единицы вероятности ошибки второго рода в предположении, что H_0 вместе с H_1 исчерпывают все возможные ситуации*. Согласно другому равноценному определению, более удобному для практического использования, *мощностью называется вероятность попадания выборочной точки x в критическую область в предположении о справедливости H_1* . Проиллюстрируем вычисление мощности на примере равномерного распределения.

Мощность полученного равномерно наиболее мощного критерия (см. п. III. 3. 2)

$$\beta(\theta) = \begin{cases} 1, & \text{если } \theta < \alpha^{1/n} \theta_0, \\ \alpha \cdot \left(\frac{\theta_0}{\theta}\right)^{1/n}, & \text{если } \alpha^{1/n} \theta_0 \leq \theta < \theta_0, \\ 1 - (1 - \alpha) \left(\frac{\theta_0}{\theta}\right)^n, & \text{если } \theta > \theta_0. \end{cases} \quad (\text{III. 3. 3})$$

Чтобы вывести указанные выражения для мощности, предположим сначала, что мы имеем дело с первым определением в (III. 3. 3), т. е.

$$\theta = \theta_1 < \alpha^{1/n} \theta_0.$$

В этом случае все наблюдения x_1, x_2, \dots, x_n должны быть в $[0; \theta_1]$. Вероятность этого события равна единице.

Рассмотрим второе определение из (III. 3. 3), т. е.

$$\theta = \theta_2 \text{ и } \alpha^{1/n} \theta_0 \leq \theta_2 < \theta_0.$$

При этом x_{\max} может находиться в сегменте $[0; \theta_2]$, но благоприятными для попадания в критическую область являются значения x_{\max} в $[0; \alpha^{1/n} \theta_0]$. Вероятность попадания x_{\max} в этот сегмент равна

$$\left(\frac{\alpha^{1/n} \theta_0}{\theta}\right)^n = \alpha \cdot \left(\frac{\theta_0}{\theta}\right)^n, \quad (\text{III. 3. 4})$$

т. е. мы получили определение мощности на втором интервале в (III.3.3).

Возьмем теперь третье положение θ в (III.3.3):

$$\theta = \theta_3 > \theta_0.$$

В этом случае благоприятными для попадания в критическую область являются положения x_{\max} в

$$(\theta_0; \theta_3] \text{ и } [0; \alpha^{1/n}\theta_0],$$

а неблагоприятными для попадания в критическую область являются все остальные положения x_{\max} внутри $[0; \theta]$.

Вычислим вероятность попадания x_{\max} в $(\theta_0; \theta_3]$:

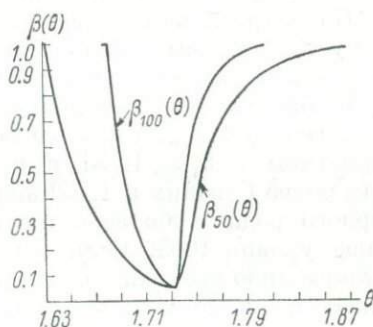
$$P(\theta_0 < x_{\max} \leq \theta_3) = P(x_{\max} \leq \theta_3) - P(x_{\max} \leq \theta_0) = \frac{\theta_3^n}{\theta^n} - \frac{\theta_0^n}{\theta^n} = 1 - \left(\frac{\theta_0}{\theta}\right)^n.$$

Далее вероятность попадания x_{\max} в сегмент $[0; \alpha^{1/n}\theta_0]$ уже вычислена нами в (III.3.4) и равна

$$\alpha \cdot \left(\frac{\theta_0}{\theta}\right)^n.$$

Складывая вероятности этих несовместных событий, получаем

Рис. III.7. Функция мощности $\beta(\theta)$ равномерно наиболее мощного критерия при объемах выборки $n=50$ и 100 .



мощность для третьего определения в (III.3.3), т. е.

$$\beta(\theta) = 1 - (1 - \alpha) \left(\frac{\theta_0}{\theta}\right)^n.$$

Следует отметить, что мощность всегда является функцией трех аргументов — n , α и θ (где θ может быть векторное).

На рис. III.7 приведены графики $\beta(\theta)$ для $\alpha = 0.05$ при $\theta_0 \approx \sqrt{3}$. Соответствующие расчеты приведены в табл. III.4.

Т а б л и ц а III.4

Расчет мощностей при различном числе наблюдений

Число наблюдений, n	50	θ	< 1.6314	1.632	1.635	1.640	1.650	1.700	1.7321
		$\beta_{50}(\theta)$	1.000	0.981	0.895	0.768	0.567	0.127	0.050
100	θ	< 1.681	1.682	1.685	1.700	1.720	1.7321	1.740	
		$\beta_{100}(\theta)$	1.000	0.941	0.788	0.325	0.101	0.050	0.397

Таблица III.4 (продолжение)

Число наблюдений, n	50	θ	1.733	1.74	1.76	1.78	1.80	1.82	1.85
		$\beta_{50}(\theta)$	0.074	0.243	0.573	0.756	0.867	0.920	0.965
	100	θ	1.760	1.780	1.800				
		$\beta_{100}(\theta)$	0.808	0.938	0.980				

На рис. III.7 видно характерное поведение функции мощности, которое должно иметь место у всех «хороших» критериев. У таких критериев мощности растут с увеличением объема выборки; они увеличиваются по мере удаления альтернативного значения $\theta = \theta_1$ от проверяемого $\theta = \theta_0$, т. е. мощность растет вместе с увеличением абсолютного значения разности $|\theta - \theta_0|$. При этом рост должен иметь место как при удалении θ влево, так и вправо от θ_0 .

Посмотрим, какие еще выводы можно сделать из рис. III.7. Допустим, что мы имеем 50 наблюдений и хотим принимать решения так, чтобы вероятности ошибок обоих родов были не более чем 0.05 при проверке гипотезы $\theta_0 = 1.732$ против двусторонних альтернатив $\theta \neq 1.732$. Тогда мы сможем отличать от 1.732 левые альтернативы с $\theta_1 \leq 1.634$ и правые альтернативы с $\theta_1 > 1.836$. Для более близких к 1.732 значений θ_1 либо вероятность ошибки первого рода, либо вероятность ошибки второго рода окажется выше уровня 0.05. Если мы не хотим этого допускать, то мы должны либо увеличить объем выборки, либо увеличить значение α (скажем, вместо $\alpha = 0.05$ брать $\alpha = 0.10$). Чтобы сориентироваться, насколько нам поможет увеличение выборки, на рис. III.7 приведен график $\beta(\theta)$ для $n = 100$. Потребуем снова, чтобы вероятности ошибок обоих родов были ≤ 0.05 . График показывает, что в этом случае мы сможем отличить от $\theta_0 = 1.732$ — с теми же вероятностями ошибок обоих родов — левые альтернативы при $\theta_1 \leq 1.682$ и правые альтернативы при $\theta_1 \geq 1.782$.

Таким образом, область эффективной работы критерия значительно расширилась.

Из рис. III.7 видно, что мощность быстро растет с увеличением объема выборки n . В зависимости от предельного поведения мощности при $n \rightarrow \infty$ различают состоятельные и несостоятельные критерии. Критерий называется состоятельным, если для его мощности мы имеем $\lim_{n \rightarrow \infty} \beta(\theta) = 1$, и несостоятельным, если тот же

предел не стремится к единице. Состоятельность критерия означает, что при проверке гипотезы о параметре $\theta = \theta_0$ против $\theta = \theta_1$, как бы близко ни располагалось θ_1 к θ_0 , ошибка второго рода может быть сделана сколь угодно малой (при фиксированной ошибке первого рода), если будет взято достаточно большое число наблюдений.

Близость понятия «состоятельность критерия» к понятию «состоятельность оценки» очевидна. Между этими понятиями суще-

ствует следующее соотношение. Если критерий, касающийся параметра θ , основывается на статистике, являющейся состоятельной точечной оценкой θ , то и критерий является состоятельным. Обратное утверждение неверно — критерий может быть состоятельным и в том случае, когда он основывается на несостоятельной оценке параметра.

Покажем теперь, что наш двухсторонний критерий для равномерного распределения (III.3.3) состоятелен. Мощность этого критерия имеет три области задания. Рассмотрим их при $n \rightarrow \infty$. Условие для первой области:

$$0 < \lim_{n \rightarrow \infty} \alpha^{1/n} \theta_0, \text{ т. е. } \theta < \theta_0.$$

Второе условие:

$$\lim_{n \rightarrow \infty} \alpha^{1/n} \theta_0 \leq \theta < \theta_0,$$

откуда

$$\theta_0 \leq \theta < \theta_0.$$

Таким образом, вторая область задания пуста (отсутствует) при $n \rightarrow \infty$. Сохраняются две области задания: $\theta < \theta_0$ (первая) и $\theta > \theta_0$ (третья).

Согласно определению $\beta(\theta)$ в первом и третьем условиях, в (III.3.3) мы сразу видим, что предел равен единице и, таким образом, критерий состоятелен.

Другой важной характеристикой критерия является его несмещенность. Критерий называется несмещенным, если его мощность нигде не падает ниже уровня значимости. Итак, для несмещенности требуется, чтобы выполнялось неравенство

$$\beta(\theta) \geq \alpha$$

(при всех значениях θ , рассматриваемых альтернативной гипотезой).

Если проверяемая гипотеза также является сложной и уровень значимости представляет $\alpha(\theta)$, то несмещенность критерия определяется неравенством $\beta(\theta'') \geq \alpha(\theta')$, справедливым для всех θ'' , принадлежащих значениям параметра в альтернативной гипотезе, и всех θ' , принадлежащих значениям параметра в проверяемой гипотезе. Напомним, что точечную оценку мы называли несмещенной, если ее математическое ожидание совпадало с истинным значением параметра. Понятие несмещенности критерия отнесено к вопросу о том, может ли мощность падать ниже уровня значимости; таким образом, казалось бы, ничего общего между этими конструкциями нет. Однако более подробное изучение этих понятий обнаруживает их близость. Действительно, допустим, что некоторый критерий является смещенным. Тогда, согласно определению, найдутся такие значения параметра θ_1 и θ_0 , для которых $\beta(\theta_1) < \alpha(\theta_0)$. Это неравенство можно записать в эквивалентной форме как

$$1 - \beta(\theta_1) > 1 - \alpha(\theta_0). \quad (\text{III. 3. 5})$$

Допустим, что мы проверяем простую гипотезу $H_0: \theta = \theta_0$ против простой альтернативы $H_1: \theta = \theta_1$.

Неравенство (III.3.5) имеет следующую интерпретацию. $1 - \beta(\theta_1)$ означает вероятность непопадания выборочной точки в критическую область, когда истинное значение параметра есть θ_1 . Иными словами, это есть вероятность принятия за истинное значение θ_0 , когда справедливо θ_1 . Правая часть (III.3.5) означает вероятность принятия θ_0 в том случае, когда истинное значение параметра равно θ_0 . Таким образом, неравенство (III.3.5) имеет следующий смысл: вероятность принятия θ_0 при справедливости θ_1 больше вероятности принятия θ_0 при справедливости θ_0 .

Итак, при смещенности критерия мы будем принимать в среднем гипотезу реже, когда она правильна, чем в том случае, когда она неправильна. Таким образом, мы будем делать систематическую ошибку в принятии решения. Такую же картину мы имеем при смещенном точечном оценивании, когда пользуемся смещенной оценкой.

Очевидно, что смещенность оказывается особенно опасной, если значение θ_1 , при котором мощность падает ниже уровня значимости, представляет для нас специальный интерес или если это значение близко к проверяемому θ_0 . Этот последний случай часто встречается на практике.

Если критерий имеет смещение, исчезающее при $n \rightarrow \infty$, то такой критерий называется асимптотически несмещенным. Асимптотически несмещенный критерий мы рассмотрим, исследуя критерий отношения правдоподобия.

Критерий, основанный на несмещенной точечной оценке, как это будет показано далее, может оказаться смещенным. В то же время критерий, основанный на смещенной точечной оценке, может оказаться несмещенным. Проиллюстрируем сказанное.

Покажем сначала, что изучаемый двусторонний критерий является несмещенным.

Действительно, если в (III.3.3) $\theta < \alpha^{1/n} \theta_0$, то $\beta(\theta) = 1 > \alpha$, т. е. здесь нет смещения мощности; если $\alpha^{1/n} \theta_0 \leq \theta < \theta_0$, то функция мощности $\alpha \cdot (\theta_0/\theta)^{1/n}$ с ростом θ монотонно убывает.

Таким образом, самое малое $\beta(\theta)$ будет при $\theta = \theta_0$, и это значение равно α . Итак, и при втором определении в (III.3.3) мощность не опускается ниже α . Наконец, при третьем определении, когда $\theta > \theta_0$, мощность является монотонной по θ и ее наименьшее значение достигается при $\theta = \theta_0$. Подставляя $\theta = \theta_0$ в третье определение (III.3.3), получим $\beta(\theta) = \alpha$. Иными словами, $\beta(\theta)$ всегда больше или равно α , и наш критерий является несмещенным. Однако этот критерий основан на статистике x_{\max} , которая является смещенной точечной оценкой параметра θ .

Отметим в заключение, что, как известно (Кендалл, Стьюарт, 1973), любой равномерно наиболее мощный критерий (а наш критерий является таковым) одновременно обладает свойствами со-

стоятельности и несмещенности, поэтому доказывать наличие этих свойств у равномерно наиболее мощного критерия излишне. Мы сделали это для иллюстрации понятий состоятельности и несмещенности.

III.3.4. Дальнейшее развитие теории проверки гипотез — сложная гипотеза против сложной альтернативы

До сих пор в качестве наиболее общей задачи, связанной с равномерным распределением, мы рассматривали простую гипотезу $H_0: \theta = \theta_0$ против сложной альтернативы $H_1: \theta \neq \theta_0$. Такая постановка задачи вызывает ряд затруднений при интерпретации, на которых мы и остановимся.

Пусть, например, мы точно не знаем проверяемого значения θ_0 , но из априорных соображений предполагаем, что θ_0 близко к $\sqrt{3}$, а фактически оно равно $\sqrt{3}$. Кроме того, предположим, что мы можем сделать неограниченное число наблюдений.

Возьмем сначала $\theta_0 = 1.73$. В силу того что критерий состоятелен, эта гипотеза практически всегда будет забракована, если мы возьмем достаточно большое число наблюдений (так как $\sqrt{3} \neq 1.73$).

Испытаем далее гипотезу $\theta_0 = 1.732$ (более точное приближение к $\sqrt{3}$). Снова при достаточно большой выборке эта гипотеза будет забракована. И так будет до бесконечности.

Таким образом, если мы работаем с очень большими выборками, мы скорее всего отвергнем правильное решение $\theta_0 = \sqrt{3}$, а примем его при пользовании выборками умеренного объема.

В итоге оказывается, что критерий лучше работает с выборками умеренного, а не большого объема. Этот эффект называется парадоксом Берксона. Чтобы обойти указанное явление, иногда рекомендуется (Кендалл, Стьюарт, 1973, с. 246) одновременно с увеличением объема выборки n уменьшать уровень значимости α . Так, например, если $n=100$, то брать $\alpha=0.05$, а при $n=1000$ брать $\alpha=0.01$. Против такой рекомендации имеются два возражения. Прежде всего, стремление α к нулю при росте n еще не обеспечивает устранения отмеченного эффекта. Нужно, чтобы это стремление происходило с достаточной скоростью, а это можно определить только с помощью специальных исследований критерия. Кроме того, работая с переменным α , мы теряем определенность суждения о вероятности ошибки первого рода.

Все эти нежелательные явления отпадут, если допустить, что θ_0 в проверяемой гипотезе может принимать множество значений. Например, что θ_0 может принадлежать некоторому сегменту $[\theta_1, \theta_2]$. Тогда мы получаем сложную гипотезу H_0 против сложной альтернативы. Дальнейшее изложение посвящено этой задаче. В ней мы проверяем $H_0: \theta \in [\theta_1, \theta_2]$ против $H_1: \theta < \theta_1$ или $\theta > \theta_2$.

Разобьем соответственно параметрическое пространство Ω на области $\omega_0 = [\theta_1; \theta_2]$ (допустимые значения параметра при H_0) и $\omega_1 = (-\infty; \theta_1) \cup (\theta_2; \infty)$ (допустимые значения параметра при H_1). Как уже отмечалось, мы имеем теперь две функции, характеризующие критерий: $\alpha(\theta)$, определенную для значения $\theta \in \omega_0$, называемую *уровнем значимости критерия*, и $\beta(\theta)$, определенную для значений $\theta \in \omega_1$, называемую *мощностью*.

Напомним, что в теории Неймана—Пирсона мы ставили задачу оптимизации критерия следующим образом — при заданном уровне значимости достичь максимальной мощности. В данной постановке вопроса такая оптимизация невозможна, так как уровень значимости не задан числом. Рассмотрим применяемые в этом случае подходы.

Прежде всего вместо понятия об уровне значимости α вводится понятие гарантированного уровня значимости α . Говорят, что *критерий имеет гарантированный уровень значимости α , если $\alpha(\theta) \leq \alpha$ для всех $\theta \in \omega_0$* (т. е. функция уровня значимости не поднимается выше некоторого заданного числа α). Будем рассматривать только критерии гарантированного уровня значимости. Если среди них имеется равномерно наиболее мощный, то он, так же как в случае простой H_0 , является одновременно несмещенным и состоятельным. Именно этот критерий признается оптимальным. Такой критерий в случае сложной H_0 встречается исключительно редко, поэтому для оптимизации применяют обычно следующий подход. Сначала класс рассматриваемых критериев сужается. При этом сужение производится по какому-либо полезному для критерия свойству или важному для данной задачи признаку. Так, например, желательным для критерия качеством является несмещенность. Поэтому сузим класс рассматриваемых критериев до одних несмещенных критериев. Среди этого суженного класса может оказаться равномерно наиболее мощный критерий. Если такой критерий существует, то он называется *«равномерно наиболее мощным несмещенным»*. Выделенный в кавычках термин часто понимается неправильно — как указание на то, что критерий является равномерно наиболее мощным и несмещенным, тогда как в действительности это равномерно наиболее мощный критерий среди несмещенных критериев.

Иногда встречаются задачи, в которых важно поддерживать ошибку первого рода постоянной. Тогда класс рассматриваемых критериев сужается лишь до таких критериев, у которых $\alpha(\theta) = \alpha_{\text{const}}$. Такие критерии называются *подобными уровня α* . Если среди них встречается равномерно наиболее мощный, то он называется равномерно наиболее мощным подобным.

Существует также целый ряд других признаков, по которым производится сужение класса критериев, но на них мы останавливаться не будем.

Допустим теперь, что мы сузили класс критериев гарантированного уровня α и нашли среди них наиболее мощный. Этот

критерий, очевидно, обладает наибольшей мощностью внутри суженного класса, но он не обязан обладать наиболее выгодным уровнем значимости внутри этого класса, поскольку известно только, что уровень значимости не превосходит числа α . В том же классе могут находиться критерии с более низкими на отдельных участках функциями уровня. Если все такие критерии имеют в каких-либо точках мощность, меньшую чем у равномерно наиболее мощного критерия, то этот равномерно наиболее мощный критерий называется *строгим равномерно наиболее мощным*.

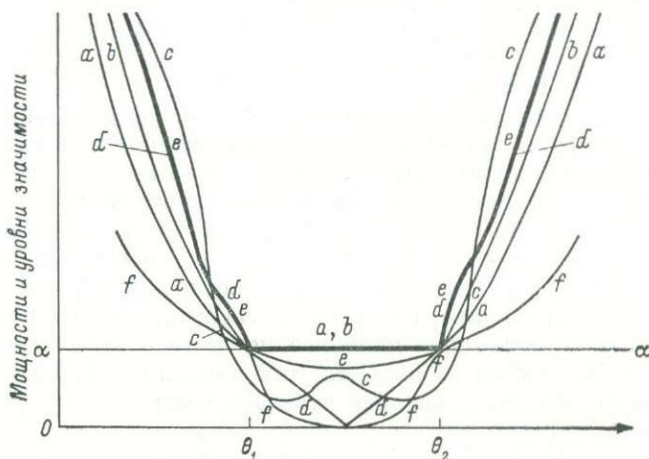


Рис. III.8. Кривые, соответствующие различным методам оптимизации критерия.

a, b, d, e, f — отвечают несмещенным критериям, c — смещенному критерию, a и b — подобным критериям уровня α , d и e — равномерно наиболее мощным среди несмещенных, d — строго равномерно наиболее мощный несмещенный.

Здесь «строгий» означает, что нельзя улучшить функцию значимости без того, чтобы в какой-либо точке не ухудшилась мощность.

Подчеркнем, что в рассмотренных операциях важен порядок, в котором проводится оптимизация. Так, например, для того чтобы получить строго равномерно наиболее мощный несмещенный критерий гарантированного уровня α , нужно сначала отделить все критерии гарантированного уровня α , затем среди них выделить несмещенные критерии, среди несмещенных отделить наиболее мощные и, наконец, среди последних выбрать критерий с наилучшей функцией уровня значимости. Если мы будем проводить оптимизацию в другом порядке, то можем прийти к другому результату.⁵

⁵ В теории проверки сложной гипотезы против сложной альтернативы, по-видимому, не существует согласованной терминологии. Так, некоторые авторы не употребляют термины «строго» и «гарантированный уровень» или определяют их иначе, чем здесь, где использована терминология Барра (1974) и Лемана (1964).

Рисунок III.8 иллюстрирует понятия, связанные с оптимизацией. Предположим, что имеется задача проверки гипотезы

$$H_0: \theta \in [\theta_1; \theta_2] \text{ против } H_1: \theta_1 \neq [\theta_1; \theta_2]$$

при уровне значимости α .

Предположим далее, что для этой задачи существует лишь шесть критериев (a, b, \dots, f). Такое допущение не реалистично, так как обычно их бесконечно много. Мы приняли его, чтобы показать на рисунке соотношение между кривыми, отвечающими равномерно наиболее мощному критерию, и кривыми некоторых других типов. Графики на рис. III.8 между точками θ_1 и θ_2 показывают функцию уровня значимости $\alpha(\theta)$, а между точками $0, \theta_1$ и θ_2, ∞ — функцию мощности $\beta(\theta)$.

Числом α отмечается гарантированный уровень значимости α в %. Все шесть критериев являются критериями с гарантированным уровнем значимости, так как ни одна из кривых на участке $[\theta_1; \theta_2]$ не поднимается выше α . Кривая c отвечает смещенному критерию, так как ее мощность левее θ_1 и правее θ_2 оказывается ниже уровня значимости. Кривые a, b, d, e, f отвечают несмещенным критериям (их мощности не опускаются ниже уровня значимости). Среди критериев, которым отвечают кривые на рис. III.8, не существует равномерно наиболее мощного с гарантированным уровнем α . Это видно из рисунка, так как для любой кривой на нем найдется абсцисса, которой отвечает мощность бóльшая, чем мощность при такой же абсциссе на какой-либо из остальных пяти кривых. Кривые a и b соответствуют подобным критериям уровня α . Уровень значимости у них совпадает между θ_1 и θ_2 и равен α , b отвечает равномерно наиболее мощному подобному критерию, так как мощность b все время больше мощности a , а остальные критерии, кривые для которых приведены на рисунке, не являются подобными. Равномерно наиболее мощными среди несмещенных критериев являются критерии, отвечающие кривым d и e . Мощности этих критериев совпадают, мощность d оказывается выше мощности a, b и f . Мощность критерия c на многих участках выше мощности d , однако c не входит в класс несмещенных критериев. Наконец, критерий, которому отвечает кривая d , является строго равномерно наиболее мощным несмещенным. Действительно, кривая f имеет лучший уровень значимости, чем d , но она обладает худшей мощностью. Таким образом, приходится выбирать между e и d — у них одинаковая мощность, но d выгоднее по уровню значимости.

Как мы видели, в задаче проверки сложной гипотезы против сложной альтернативы не существует общего метода, гарантирующего получение критерия с оптимальными свойствами. В такой ситуации важно иметь некоторый стандартный метод, который обычно дает хорошие результаты, хотя и не гарантирует их всегда. Таким методом является метод отношения правдоподобия, он часто будет использоваться нами в дальнейшем.

III.3.5. Критерий отношения правдоподобия

Рассматриваемый критерий является состоятельным и асимптотически несмещенным, т. е. обладает хорошими свойствами при больших выборках (при некоторых ограничениях, которые обычно имеют место). К р и т е р и й о т н о ш е н и я п р а в д о п о д о б и я строится так.

Пусть $p_0(x; \theta)$ — выборочная плотность при H_0 (когда $\theta \in \omega_0$). Найдем для этого случая точечную оценку максимального правдоподобия $\hat{\theta}_0(x)$ (как это делается, см. в примере III.1) и подставим ее в выборочную плотность вместо параметра θ . Тогда получим

$$p_0(x; \hat{\theta}_0(x)).$$

Пусть, далее, $p_1(x; \theta)$ — выборочная плотность при $\theta \in \Omega$ (обратим внимание, что $\theta \in \Omega$, а не $\theta \in \omega_1$).

Для этого случая найдем другую оценку максимального правдоподобия $\hat{\theta}_1(x)$. Подставим вместо θ эту оценку в p_1 и получим

$$p_1(x; \hat{\theta}_1(x)).$$

Затем образуем отношение

$$R(x) = \frac{p_0(x; \hat{\theta}_0(x))}{p_1(x; \hat{\theta}_1(x))},$$

где $R(x)$ не зависит от параметра θ , однако функция распределения $R(x)$ зависит от θ .

В критическую область мы будем помещать такие точки выборочного пространства, для которых $R(x)$ по возможности мало, и отбирать их в критическую область w до тех пор, пока не достигнем уровня α , т. е. до тех пор, пока не наберется столько точек внутри w , что

$$\int_{\{w\}} p_0(x; \theta_0(x)) dx = \alpha.$$

Для этого область w определяем следующим образом: $x \in \{w\}$ в том и только в том случае, когда $R(x) < c_\alpha$, где константа c_α определяется из соотношения

$$\int_{\{R(x) < c_\alpha\}} p_0(x; \hat{\theta}_0(x)) dx = \alpha.$$

Конструкция такой критической области близка к конструкции наилучшей критической области по Нейману—Пирсону, подробно изученной нами в III.3.2. Однако имеются также и различия. В п. III.3.2 мы рассматривали отношение $p_0(x; \theta)/p_1(x; \theta)$, где θ в числителе было неизвестным параметром из ω_0 , θ в знаменателе — неизвестным параметром из ω_1 . Теперь неизвестный параметр не входит в R , а вместо него стоят оценки. При этом θ_0 соответствует

оценка при положении параметра в ω_0 , тогда как θ_1 соответствует оценка из всего Ω , а не из ω_1 .

Проиллюстрируем сказанное, рассматривая равномерное распределение с $\theta > 0$. Изучим случай, когда Ω отвечает $(0; \infty)$, а ω_0 есть сегмент $[\theta_1; \theta_2]$. Тогда

$$p_0(x; \theta) = \begin{cases} \frac{1}{\theta^n} & \text{при } x_{\max} \leq \theta \\ 0, & \text{если } x_{\max} > \theta \end{cases} \quad (\theta_1 \leq \theta \leq \theta_2),$$

$$p_1(x; \theta) = \begin{cases} \frac{1}{\theta^n} & \text{при } x_{\max} \leq \theta \\ 0, & \text{если } x_{\max} > \theta \end{cases} \quad (0 < \theta < \infty).$$

Оценка максимального правдоподобия в первом случае

$$\hat{\theta}(x) = \begin{cases} \theta_1, & \text{если } x_{\max} \leq \theta_1, \\ x_{\max}, & \text{если } x_{\max} > \theta_1. \end{cases}$$

Во втором случае эта оценка

$$\hat{\theta}_1(x) = x_{\max}.$$

Таким образом,

$$p_0(x; \hat{\theta}_0(x)) = \begin{cases} \frac{1}{\theta_1^n}, & \text{если } x_{\max} < \theta_1, \\ \frac{1}{x_{\max}^n}, & \text{если } \theta_1 \leq x_{\max} \leq \theta_2, \\ 0, & \text{если } x_{\max} > \theta_2, \end{cases}$$

и

$$p_1(x; \hat{\theta}_1(x)) = \frac{1}{x_{\max}^n} \text{ для всех } x.$$

Из сказанного ясно, что отношение правдоподобия

$$R(x) = \begin{cases} \left(\frac{x_{\max}}{\theta_1}\right)^n, & \text{если } x_{\max} < \theta_1, \\ 1, & \text{если } \theta_1 \leq x_{\max} \leq \theta_2, \\ 0, & \text{если } x_{\max} > \theta_2. \end{cases} \quad (\text{III. 3.6})$$

Предположим теперь, что одних таких точек, у которых $x_{\max} < \theta_1$ (т. е. точек, удовлетворяющих первому определению $R(x)$ в (III.3.6)), будет достаточно, чтобы образовать критическую область w_α уровня α . Это всегда будет иметь место, если

$$P(x_{\max} < \theta_1 | H_0) \geq \alpha. \quad (\text{III. 3.7})$$

Действительно, вероятность (III.3.7) равна $\left(\frac{\theta_1}{\theta}\right)^n$, а минимум этого выражения есть $\left(\frac{\theta_1}{\theta_2}\right)^n$.

Таким образом, условие (III.3.7) выполнено, если

$$\left(\frac{\theta_1}{\theta_2}\right)^n \geq \alpha.$$

Возьмем, например, в качестве $[\theta_1; \theta_2]$ сегмент [1.640; 1.733] при $n = 50$. Тогда $\left(\frac{\theta_1}{\theta_2}\right)^n$ приблизительно равно 0.063 и условие (III.3.7) будет выполняться при любом уровне значимости $< 6\%$. В дальнейшем мы будем сталкиваться лишь с такими задачами, в которых (III.3.7) выполнено. В этом случае критическая область определяется неравенством $\left(\frac{x_{\max}}{\theta_1}\right)^n \leq c_\alpha$, где c_α отвечает соотношению

$$\frac{1}{\theta_1^n} \int_{\left\{\left(\frac{x_{\max}}{\theta_1}\right)^n \leq c_\alpha\right\}} dx = \alpha,$$

откуда получаем $c_\alpha = \alpha$. Кроме того к критической области нужно отнести все точки, для которых $x_{\max} > \theta_2$.

Итак, критерий отношения правдоподобия для проверки сложной гипотезы против сложной альтернативы, т. е.

$$H_0: \theta \in [\theta_1; \theta_2] \text{ против } H_1: \theta_1 \neq [\theta_1; \theta_2],$$

при дополнительном условии

$$\left(\frac{\theta_1}{\theta_2}\right)^n \geq \alpha$$

имеет критическую область w_α следующего вида:

$$x_{\max} \in (0; \alpha^{1/n} \theta_1] \cup (\theta_2; \infty). \quad (\text{III. 3. 8})$$

Как видно из сравнения критерия отношения правдоподобия (III.3.8), для интервальной проверяемой гипотезы с двусторонним равномерно наиболее мощным критерием (III.3.2) эти критерии весьма близки и полностью совпадают, если $\theta_1 = \theta_2 = \theta_0$.

Нетрудно показать, что критерий (III.3.8) имеет уровень значимости

$$\alpha(\theta) = \alpha \cdot \left(\frac{\theta_1}{\theta}\right)^n \quad (\theta_1 \leq \theta \leq \theta_2) \quad (\text{III. 3. 9})^a$$

и функцию мощности

$$\beta(\theta) = \begin{cases} 1, & \text{если } \theta \leq \alpha^{1/n} \theta_1, \\ \alpha \cdot \left(\frac{\theta_1}{\theta}\right)^n, & \text{если } \alpha^{1/n} \theta_1 < \theta < \theta_1, \\ 1 - \frac{\theta_2^n - \alpha \theta_1^n}{\theta^n}, & \text{если } \theta > \theta_2. \end{cases} \quad (\text{III. 3. 10})$$

^a Слева в (III.3.9) через $\alpha(\theta)$ обозначена функция от θ , а справа α без указания в скобках аргумента является числовым множителем.

Уровень значимости $\alpha(\theta)$ достигает максимального значения при $\theta = \theta_1$. Таким образом, рассматриваемый критерий принадлежит к критериям с гарантированным уровнем значимости α .

Из (III.3.10) видна состоятельность критерия; очевидно также, что при $n \rightarrow \infty$ смещения нет (асимптотическая несмещенность). Если же взять конечное n , то на основании третьего определения и (III.3.10) при $\theta = \theta_2$ получаем максимальную мощность, равную $\alpha \cdot (\theta_1/\theta_2)^n$. Так как θ_2 всегда больше, чем θ_1 , то это выражение меньше, чем α , и критерий является смещенным. Так, например, если мы возьмем $\alpha = 0.05$, $\theta_1 = 1.640$, $\theta_2 = 1.733$ и $n = 50$, то получим

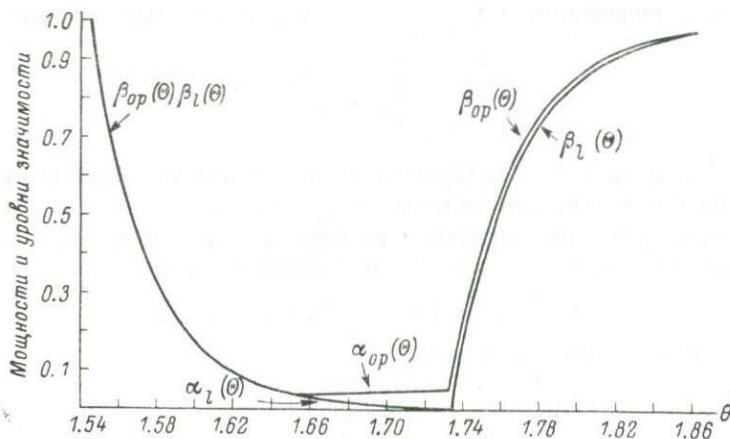


Рис. III.9. Кривые критериев отношения правдоподобия (с нижним индексом l) и оптимального (с нижним индексом op) в случае проверки интервальной гипотезы $H_0: \theta_1 \leq \theta \leq \theta_2$ против $H_1: \theta < \theta_1$ или $\theta > \theta_2$.

значение мощности $\beta(\theta)$ при $\theta \rightarrow \theta_2$, равную 0.003, вместо минимальной мощности 0.05 для несмещенных критериев. Поскольку отношение 0.05 и 0.003 велико, для данных числовых значений смещение весьма существенно. Отметим, наконец, что если бы мы использовали в качестве статистики критерия исправленную на смещение точечную оценку правдоподобия $\frac{n+1}{n} x_{\max}$ вместо x_{\max} , то критерий тем не менее остался бы смещенным.

Таким образом, смещение в рассматриваемом случае связано со спецификой метода отношения правдоподобия.

На рис. III.9 показаны уровень значимости и мощность для рассматриваемого критерия — для $\theta_1 = 1.640$, $\theta_2 = 1.733$, $\alpha = 0.05$, $n = 50$.

Перейдем теперь к разбору конкретного примера, представляющего интерес для геохимии. В нем используются приведенные ранее понятия и даются некоторые общие разработки, возникающие в процессе работы с конкретным материалом.

Пример III.4. О распределений
вероятностей концентраций
Na₂O в базальтах мира

В 1960 г. автором настоящих строк был исследован вопрос о скошенности распределений вероятностей концентраций различных химических элементов в горных породах (Vistelius, 1960). При этом исходное предположение в терминах, принятых в этой книге, сводилось к тому, что распределение вероятностей весов или чисел атомов Na в единице объема базальта нормально и после перехода к процентным величинам сохраняет нормальность. Таким образом, не проводилось различия между функциями распределения исходных величин и величин, пересчитанных на доли (проценты, промилле, граммы на тонну и т. п.). Возникшая при этом задача о типе распределения Na была решена с помощью классических статистических методов (критерий Колмогорова). При этом было показано, что гипотеза нормальности процентных содержаний Na₂O не опровергается.

За прошедшее время выяснилось, что нормальное распределение «допроцентных» величин при процентном пересчете в подавляющем большинстве случаев переходит в распределение Бернштейна, как это было показано в гл. II.

Таким образом, исходные предпосылки, на которых строилось исследование, начали вызывать сомнение. Так как рассматриваемая задача весьма типична и имеет большое значение для геохимии, мы переработали материал заново.

Решаемая нами задача имеет следующую специфику. Базальты в разных пунктах земного шара имеют различный химический состав. В каждом пункте сделано мало химических анализов, но общее число пунктов велико. Если мы смешаем весь материал, то вид выборочной функции распределения будет определяться весами, с которыми введены анализы из различных пунктов опробования. В этих условиях проверить вид функции сложно. Таким образом, остается единственный путь — проверка распределения по многим малым выборкам. При этом имеется метод (Крамер, 1948), позволяющий решить задачу о проверке нормальности распределения. Для этого вместо изучаемой величины x_{ij} (в нашем случае — содержание Na₂O, в %) вводится величина

$$\tau_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j},$$

где x_{ij} — содержание Na₂O в i -той пробе из j -того пункта опробования; \bar{x}_j — среднее по 4 анализам в j -том пункте опробования; s_j — несмещенная оценка стандарта для того же пункта по тем же четырем анализам из пункта j . В каждом пункте j берется по 4 анализа, для каждого анализа рассчитывается τ_{ij} и затем случайно выбирается одно из них. Отобранные указанным способом τ_{ij} распре-

делены равномерно на сегменте $[-\sqrt{3}; \sqrt{3}]$, если значения x_{ij} нормальны и взаимно стохастически независимы, а математические ожидания (a_{ij}) и стандарты σ_{ij} для Na_2O одни и те же для каждого наблюдения (i) в пределах фиксированного j (Крамер, 1948, с. 426).

Равномерное распределение случайной величины τ на сегменте $[-\sqrt{3}; \sqrt{3}]$ преобразуется к равномерному распределению на $[0; \sqrt{3}]$ при замене случайной величины τ на ее абсолютное значение. Так как в дальнейшем окажется, что удобнее работать с $|\tau|$, то ее значения мы приводим в табл. III.5.

Таблица III.5

Точки опробования, среднее содержание Na_2O (‰) и значения $|\tau|$ (Vistelius, 1960)

Место опробования	Na_2O	$ \tau $	Место опробования	Na_2O	$ \tau $
Южная Африка	2.970	0.6141	Остров Хуан Фернандес	2.868	1.3427
Южная Родезия	1.915	0.7607	Остров Мангарева	2.260	0.4092
Кения	2.678	1.2605	Острова Таити	7.312	1.3253
Остров Мадагаскар	2.140	1.3077	Вулкан Мауна-Лоа	2.272	0.9931
Индокитай	3.050	0.0343	(исторический поток)		
Северный Китай	3.285	0.2890	Вулкан Килауэа (доисторический поток)	2.198	1.3840
Центральная Ява	2.362	0.9730	Вулкан Килауэа (историческое время)	2.410	0.3247
Алеутские острова	3.562	0.0487	Остров Хаулаилаи	2.985	0.1759
Остров Суматра	3.345	1.2006	Вулкан Мауна-Кеа	2.748	0.2144
Вулкан Кракатау	2.692	0.8424	Кохала	3.075	0.4275
Армения, СССР	2.865	0.9250	Фарерские острова	2.688	0.5748
Якутия, СССР	2.025	0.0609	Азорские острова	3.030	0.6142
Камчатка, СССР	2.265	1.2619		2.722	0.5381
Земля Франца-Иосифа, СССР	1.670	0.2058	Архипелаг Шпицберген	2.650	1.4110
Река Уссури, СССР	3.185	0.5277	Остров Оаху	4.585	0.0903
Река Хиллок, СССР	3.108	0.3833	Остров Ян-Майен	3.180	1.2247
Нижняя Силезия	3.010	1.0605	Гора Дракенберг,	1.712	1.2500
Австралия	3.355	0.6142	Южная Африка		
Словакия, ЧССР	2.945	1.2302	Южная Африка	1.742	0.3405
Вулкан Парикутин, Мексика	3.832	0.3958	Южная Африка	2.810	0.2665
Армения, СССР	2.578	1.4394	Маскаренские острова	2.725	0.5967
Вулкан Авача, Камчатка, СССР	2.860	1.3886		2.930	1.2298
Вулкан Фудзияма, Япония	2.640	1.0518	Острова Тристан-да-Кунья и Гоуф	5.845	0.8201
Венгрия	3.322	1.4088	Япония	2.300	0.9929
То же	3.690	0.5623	Западная Гренландия	1.252	0.6311
Остров Реюньон	1.098	0.6498			
Калифорния, США	2.488	0.6764			

Итак, мы имели такую исходную задачу: необходимо проверить гипотезу о наличии нормальных распределений процентных кон-

центраций Na_2O против гипотезы о том, что смешиваемые распределения могли быть какими угодно, за исключением нормальных. Эту задачу мы преобразовали к другой, заключающейся в следующем: требуется проверить гипотезу о равномерном распределении безразмерной величины τ на $[0; \sqrt{3}]$ против любого другого распределения (неравномерного или равномерного с $\theta \neq \sqrt{3}$). Последняя задача, согласно указанной работе Крамера (1948), эквивалентна предыдущей, если x_{ij} стохастически независимы и параметры у смешанных нормальных распределений постоянны в пределах каждого опробываемого участка. Оба эти предположения из геологических соображений кажутся реальными.

Задача о проверке равномерного распределения поставлена в непараметрической форме, так как альтернатива не определяется значением параметра. Поэтому для принятия решения мы не имеем ничего лучшего, чем общепринятую и широко известную в геологии методику проверки гипотез с помощью критериев согласия (χ^2 -критерий и критерий Колмогорова, использующий максимальное отклонение d_{\max}^* эмпирического распределения от теоретического; мы предполагаем их известными читателю, см. Ван дер Варден, 1960). В результате использования обоих этих критериев согласия было выяснено, что гипотеза равномерного распределения на $[0; \sqrt{3}]$ не противоречит наблюдениям, если брать 5%-й уровень значимости.

Если бы гипотеза была забракована, то мы могли бы забраковать и исходное предположение о смеси нормальных распределений. При этом вероятность того, что наше заключение ошибочно, была бы около $1/20$. Так как в нашем случае гипотеза принята, то существует опасность сделать неправильное заключение из-за двух обстоятельств. Прежде всего, при переходе от исходной задачи о смеси нормальных распределений концентраций Na_2O к задаче о равномерном распределении резко теряется эффективность метода, если хотя бы немного нарушены условия о независимости наблюдений и постоянстве параметров (Петров, 1954). Далее, если эти условия соблюдены точно, то при применении критерия согласия можно сделать ошибку второго рода. Это означает, что можно ошибочно принять гипотезу о том, что выборка происходит из равномерной совокупности на $[0; \sqrt{3}]$, в то время как в действительности эта какая-то иная совокупность. Поскольку мощность критерия согласия, как правило, низка и не поддается вычислению на практике, то с возможностью такой ошибки следует считаться. Поэтому заменим непараметрическую постановку задачи о равномерном распределении параметрической. При этом мы будем проверять ту же гипотезу равномерного распределения на $[0; \sqrt{3}]$ против равномерного же распределения с $\theta \neq \sqrt{3}$. Последняя постановка задачи не эквивалентна предыдущей, но в случае браковки гипотезы может быть обоснована. Действительно, если

мы забракуем равномерные распределения с $\theta = \sqrt{3}$ против равномерного же с $\theta \neq \sqrt{3}$, то тем более мы должны забраковать гипотезу, если допускать в класс альтернатив еще неравномерные распределения. Если же гипотеза принималась, то такое сведение непараметрической задачи к параметрической недопустимо.

Отметим, наконец, что предположение о равномерном распределении представляется осмысленным, поскольку применение критерия согласия χ^2 без фиксации значения параметра θ не опровергает этого предположения на очень высоком уровне.

Итак, мы хотим проверить

$$H_0: \theta = \sqrt{3} \text{ против } H_1: \theta \neq \sqrt{3}$$

в предположении равномерного на $[0; \theta]$ распределения.

В отмеченной ситуации, как указывалось в п. III.3.5, наиболее реалистичной является постановка задачи с интервальной H_0 против интервальной H_1 (т. е. что $\theta \in [\theta_1; \theta_2]$, где сегмент включает $\sqrt{3}$ вместо частного случая $\theta = \sqrt{3}$). Задача с интервальной H_0 против двусторонней сложной альтернативы H_1 может быть решена по наблюдениям с помощью критерия отношения правдоподобия (III.3.8) с уровнем значимости (III.3.9) и мощностью (III.3.10). Там же отмечалось, что этот критерий имеет смещение вблизи правой точки θ_2 , что делает использование его неуместным, если есть основания предполагать, что истинное значение θ может находиться вблизи θ_2 . Этот случай имеет место в нашем примере, так как мы должны выбрать θ_2 вблизи $\sqrt{3}$. Брать далеко отстоящие вправо от $\sqrt{3}$ значения θ_2 нерационально, поскольку максимальное выборочное значение 1.4394 находится значительно левее $\sqrt{3}$.

Далее, мы не проводили исследования мощности различных критериев при интервальной H_0 , без чего нельзя выяснить, насколько удовлетворительна мощность отношения правдоподобия. В связи с этим попытаемся разыскать оптимальный в определенном смысле критерий. Для этого докажем следующую теорему.

Т е о р е м а III.2. *Для семейства распределений (R_θ) , где R_θ — равномерное распределение, сосредоточенных на $[0; \theta]$, в задаче проверки гипотезы $\theta \in [\theta_1; \theta_2]$ против альтернативы $\theta \notin [\theta_1; \theta_2]$ существует строгий, равномерно наиболее мощный критерий гарантированного уровня α . Критическая область w_α этого критерия такова, что гипотеза отвергается в каждом из следующих двух случаев:*

1) все координаты (наблюдения) x_i ($i=1, 2, \dots, n$) выборочной точки x принадлежат множеству

$$\{[0; \alpha^{1/n} \theta_1] \cup [\alpha^{1/n} \theta_1 + (1 - \alpha^{1/n}) \theta_2; \theta_2]\};$$

2) $x_{\max} > \theta_2$.

Напомним свойства указанного в теореме критерия: его уровень значимости между θ_1 и θ_2 нигде не поднимается выше α . Из всех критериев с таким свойством он имеет наибольшую мощность.

Если какой-либо другой критерий имеет более низкий уровень значимости, то этот другой критерий где-либо имеет меньшую мощность. Наконец, так как критерий является равномерно наиболее мощным с гарантированным уровнем, то он обладает свойством несмещенности.

Предварительные замечания. Пусть имеется равномерное на $[0; \theta]$ распределение и некоторое множество \mathcal{A} , все точки которого расположены между 0 и θ . Вероятность попадания отдельного наблюдения в \mathcal{A} , т. е. $p(x \in \mathcal{A})$, зависит лишь от меры множества \mathcal{A} и равна ⁷

$$\frac{\text{mes } \mathcal{A}}{\theta},$$

а вероятность попадания выборочной точки x в \mathcal{A}^n , где $\mathcal{A}^n = \mathcal{A} \times \mathcal{A} \times \dots \times \mathcal{A}$, есть

$$\left(\frac{\text{mes } \mathcal{A}}{\theta}\right)^n.$$

Рассмотрим следующую задачу. Пусть $a > \theta$ — некоторое число. Необходимо найти множество со значением лебеговой меры, равной α ($\alpha < a$), все точки которого лежат между нулем и a , такое, чтобы обеспечить наибольшую возможную вероятность попадания наблюдения x в это множество (при неопределенном положении параметра θ между нулем и a). Далее, нужно указать множество меры α с точками между нулем и a такое, чтобы обеспечить наименьшую вероятность попадания наблюдения в это множество (для неопределенного параметра θ между нулем и a). Множеством с наибольшей вероятностью будет сегмент $[0; \alpha]$, а множеством с наименьшей вероятностью — сегмент $[a - \alpha; a]$. В этом нетрудно убедиться, повторяя рассуждения, использованные при построении односторонних равномерно наиболее мощных критериев.

Доказательство теоремы III.2. Покажем сначала, что наш критерий принадлежит к классу критериев с гарантированным уровнем α . Вычисления дают

$$\alpha(\theta) = p(x \in w_\alpha | \theta_1 \leq \theta \leq \theta_2) = \begin{cases} \alpha \cdot \left(\frac{\theta_1}{\theta}\right)^n, & \text{если } \theta_1 \leq \theta \leq \alpha^{1/n}\theta_1 + (1 - \alpha^{1/n})\theta_2, \\ \left[1 - (1 - \alpha^{1/n})\left(\frac{\theta_2}{\theta}\right)\right]^n, & \text{если } \alpha^{1/n}\theta_1 + (1 - \alpha^{1/n})\theta_2 \leq \theta \leq \theta_2, \end{cases} \quad (\text{III. 3. 11})$$

Из (III. 3. 11) ясно, что $\alpha(\theta_1) = \alpha(\theta_2) = \alpha$ и что $\alpha(\theta) < \alpha$ для $\theta_1 < \theta < \theta_2$.

⁷ Напомним, что $\text{mes } \mathcal{A} = \int_{\mathcal{A}} dx$, где интеграл является интегралом Лебега.

Заметим далее, что оптимальная критическая область должна содержать точки x' , для которых $x_{\max} > \theta_2$.

Рассмотрим остающуюся после удаления точек x' часть критической области. Эта часть, согласно условиям теоремы, есть множество \mathcal{A}^n ,

$$\mathcal{A} = \{[0; \alpha^{1/n}\theta_1] \cup [\alpha^{1/n}\theta_1 + (1 - \alpha^{1/n})\theta_2; \theta_2]\},$$

с мерой

$$\text{mes } \mathcal{A} = \alpha^{1/n}\theta_2.$$

Покажем, что так выбранное \mathcal{A} обеспечивает максимальную мощность. Пусть, вопреки этому, существует некоторое другое множество \mathcal{A}' , приводящее к мощности $\beta'(\theta) > \beta(\theta)$ для некоторого значения θ . Для θ в этом неравенстве возможно положение либо левее θ_1 , либо правее θ_2 . Если $\theta > \theta_2$, то из $\beta'(\theta) > \beta(\theta)$, согласно предварительному замечанию на с. 209, следует, что

$$\text{mes } \mathcal{A}' > \text{mes } \mathcal{A}.$$

Последнее же неравенство влечет за собой неравенство $\alpha(\theta_2) > \alpha$, что противоречит свойству критерия иметь гарантированный уровень α .

Пусть, далее, $\theta < \theta_1$. Согласно тем же предварительным замечаниям на с. 209, сегмент $[0; \alpha^{1/n}\theta_1]$ дает наилучшую мощность. Удлинение этого сегмента привело бы снова к нарушению гарантированного уровня α .

Итак, критерий является равномерно наиболее мощным среди критериев с гарантированным уровнем α .

Теперь покажем, что критерий является строгим среди указанного класса.

Критическая область любого равномерно наиболее мощного критерия с гарантированным уровнем α такова, что:

- 1) в критическую область входят точки x , для которых $x_{\max} > \theta_2$;
- 2) в нее входит также некоторое множество \mathcal{B}^n следующего строения: множество \mathcal{B}^n содержит в качестве своего подмножества сегмент $[0; \alpha^{1/n}\theta_1]$ и еще некоторое подмножество \mathcal{C} произвольного строения, имеющее меру $\text{mes } \mathcal{C} = \alpha^{1/n}\theta_2$. Согласно определению на с. 199, критерий будет строгим, если \mathcal{C} выбрано так, чтобы вероятность попадания в него выборочной точки x для $\theta_1 < \theta < \theta_2$ была наименьшей.

Согласно предварительному замечанию на с. 209, \mathcal{C} надо составить из некоторого отрезка, отложенного влево от θ_2 , и сегмента $[0; \alpha^{1/n}\theta_1]$. Поскольку общая мера этого отрезка и указанного сегмента должна быть $\alpha^{1/n}\theta_2$, отрезок выбирается единственным способом как

$$[\alpha^{1/n}\theta_1 + (1 - \alpha^{1/n})\theta_2; \theta_2].$$

Этим завершается доказательство теоремы III. 2.

Оптимальный критерий, полученный в теореме III. 2, имеет мощность

$$\beta(\theta) = P(x \in w_\alpha / \theta \in \{(0; \theta_1) \cup (\theta_2; \infty)\}) =$$

$$= \begin{cases} 1, & \text{если } \theta \leq \alpha^{1/n} \theta_1, \\ \alpha \cdot \left(\frac{\theta_1}{\theta}\right)^n, & \text{если } \alpha^{1/n} \theta_1 \leq \theta < \theta_2, \\ 1 - (1 - \alpha) \left(\frac{\theta_2}{\theta}\right)^n, & \text{если } \theta > \theta_2 \end{cases} \quad (\text{III. 3. 12})$$

и уровень значимости $\alpha(\theta)$, даваемый (III. 3. 11).

Ниже приведены значения $\alpha(\theta)$ и $\beta(\theta)$ для полученного оптимального критерия и критерия отношения правдоподобия (табл. III. 6 и III. 7).

Таблица III.6

Значения $\alpha_{оп}(\theta)$ и $\alpha_l(\theta)$
($\alpha_{оп}$ — оптимальный критерий, l — критерий правдоподобия)

θ	1.640	1.645	1.650	1.700	1.733
$\alpha_{оп}(\theta)$	0.050	0.043	0.043	0.047	0.050
$\alpha_l(\theta)$	0.050	0.043	0.037	0.008	0.003

Таблица III.7

Значения $\beta_{оп}(\theta)$ и $\beta_l(\theta)$
(обозначения те же, что и в табл. III.6)

θ	1.545	1.547	1.550	1.560	1.580	1.600	1.733
$\beta_{оп}(\theta)$	1.00	0.926	0.841	0.609	0.322	0.172	0.050
$\beta_l(\theta)$	1.00	0.926	0.841	0.609	0.322	0.172	0.003
θ	1.736	1.740	1.750	1.760	1.780	1.810	1.850
$\beta_{оп}(\theta)$	0.129	0.223	0.417	0.562	0.751	0.899	0.964
$\beta_l(\theta)$	0.086	0.185	0.388	0.540	0.738	0.887	0.962

На рис. III.9 приведены графики, построенные по данным табл. III.6 и III.7 для оптимального критерия и приводимого здесь для сравнения критерия отношения правдоподобия. В обоих случаях $\alpha=0.05$, $n=50$, $\theta_1=1.640$ и $\theta_2=1.733$ (значение θ_2 , равное 1.733, было взято из-за того, что было очень маловероятно, что при наблюдаемом значении $x_{\max}=1.4394$ величина θ окажется много больше $\sqrt{3}$; θ_1 произвольно принято равным 1.640 как одно из значений между наблюдаемым x_{\max} и $\sqrt{3}$). Как видно из рис. III.9, оптимальный критерий для левых альтернатив имеет функцию мощности, совпадающую с функцией мощности критерия отношения правдоподобия; для правых альтернатив мощность оптимального критерия несколько выше. В то же время критерий отношения правдоподобия имеет лучший уровень значимости. Таким образом, для значений α , θ_1 , θ_2 и n , фигурирующих в нашем при-

мере, мощность и уровень значимости сравниваемых критериев близки. Однако оптимальный критерий обладает важным свойством несмещенности, существенность чего в нашей задаче подчеркивалась ранее.

Рассмотрим теперь результат применения оптимального критерия к примеру о распределении Na_2O . Будем считать решение обоснованным, если оно принимается при вероятности ошибок обоих родов < 0.05 . Тогда, как видно из рис. III.9, для примера с базальтами сегмент $[1.640; 1.733]$, содержащий предполагаемое значение θ , отличается от значений < 1.545 и > 1.800 и не может быть отличен от значений θ между 1.545 и 1.640 , с одной стороны, и значений от 1.733 до 1.800 — с другой. Учитывая, что выборка дала $x_{\max} = 1.4394$, заключаем, что $\theta_0 \neq \sqrt{3}$ и что скорее всего θ_0 находится между 0 и 1.545 , либо она > 1.800 .

Применение критерия для правых односторонних альтернатив, как легко убедиться с помощью расчетов, изложенных в п. III.3.2 (с. 189, 190), приводит к браковке гипотезы $\theta > 1.800$ с чрезвычайно малыми вероятностями ошибок обоих родов. Таким образом, если распределение $|\tau|$ действительно равномерно, то истинное значение параметра θ скорее всего располагается между 0 и 1.545 .

Итак, предположение о том, что процентные концентрации Na_2O в базальтах распределены нормально, не подтверждается. Это могло бы иметь место, если допроцентные значения Na в единице объема базальта были распределены нормально и это нормальное распределение было деформировано процентным пересчетом, как это подробно рассмотрено в гл. II.

III.3.6. О доверительных интервалах

Только что мы закончили исследование, в котором проверили гипотезу о том, что параметр $\theta \in [\theta_1; \theta_2]$. При исследовании этой гипотезы мы пришли к заключению, что скорее всего $\theta \in (0; 1.545)$ и, вероятно, $\theta \notin [1.640; 1.733]$. Таким образом, мы составили суждение о принадлежности θ некоторому интервалу. В связи с этим рассмотрим подробнее вопрос о том, как разумно строить по выборке x_1, \dots, x_n интервал $(\theta_1(x); \theta_2(x))$, предположительно содержащий истинное значение параметра θ . Такие интервалы называются *доверительными интервалами для θ* и широко используются на практике.

Границы доверительных интервалов являются случайными величинами. Поэтому интервалы, о которых до сих пор шла речь, не являлись доверительными. Границы этих интервалов (θ_1 и θ_2) указывались до того, как была взята выборка. Так, сегмент $[1.640; 1.733]$ мы взяли из априорных соображений. Интервал $(0; 1.535)$ был найден по результатам исследования функций мощности и уровня значимости. Обе эти функции были построены независимо от того, что дала конкретная выборка.

Случайными до сих пор были не интервалы, а принятые решения о принадлежности θ интервалу. Эти решения принимались по выборке x_1, x_2, \dots, x_n . Оптимальность статистической процедуры означала, что при плохо выбранном интервале вероятность отрицательного решения, т. е. $\theta \notin (\theta_1; \theta_2)$, была очень велика, но границы интервала мы определяли независимо от результатов выбора.

Рассмотрим теперь, как находить границы интервалов по выборке (в более общей постановке рассматривается задача о том, как находить оптимальное случайное множество, называемое доверительным, предположительно содержащее истинное значение θ ; здесь, преследуя иллюстративные цели, мы коснемся лишь простейшего случая оценивания интервалом одномерного параметра). Допустим, что указаны две функции от выборки: $\theta_1(x)$ — нижняя граница и $\theta_2(x)$ — верхняя граница, причем $\theta_1(x) \leq \theta_2(x)$.

Границы $\theta_1(x)$ и $\theta_2(x)$ представляют собой два числа. Таким образом, одна выборка дает один интервал, т. е. одну реализацию. Допустим, что многократно извлекается выборка

$$\begin{aligned} X_1 = x_{11}, x_{12}, \dots, x_{1n}; X_2 = x_{21}, x_{22}, \dots, x_{2n}; \dots; \\ X_i = x_{i1}, x_{i2}, \dots, x_{in}. \end{aligned} \quad (\text{III. 3. 13})$$

осуществляемая в фиксированных условиях, и по данным этих выборок строятся доверительные интервалы

$$(\theta_1(x_1), \theta_2(x_1)), \dots, (\theta_1(x_i), \theta_2(x_i)).$$

Каждый отдельный интервал $(\theta_1(x_n); \theta_2(x_n))$, после того как n -я выборка реализована, может либо включать, либо не включать истинного значения параметра θ_0 . Попадание θ_0 в отдельный интервал — не случайное событие. Если же рассматривается процесс опробования в целом, то появляется вероятность того, что в случайно взятой из последовательности (III.3.13) выборке соответствующий ей доверительный интервал будет включать θ_0 . Обозначим эту вероятность

$$p(\theta_1(x) \leq \theta_0 \leq \theta_2(x)). \quad (\text{III. 3. 14})$$

Напомним, что в (III.3.14) $\theta_1(x)$ и $\theta_2(x)$ — случайные величины. Очевидно, что вероятность (III.3.14) желательнее получить при прочих равных условиях возможно большей. Если вероятность (III.3.14) равна $1-\alpha$, то говорят, что имеется 100 α %-й доверительный интервал или интервал ⁸ уровня значимости α . Таким образом, 5%-й доверительный интервал в 95% всех выборок охватывает истинное значение параметра, а в 5% не охватывает.

⁸ В литературе нет единой точки зрения на то, как называть доверительный интервал, охватывающий неизвестный параметр с вероятностью $1-\alpha$. Иногда такой интервал называют 100 α %-м, как это принято нами, иногда тот же интервал называют 100 (1- α)%-м (Кендал, Стьюарт, 1973, с. 141).

Хорошее качество интервала не обеспечивается наличием одного высокого уровня значимости. Отметим еще некоторые важнейшие желательные свойства доверительного интервала.

Желательно, чтобы доверительный интервал при многократном опробовании чаще включал точку θ_0 , чем любую другую точку $\theta \neq \theta_0$. Это свойство называется наибольшей селективностью. Далее, если

$$P(\theta_1(x) \leq \theta \leq \theta_2(x)) \leq 1 - \alpha \text{ для всех } \theta \neq \theta_0, \quad (\text{III. 3. 15})$$

то такой доверительный интервал называется несмещенным.

Желательно также, чтобы при прочих равных условиях средняя длина доверительного интервала $E\theta_2(x) - E\theta_1(x)$ была кратчайшей. Интервал уровня α , обладающий таким свойством, называется кратчайшим доверительным интервалом.

Наконец, естественно требовать, чтобы при неограниченном росте объема выборки доверительный интервал стягивался в точку θ_0 , т. е. чтобы

$$\lim_{n \rightarrow \infty} \theta_1(x) \rightarrow \theta_0 \text{ и } \lim_{n \rightarrow \infty} \theta_2(x) \rightarrow \theta_0.$$

Последнее свойство называется состоятельностью доверительного интервала.

Отметим, что оптимальный доверительный интервал может не быть центральным, т. е. вероятность попадания θ_0 в правую и левую половины интервала могут быть неравными.

Практические способы построения функций $\theta_1(x)$ и $\theta_2(x)$ могут быть разными. Рассмотрим простейшую ситуацию, когда их построение осуществляется с помощью теории Неймана—Пирсона — проверки простой гипотезы против сложной альтернативы.

Допустим, что нам известен вид выборочной плотности $f(x; \theta)$ и что эта плотность сохраняется во всех выборках объема n . Примем далее, что имеется критерий для проверки $H_0: \theta = \theta_0$ против $H_1: \theta \neq \theta_0$ и что этому критерию соответствует критическая область w_α уровня α .

Пусть произведена единичная выборка. Соберем теперь все отдельные значения θ , при которых гипотеза принимается на основании данной выборки. Это дает некоторое множество значений θ , которые и представляют *доверительное множество уровня α для параметра θ* . Если множество оказывается интервалом, то это доверительный интервал уровня α .

Существует известный параллелизм между свойствами критерия и свойствами доверительного интервала, построенного с помощью этого критерия по изложенному способу.

Если критерий был уровня значимости α , то доверительный интервал оказывается уровня α .

Если критерий был несмещенным, то доверительный интервал будет несмещенным.

Если критическая область критерия была построена по состоятельной точечной оценке, то доверительный интервал состоятелен.

Наконец, если критерий был равномерно наиболее мощным, то доверительный интервал будет наиболее селективным и несмещенным.

Отметим также, что равномерно наибольшая мощность критерия — без некоторых дополнительных условий — еще не влечет наличия у доверительного интервала наименьшей средней длины.

Проиллюстрируем сказанное о доверительных интервалах некоторыми расчетами по данным примера III.4.

Напомним, что равномерно наиболее мощный двусторонний критерий уровня α имел следующую критическую область:

$$w_\alpha: x_{\max} < \alpha^{1/n} \theta_0 \text{ и } x_{\max} > \theta_0.$$

Эта область охватывает значения θ , бракуемые выборкой. Дополнительная область, где расположены принимаемые при данной выборке значения θ , будет, очевидно,

$$W - w_\alpha: \alpha^{1/n} \theta_0 \leq x_{\max} \leq \theta_0.$$

Записывая последнее неравенство в виде доверительного множества для θ_0 , получаем

$$x_{\max} \leq \theta_0 \leq \frac{x_{\max}}{\alpha^{1/n}}. \quad (\text{III. 3. 16})$$

Таким образом, в данном случае доверительное множество оказалось интервалом (III.3.16).

Подчеркнем еще раз, что

$$P \left(x_{\max} \leq \theta_0 \leq \frac{x_{\max}}{\alpha^{1/n}} \right) = 1 - \alpha, \quad (\text{III. 3. 17})$$

и эта вероятность может быть отнесена к опробованию повторными выборками. В этом случае указанная вероятность имеет частотную интерпретацию. Если же выборка является однократной, то частотной статистической интерпретации придать (III.3.17) нельзя.

После того как в примере III.4 выборка была извлечена, оказалось, что $x_{\max} = 1.4394$ и мы получаем интервал

$$1.4394 \leq \theta \leq 1.5282.$$

С этим детерминированным интервалом уже не связана какая-либо вероятность, основанная на частотной интерпретации. Последний раз вероятность, допускающая частотную интерпретацию, появлялась для интервала (III.3.16).

Возвращаясь к примеру III.4, мы еще раз подтвердим, что истинное значение θ_0 при 5%-м уровне значимости нельзя принять равным $\sqrt{3}$. Выясним теперь, при каком уровне значимости можно было бы признать $\theta_0 = \sqrt{3}$. Для этого нужно решить уравнение

$$\frac{1.4394}{\alpha^{1/50}} = \sqrt{3},$$

откуда $\alpha \approx 0.0001$, т. е. должен быть взят доверительный интервал уровня 0.01%. Так как кажется невероятным, что в нашем случае реализовалось событие, происходящее 1 раз среди 10 000 испытаний, то мы полагаем, что $\theta \neq \sqrt{3}$.

Посмотрим при каком объеме выборки n_0 значение $\sqrt{3}$ попало бы в 5%-й доверительный интервал, если бы мы предположили, что в выборке объема n_0 реализовалось значение $x_{\max} = 1.4394$. Подставляя значения $\alpha = 0.05$ и $n = n_0$, при $x_{\max} = 1.4394$ из уравнения (III.3.17) получим $n_0 \leq 16$.

Природный процесс, порождающий характеристики, изучаемые геологами, протекает так, что исследователь вынужден, как правило, иметь дело не с самими этими характеристиками, а с распределениями вероятностей их значений. Суждение о распределении соответствующих вероятностей возникает на основе совокупности данных, которыми обладает исследователь. Проверка же правильности суждений осуществляется с помощью наблюдений. При этом возникают две группы задач. Одна группа связана с оценкой значений параметров функций распределения по данным наблюдений. Эта оценка производится в предположении, что функция распределения вероятностей известна с точностью до неизвестного параметра.

Задача оценивания параметров имеет много специфических особенностей, зачастую неожиданных для человека, никогда ранее с ними не сталкивавшегося. Вторая группа задач возникает тогда, когда требуется установить, насколько высказанное суждение о функции распределения не противоречит наблюдениям. Для решения этих задач существуют различные теории. В настоящей очерке мы рассмотрели классическую теорию, созданную Нейманом и Пирсоном. Изложенная теория позволяет наглядно показать характерную постановку вопроса и выяснить возникающие при этом трудности. Все это было проиллюстрировано разбором гипотезы о равномерном распределении. В конце главы дан пример, посвященный задаче о распределении вероятностей концентраций Na_2O в базальтах. Пример весьма симптоматичен — несмотря на то что задачу удалось свести к проверке гипотезы о равномерном распределении, решение ее в данной конкретной ситуации потребовало новых математических разработок. Полученные результаты и позволили решить геохимическую задачу оптимальным методом.

Л и т е р а т у р а

- Б а р р а Ж. Р. Основные понятия математической статистики. М., «Мир», 1974. 275 с.
- Б а ш а р и н Г. П. О статистическом оценивании энтропии. — Теория вероятностей и ее приложения. 1959, т. 4, вып. 3, с. 361—364.
- В а н - д е р - В а р д е н Б. Л. Математическая статистика. М., ИЛ, 1960. 434 с.
- В и с т е л ю с А. Б. Задачи геохимии и информационные меры. — Советская геология, 1964, № 12, с. 5—26.

- Грязнова Т. Е. Ориентированные структуры песчаников продуктивной толщи Ашшеронского полуострова. Л.—М., Гостоптехиздат, т. II, 1953, с. 224—241.
- Кендалл М., Стьюарт А. Теория распределений. Т. I. М., «Наука», 1966. 587 с.
- Кендалл М., Стьюарт А. Статистические выводы и связи. Т. II. М., «Наука», 1973. 899 с.
- Крамер Г. Математические методы статистики. М., ИЛ, 1948. 631 с.
- Кульбак С. Теория информации и статистика. М., «Наука», 1967. 408 с.
- Леман Э. Проверка статистических гипотез. М., «Наука», 1964. 498 с.
- Петров В. В. Обобщение одной предельной теоремы Крамера. — Усп. математ. наук, 1954, т. 9, вып. 4, с. 195—202.
- Рао С. Р. Линейные статистические методы и их применение. М., «Наука», 1968. 547 с.
- Романова М. А. Марковские свойства последовательностей зерен в редкометальных гранитах, их использование при поисковых работах и петрологических исследованиях. — В кн.: Геологическая информация и математическая геология. М., «Недра», 1976, с. 53—66.
- Феллер В. Введение в теорию вероятностей и ее приложения. М. «Мир». Т. I, 1964, 498 с. Т. II, 1967, 752 с.
- Cochran W. G. Sampling Techniques. New York, J. Wiley, 1961. 330 p.
- Vistelius A. B. The skew frequency distributions and the fundamental law of Geochemical processes. — J. Geol., 1960, v. 68, No. 1, p. 1—22.
- Vistelius A. B. Some lessons of G-1-W-1 investigations. — J. Int. Assoc. Math. Geol., 1971, v. 3, No. 3, p. 323—326.

Глава IV

СЛУЧАЙНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ И ИХ МАРКОВСКИЕ МОДЕЛИ

Случайные последовательности исходов испытаний, их различные типы и методы изучения, определения. Марковские цепи, их свойства и связанные с ними расчеты. Исследование марковских цепей матричными методами. Восстанавливающие события, частные и ограниченно марковские переходы. Примеры из геологической практики.

Ключевые слова:

случайная последовательность, цепь Маркова, восстанавливающие события, типы марковских переходов, существенно немарковские последовательности.

IV.1. ВВЕДЕНИЕ

Как известно, одним из основных методов геологии является профилирование и построение разрезов. При этом профили показывают изменение какой-либо характеристики вдоль некоторой линии. Эта характеристика может быть либо составом слоев, чередующихся в разрезе, либо содержанием в породе той или иной химической или физической величины (каротажная диаграмма). Так как указанные профили (или разрезы) являются одним из важнейших материалов, по которым строит заключение геолог, то интересно отметить те специфические черты, которые заставляют пользоваться профилями по крайней мере уже 200 или более лет.

1. Геолога интересует не только само по себе значение изучаемой им характеристики, но и значения характеристик в последовательных пунктах наблюдений, к которым они точно привязаны. Важна не только сама характеристика, но и размещение ее значений, упорядоченных по значениям некоторого параметра (времени, мощности, глубины и т. п.).

2. Для решения задач, где используется профилирование, важно не только знать размещение характеристики по значениям некоторого параметра, но и отчетливо представлять, как ведут себя по отношению друг к другу значения характеристик, находящиеся в соседних точках наблюдений, через точку, через две точки и т. д.

Из сказанного следует, что геологов в целом интересуют некоторые последовательности вида

$$\dots a_I(h-1), a_J(h), a_K(h+1), \dots \quad (\text{IV.1.1})$$

Обозначение $a_J(h)$ читается так: в h -й точке последовательности наблюдалась характеристика a ; эта характеристика приняла некоторое значение J .

Значения, принимаемые характеристикой, могут быть как численными, так и качественными. Если, скажем, изучается чередование глин и песков, то $J \in \{\pi, \gamma\}$. Последняя запись означает — состав слоя мог быть песчаным (π) или глинистым (γ), но не мог быть никаким другим. Индекс h , отмечающий положение точки на профиле, указанный в данном случае в скобках и иногда приводимый как верхний индекс ($a_J^{(h)}$), отсчитывается от начальной точки, которая удобна для данной задачи.

Последовательность вида (IV.1.1) иногда кратко будет изображаться в виде $(a_I(h))$ или $(a_J^{(h)})$. Результат (реализация) отдельного испытания в момент h называется далее *исходом испытания*. Таким образом, исходом может быть появление того или иного значения $I, J, K \dots$, принятого изучаемой характеристикой.

Всякое событие, отнесенное к последовательности, которому мы можем приписать некоторую вероятность (в том числе 0 и 1), будет называться случайным событием. Например, то, что более половины слоев в некоторой пачке сложено песчаными слоями, мы будем называть событием; если же обнаружено, что третий от подошвы слой образован песчаником, то говорится о событии, связанном с исходом третьего испытания.

В зависимости от того, насколько мы можем предсказать появление события в h -й точке наблюдений, зная как реализовались предыдущие испытания, мы можем разделить последовательности на два типа: детерминированные и случайные. *Детерминированной называется последовательность, в которой для любого h можно однозначно предсказать исход испытания, зная предыдущие исходы испытаний.* Если этого сделать нельзя, то в вероятностной модели мы должны довольствоваться только знанием вероятности того, каков исход испытания h , если известны исходы предыдущих испытаний. *Последовательности, у которых исходы испытаний можно предсказать только с некоторой вероятностью, носят название случайных* (см. гл. II).

При характеристике исходов испытаний мы говорили до сих пор о значениях, принимаемых изучаемой характеристикой. В тех случаях, когда характеристика носит качественный характер, результаты исходов испытаний, классифицированные по их качеству, называются *состояниями*. Так, если разрез сложен песками, глинами и известняками без каких-либо промежуточных разностей отложений, то мы говорим о трех состояниях — песках, глинах и известняках. Если горная порода состоит из мелилита,

гаюина и биотита, то можно говорить, что в ее сечении могут быть встречены только три состояния — мелилит, гаюин и биотит. Если изучаемая характеристика может принимать любое значение в некотором интервале (скажем, содержание ванадия в известняке) и мы хотим представить ее в виде чередования небольшого числа

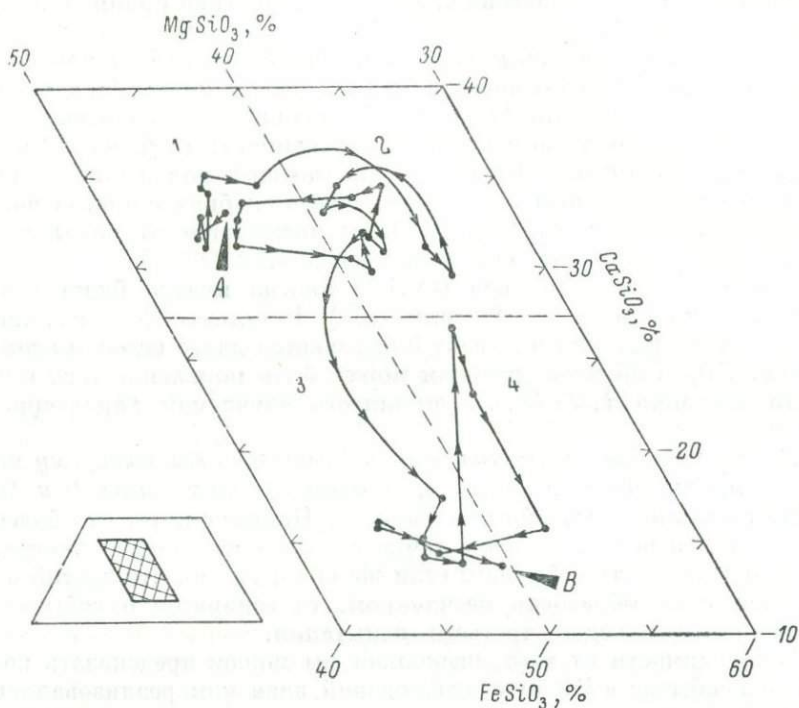


Рис. IV.1. Изменение состава фенокриста пироксена в лунном базальте по мере движения от центральной части (показана жирной стрелкой A) к краю (показан жирной стрелкой B) (Boyd, Smith, 1971).

Измерения через 10 ммк.

состояний, то эта характеристика может быть закодирована и представлена как последовательность кодовых номеров.

Вероятности появления различных состояний могут зависеть от номера испытания. Большую информацию о случайной последовательности несут условные вероятности различных состояний в испытаниях с фиксированным номером относительно состояний в испытаниях с некоторыми соседними номерами. В зависимости от исхода одного или нескольких предыдущих испытаний могут меняться вероятности появления тех или иных состояний в последующем испытании.

Посмотрим, как выглядят некоторые случайные последовательности у типичных геологических объектов,

Пример IV.1. На рис. IV.1 приведена диаграмма, заимствованная из работы Бойда и Смита (Boyd, Smith, 1971), на которой показано изменение состава пироксенов при переходе через периферическую часть вкрапленника пиджонита в лунном базальте через 10 ммк. Из диаграммы виден случайный характер соотношений, особенно внутри клеток, показанных пунктиром. Арифметизируя состав пироксена путем приписки ему номера по принадлежности к одной из клеток, выделенных пунктиром на рис. IV.1, мы получим следующую классификацию:

	En	Wo	Fs
1)	$40 < x < 50$	$27 < x < 40$	$40 < x < 50$
2)	$30 \leq x \leq 40$	$27 < x < 40$	$50 < x < 60$
3)	$40 < x < 50$	$10 < x \leq 27$	$40 < x < 50$
4)	$30 \leq x \leq 40$	$10 < x \leq 27$	$50 < x < 60$

где En — энстатитовый ($MgSiO_3$), Wo — волластонитовый ($CaSiO_3$) и Fs — фоссаитовый ($FeSiO_3$) миналы в пиджоните. Рисунку IV.1 отвечает следующая последовательность:

... 11111111222222211122233334443333 ... (IV. 1. 2)

Из вида последовательности легко сделать заключение о ее гетерогенном (неоднородном) строении. Вероятности состояний в ней сильнейшим образом зависят от номера испытания в последовательности.

Пример IV.2. Случайная последовательность типов железных метеоритов по дате их падения за период с 1835 г. по 1947 г. имеет следующий вид (Кринов, 1948):

... OGOOAOOOOOOOOGOGOOOOOOGOOOOAOOOOOOAOO ... (IV. 1. 3)

Здесь O — октаэдрит, Г — гексаэдрит и А — атаксит. Из последовательности видно, что в ней резко доминируют октаэдриты. Никакой зависимости типа железного метеорита от даты падения не заметно. Скорее всего это случайная последовательность, в которой любое последующее испытание не зависит от предыдущих.

Пример IV.3. На рис. IV.2 приведена структура слюды типа мусковита (Дир и др., 1966). Основной мотив этой структуры определяется чередованием трех типов слоев. При этом первый слой образован кремнекислородными тетраэдрами с вершинами, повернутыми вверх (слой А), затем идет слой из октаэдрически упакованных кислородов, в центре октаэдров которых располагаются атомы алюминия или магния (слой В), затем снова идет слой кремнекислородных тетраэдров (А), вершины которых обращены к октаэдрическому слою. Пакеты слоев, образованных указанным способом, соединены между собой атомами калия, располагающимися во впадинах в основании кремнекислородных тетраэдров. При этом каждый атом калия принадлежит одновременно двум тетраэдрам из разных пакетов. Указанная структура характеризуется непрочным положением атомов К. Они могут

изоморфно замещаться, скажем, Rb, на их место может становиться Na и т. п.

Наконец, возможна гидратация Mu, при которой затрагивается структура пакета АВА, а места, занятые К, частично случайно замещаются либо гидроксонием (при образовании гидромусковита), либо H₂O (при образовании монтмориллонита). Это приводит к большому разнообразию стохастических структур при изучении последовательностей структурных элементов слюд в направлении, перпендикулярном спайности. При этом основной

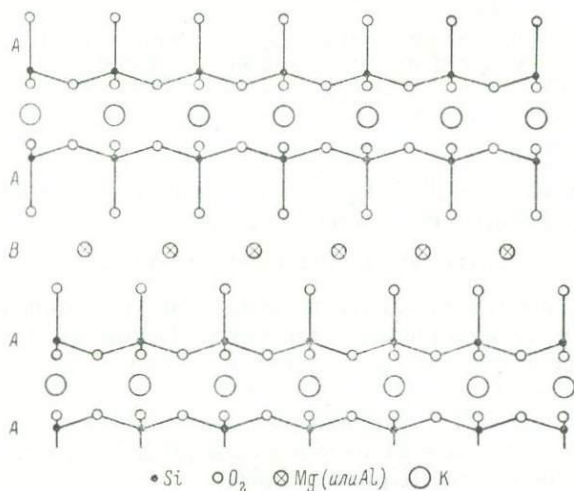


Рис. IV.2. Структура слюды.

Вид вдоль оси y . Рубидий или гидроксоний, замещающие К, в примере IV.3 обозначены \mathcal{E}_i .

структурный мотив — при стохастическом его рассмотрении — может быть представлен случайной последовательностью типа

$$\dots, \Pi, \mathcal{E}_1, \Pi, \mathcal{E}_2, \Pi, \dots, \quad (IV.1.4)$$

где Π — детерминированно построенный пакет вида АВА, а \mathcal{E}_i — калий и различные элементы, частично становящиеся случайным образом на его место. В гидромусковите ведущую роль при этом в $\{\mathcal{E}_1, \mathcal{E}_2, \dots\}$ играют случайно чередующиеся К и гидроксоний.

Таким образом, для гидромусковита в первом приближении последовательность (IV.1.4) оказывается случайной последовательностью на трех состояниях (Π, \mathcal{E}_1 и \mathcal{E}_2). Зная, что имеет место Π , мы можем вынести только вероятностное суждение о появлении К или Н в следующем испытании. Зная, что наступило одно из \mathcal{E}_i , мы детерминированно предсказываем появление Π . Сведения о более ранних результатах испытаний не влияют на это предсказание.

Если перенумеровать состояния и развернуть структуру П, введя пакет АВА на место П, то выявятся новые черты случайной последовательности. Действительно, в последовательности

$$\dots, АВА\mathcal{E}_1АВА\mathcal{E}_2АВА\mathcal{E}_2, \dots$$

для определения перехода с В или с \mathcal{E}_i достаточно знания только того, что в настоящий момент реализовалось В или \mathcal{E}_i . Однако, когда мы переходим с А, то существенно знание предшествующего исхода испытания. Если перед А было случайное событие \mathcal{E}_i , то переход на В детерминирован. Если перед А было В, то переход на \mathcal{E}_i носит вероятностный характер, а переход на В запрещен. Таким образом, событие, наступившее после А, существенно зависит от того, что было перед А.

Наконец, мы можем детализировать состояния далее. При этом можно различать А в зависимости от положения отвечающих им кремнекислородных тетраэдров по отношению к направлению движения поперек спайности (направление вершины тетраэдра). В этом случае мы получим последовательность на пяти состояниях A_1, A_2, B и $\mathcal{E}_1, \mathcal{E}_2$ (условно мы принимаем, что \mathcal{E} случайно принимает значения либо К, либо Н); A_1 всегда лежит в основании пакета, A_2 — его заключает. В этом случае в последовательности

$$\dots, A_1BA_2\mathcal{E}_1A_1BA_2, \mathcal{E}_2, \dots$$

снова появляется специфическое свойство — знание настоящего целиком определяет вероятностную информацию о будущем.

Рассматривая приведенные примеры, можно обнаружить, что фигурирующие в них последовательности имеют принципиально различно определенные интервалы между двумя соседними испытаниями. В примере IV.1 расстояние между испытаниями (10 ммк) было выбрано совершенно произвольно: с таким же успехом можно было брать интервалы 9 или 11 ммк, это вопрос реализации эксперимента. В примере IV.2 имелась только упорядоченность в смысле событий до и после, но никакого фиксированного интервала времени между ними задано не было. Наконец, в примере IV.3 мы имели расстояния между элементами последовательности, заданные природными соотношениями, т. е. структурой слюды.

При работе с природными объектами могут встретиться случаи, когда придется столкнуться с ранжировкой случайных событий по любому из изучаемых признаков. Однако, когда дело идет о построении случайных последовательностей как реализаций моделей, всегда нужно стремиться, чтобы интервал между событиями имел какой-то внутренний смысл, связанный с явлением. Если это сделать невозможно, то нужно тщательно продумать технику анализа последовательности, возникающей на искусственно заданных интервалах.

IV.2. ВЕРОЯТНОСТНЫЕ СТРУКТУРЫ МАРКОВСКИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Ниже дается краткая характеристика случайных последовательностей вообще и относительно детально рассматривается структура марковских цепей.

IV.2.1. Случайные последовательности и связанные с ними вероятности

Предположим, что параметр t , по которому производится упорядочение процесса, является одномерной величиной. Рассматриваемые наблюдения $\chi(t)$ предположим также одномерными. При этом исследуется случай, при котором как t , так и $\chi(t)$ дискретны, т. е. параметр t последовательно принимает значения из ряда $0, 1, 2, \dots$. Наблюдения $\chi(t)$ также принимают значения из некоторого дискретного множества \mathcal{S} , содержащего лишь конечное число элементов s . Множество \mathcal{S} называется множеством состояний случайной последовательности. Элементы множества \mathcal{S} называются состояниями последовательности. При этом безразлично, будут ли состояния числами, качественными признаками или объектами еще какой-либо природы. Важно только, чтобы любое состояние всегда одинаково идентифицировалось.

Последовательность $0, 1, 2, \dots$ значений параметра t может быть конечной или бесконечной. В случае конечной последовательности мы имеем дело с многомерным вектором, который часто называют конечным случайным процессом. Конечную случайную последовательность можно полностью охарактеризовать с помощью так называемого дерева (Кемени, Снелл, 1970).

Рассмотрим, например, описание пачки из четырех слоев, которые могут состоять либо из песка (π), либо из глины (γ). Пачка может начинаться как с π , так и с γ . Это начальные состояния для дерева. В результате следующего испытания может снова появиться случайное событие π или γ , и т. д. После третьего испытания появляется четвертый слой, и этим процесс заканчивается (так как в нашем примере пачка состояла из четырех слоев). Каждая возможная четверка слоев представляет собой отдельную реализацию или «ветку дерева». Употребляется также термин «траектория». В целом в нашем примере получается дерево, изображенное на рис. IV.3.

Чтобы получить вероятностное пространство (Ω, \mathcal{F}, P) (см. гл. II) для конечной случайной последовательности, достаточно задать вероятность каждой отдельной ветви. Например, для рассмотренной пачки слоев достаточно задать вероятность каждой из 16 возможных ветвей. Эти вероятности могут быть произвольными, но их сумма должна быть равна единице. Так, в нашем примере можно определить случайную последователь-

ность следующим заданием вероятностей ветвей (см. рис. IV.3). Для первой и последней траектории (ветви, реализации) примем нулевые вероятности, а остальным припишем вероятности, равные для каждой из оставшихся ветвей, т. е. $1/14$. Это соответствует такому представлению о формировании случайной пачки, при котором не могут возникать пачки, целиком сложенные только песчаными или только глинистыми слоями, и при котором состояния π и γ симметричны. Каждая траектория представляет собой элементарное событие из пространства элементарных событий, т. е. точку в пространстве Ω (см. гл. II). Размерность вектора, отвечающего конечному случайному процессу, в данном случае равна четырем, пространство же Ω элементарных событий содер-

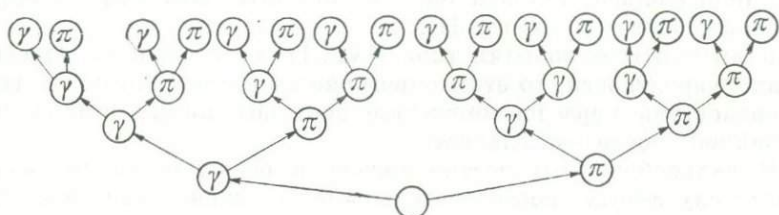


Рис. IV.3. Вероятностное дерево.

Внизу условно показано общее начало (последний слой перед рассматриваемой пачкой в рассматриваемом примере). Вероятности конкретных переходов зависят от структуры вероятностного пространства.

жит 16 точек. Вероятность какого-либо случайного события $A \subset \Omega$ (см. гл. II), охарактеризованная принятыми условиями формирования рассмотренной пачки, всегда может быть найдена с помощью операций объединения, дополнения и пересечения элементарных событий, как это было показано в гл. II.

Рассмотрим, например, такое событие A . Случайная пачка содержит не менее одного и не более двух песчаных слоев. В то же время эта пачка содержит какую-либо серию из слоев произвольного состава (т. е. песчаных или глинистых) длины 2.

Чтобы получить вероятность A , нужно сложить вероятности всех ветвей, где это событие осуществилось. Оно реализуется, как видно из рис. IV.3, лишь для 4, 7, 10, 12, 13 и 14-й ветвей, считая ветви слева направо (реализаций). Таким образом, $p(A) = 3/7$.

Аналогично можно определить вероятность любого случайного события, связанного с формированием рассматриваемой пачки слоев. Итак, вероятностная мера P определена.

В случае бесконечной последовательности (с бесконечным множеством шагов) задание вероятностной меры с помощью перечисления вероятностей всех ветвей невозможно или нерационально. Здесь мера задается обычно другим способом.

Рассмотрим событие

$$\{a_I(t_1) a_J(t_2), \dots, a_K(t_{k-1}), a_L(t_k)\}, \quad (IV. 2.1)$$

означающее, что в моменты времени t_1, t_2, \dots, t_k (не обязательно смежные) реализовались соответственно состояния I, J, \dots, K, L . Пусть для каждого события типа (IV.2.1), т. е. для любого конечного набора $t_1, \dots, t_k \in \{0, 1, 2, \dots\}$ и для любого набора $I, J, \dots, K, L \in \mathcal{S}$, определена вероятность

$$p \{a_I(t_1), a_J(t_2), \dots, a_K(t_{k-1}), a_L(t_k)\}. \quad (\text{IV. 2. 2})$$

Очевидно, что все вероятности (IV.2.2) нельзя указать перечислением, поскольку k может быть сколь угодно велико. Такие вероятности задаются в виде формул с параметром k . Очевидно также, что вероятности разных событий вида (IV.2.2) нельзя задавать произвольно, так как они должны быть согласованы определенным образом (см. гл. II).

Если для всех событий типа (IV.2.1) согласованным образом заданы вероятности, то этим определяется единственным образом вероятностная мера на множестве всех возможных реализаций случайной последовательности.

В дальнейшем мы рассматриваем в основном специальный класс случайных последовательностей — марковские последовательности. Для этих последовательностей вероятности вида (IV.2.2) вводятся так, что согласованность и однозначное задание вероятностей меры оказываются, очевидно, выполненными.

Простейшим событием вида (IV.2.1) является $\{a_I(h)\}$, т. е. событие, заключающееся в том, что в h -м испытании или, что то же, после $(h-1)$ -го шага в случайной последовательности реализуется состояние I . Вероятность $p \{a_I(h)\}$, которую мы будем сокращенно обозначать $p_I(h)$, часто называется *безусловной вероятностью состояния I в момент h* . Если фиксировать h и собрать такие вероятности для всех $I \in \mathcal{S}$, то получим вектор с s компонентами (количество элементов во множестве состояний \mathcal{S}), представляющий дискретное распределение вероятностей исходов в h -м испытании. Этот вектор обозначается далее $p^{(h)}$. В частности, $p^{(0)}$ представляет *начальное распределение вероятностей для случайной последовательности*.

Следующим событием из (IV.2.1) по простоте строения после $\{a_I(h)\}$ является событие вида $\{a_I(h), a_J(h+r)\}$ ($r > 0$), т. е. событие, заключающееся в том, что в h -м испытании появилось состояние I , а затем через r шагов наступило состояние J . Вероятность такого события мы будем обозначать как $\pi_{I;J}(h, h+r)$. Если рассмотренная нами ранее безусловная вероятность первого состояния в момент h , т. е. $p_I(h)$, больше нуля, то, как отмечалось в гл. II, можно найти условную вероятность того, что в $h+r$ -м испытании появится состояние J при условии, что в h -м испытании появилось состояние I . Такая условная вероятность называется *переходной вероятностью из состояния I в состояние J для моментов h и $h+r$* . Эту вероятность мы обозначаем как $p_{I;J}(h, h+r)$.

Как известно, эта вероятность вычисляется с помощью формулы

$$p_{I;J}(h, h+r) = \frac{\pi_{I;J}(h, h+r)}{p_I(h)}. \quad (\text{IV.2.3})$$

Если $p_I(h) = 0$, то переходная вероятность (IV.2.3), как отмечалось в гл. II, будет неопределенной. Если фиксировать h и r , то все переходные вероятности вида (IV.2.3) можно собрать в матрицу $s \times s$ с общим элементом $p_{I;J}(h, h+r)$ при $I, J \in \mathcal{S}$, $h \in \{0, 1, 2, \dots\}$, $r > 0$. Эту матрицу мы будем обозначать $P^{(h; h+r)}$. Каждая строчка этой матрицы представляет некоторое условное дискретное распределение вероятностей.

Задание вероятностей $p_I(h)$ для всех $I \in \mathcal{S}$ и $h \in \{0, 1, 2, \dots\}$, очевидно, еще не определяет вероятностную меру для случайной последовательности. Это значит, что существуют случайные события, отнесенные к случайной последовательности, и такие, что для их вычисления недостаточно знания только $p_I(h)$. Однако имеются случайные последовательности, для которых все определяется вероятностями $p_I(h)$. К таким последовательностям относятся изучаемые далее последовательности независимых испытаний.

Точно так же задание всех $p_I(h)$ и всех $p_{I;J}(h; h+1)$ еще не определяет случайной последовательности. Однако имеется важный для геологии класс последовательностей, для которых задание этих вероятностей полностью определяет случайную последовательность. К этому классу относятся случайные последовательности, называемые простыми цепями Маркова. Эти последовательности мы будем подробно изучать далее.

IV.2.2. Стационарность, однородность и обратимость

При изучении случайной последовательности важно знать, насколько ее строение является неизменным на всем протяжении. В связи с этим вводятся понятия стационарности и однородности. Для этого рассмотрим вместе с некоторым конечным набором точек t_1, t_2, \dots, t_k набор точек $t_1+r, t_2+r, \dots, t_{k+r}$, сдвинутый вправо на r точек (напомним, что t_1, t_2, \dots, t_k являются числами из $0, 1, 2, \dots$ и что r — целое положительное число). Если вероятность любого события вида (IV.2.1) не зависит от сдвига, т. е. если

$$\begin{aligned} & p[a_I(t_1), a_J(t_2), \dots, a_K(t_{k-1}), a_L(t_k)] = \\ & = p[a_I(t_1+r), a_J(t_2+r), \dots, a_K(t_{k-1}+r), a_L(t_k+r)], \end{aligned} \quad (\text{IV.2.4})$$

то случайная последовательность называется стационарной.

Равенство (IV.2.4) предполагается выполненным для любого конечного k при произвольных t_1, t_2, \dots, t_k , каких угодно $I, J, K, \dots, L \in \mathcal{S}$ и любого целого r . Наличие стационарности влечет отсутствие какого-либо тренда во времени.

Из требования стационарности, в частности, следует, что

$$p_I^{(0)} = p_I^{(1)} = \dots = p_I^{(k)} = p_I, \quad I \in \mathcal{S}, \quad k \geq 1,$$

т. е. вектор, представляющий распределение вероятностей исходов, один и тот же в любой момент времени. Такой постоянный вектор и отвечающее ему распределение называют стационарными.

Переходные вероятности стационарной последовательности, т. е. $p_{I;J}(h, h+r)$, не меняются при изменении h . Эту вероятность мы будем обозначать $p_{I;J}(r)$. Соответственно матрица переходных вероятностей обозначается $P^{(r)}$.

Стационарность является сильным ограничением. Однако подчас неизменность строения последовательности имеет место в более широком смысле слова. В этом случае используется понятие «однородность». Это понятие вводится по-разному в зависимости от типа строения случайной последовательности. В дальнейшем определение понятия «однородность» будет отнесено к марковским последовательностям, в связи с чем возникает термин «однородная марковская последовательность».

Стационарные последовательности однородны, но однородные необязательно стационарны.

Как наличие стационарности, так и однородности у случайной последовательности часто с разумным приближением можно гарантировать заранее. Так, например, последовательность зерен различных минералов одного удельного веса, возникшую в центральной части раскристаллизовавшегося гомогенного расплава, естественно полагать стационарной, так как она статистически удовлетворяет (IV.2.4). Исследование первично магматических гранитов во многих случаях подтверждает эту точку зрения (Иванов, 1978; Романова, 1978). Примерами последовательностей, которые могут считаться однородными, наблюдаемыми непосредственно на геологических объектах, могут служить разрезы флиша Южного Урала и северо-западного Кавказа, которые не удовлетворяют соотношению (IV.2.4) при $k=1$ и удовлетворяют ему при остальных $k=2, 3, \dots$. Это происходит из-за ритмического строения флишевых толщ, при котором вероятность появления, скажем, глинистого слоя резко различна в начале и в конце ритма. Последовательности, обладающие отмеченными свойствами, могут в определенном смысле считаться однородными, но не стационарными.

Неоднородные последовательности также достаточно обычны. Пример, уже приведенный нами при разборе зональности лунных пироксенов, очень типичен для петрологии. Чередование слоев различного состава, сформированных в пределах нормального фациального ряда трансгрессирующей серии, дает неоднородную последовательность. Создается впечатление, что оценка однородности чередования слоев в разрезах осадочных толщ могла бы быть весьма эффективно использована при разделении осадочных толщ на различные типы формаций.

Другой важной характеристикой случайной последовательности является обратимость. Если для всех событий типа (IV.2.1) (конечные наборы) в стационарной последовательности имеет место

$$p[a_I(t_1), a_J(t_2), \dots, a_K(t_{k-1}), a_L(t_k)] = \\ = p[a_L(T-t_k), a_K(T-t_{k-1}), \dots, a_J(T-t_2), a_I(T-t_1), (T \geq t_k), (IV.2.5)$$

то случайная последовательность называется обратимой. Сочетание исходов испытаний в правой части равенства (IV.2.5) получается из сочетания исходов испытаний в левой части равенства, если в сочетании из левой части равенства испытание a_L , происшедшее на месте t_k , поместить на любое место $T-t_k$, а затем через соответствующие интервалы разместить и другие испытания. Иными словами, сочетание в (IV.2.5) справа для конечномерного набора с неравными временными интервалами является тем же самым, что и сочетание (IV.2.5) слева, прочитанное в обратном порядке.

Например, пусть мы имеем обратимую стационарную последовательность с состояниями π , α и γ . Тогда вероятность набора, скажем, $p[a_\pi(8) a_\gamma(17) a_\alpha(21)]$ должна быть равна $p[a_\gamma(8) a_\alpha(12) a_\pi(21)]$. В силу того что сдвиг не меняет вероятностей событий в однородной последовательности, ту же вероятность $p[a_\pi(8) a_\gamma(17) a_\alpha(21)]$ должно иметь и всякое событие $p[a_\alpha(r), a_\gamma(r+4), a_\pi(r+13)]$.

Случайная последовательность, все реализации которой даются в обратном направлении, порождает новую последовательность, называемую *обращенной*. Для последовательности, бесконечной в обе стороны, обратимость означает, что все вероятностные характеристики обращенной последовательности совпадают с соответствующими характеристиками исходной последовательности.

Существует ряд геологических явлений, где заранее можно ожидать наличия обратимости. Большой опыт исследования последовательностей зерен калиевого полевого шпата, кварца и плагиоклаза в гранитах показывает, что эти последовательности почти всегда обратимы. В то же время, скажем, последовательность слоев в кавказском флише, очевидно, необратима, так как вероятность $p[a_\pi(h) a_\gamma(h+1)]$, где π — песчаный, а γ — глинистый слой, заведомо близка к единице, а $p[a_\gamma(h) a_\pi(h+1)]$ — близка к нулю (Вистелнус, 1949).

IV.2.3. Марковские последовательности

Рассмотрим r последовательных моментов времени

$$h-r, h-r+1, \dots, h-2, h-1.$$

Назовем набор этих моментов времени «настоящим». Мы знаем, что произошло в настоящем, если известны исходы испытаний, отвечающие этим моментам времени, иными словами, если мы

знаем, какое конкретное событие $\{a_I(h-r), a_J(h-r+1), \dots, a_K(h-2), a_L(h-1)\}$ реализовалось в это время.

О любом событии, связанном лишь с исходами испытаний после $h-1$ -го, мы будем говорить, что оно относится к «будущему». Для таких событий из будущего мы используем обозначения вида $B \in \mathfrak{B}_h$. Всякое событие, зависящее лишь от исходов испытаний, с номерами, меньшими, чем $h-r$, мы будем относить к «прошлому» и использовать для обозначения таких событий обозначения вида $A \in \mathfrak{A}_{h-r-1}$.

Пусть \mathfrak{A}_{h-r-1} — это σ -алгебра событий, связанных с ходом последовательности, до момента $h-r-1$ включительно, а \mathfrak{B}_h — σ -алгебра событий, связанных с ходом последовательности, начиная с момента h (включая h). *Случайная последовательность называется марковской цепью r -го порядка, если соотношение*

$$P[B | a_I(h-1), a_J(h-2), \dots, a_K(h-r+1), a_L(h-r); A] = \\ = P[B | a_I(h-1), a_J(h-2), \dots, a_K(h-r+1), a_L(h-r)] \quad (\text{IV.2.6})$$

выполняется для любых событий $B \in \mathfrak{B}_h$, $A \in \mathfrak{A}_{h-r-1}$ при всех целых $h \geq r$ и каких угодно состояниях $I, J, \dots, K, L \in \mathcal{S}$ (вертикальная черта отделяет условие). Следует иметь в виду, что событие B из будущего может быть отнесено к любой совокупности исходов испытаний с номерами $l \geq h$. Такому событию может отвечать исход одного испытания, например $a(h)$ или $a(h+10)$. Подобному событию может отвечать исход нескольких испытаний. Наконец, это событие может охватывать исходы бесконечного числа испытаний. Подобными способами может быть выбрано также событие из прошлого A .

Вместо термина «марковская цепь r -го порядка» можно говорить о случайной последовательности с марковским свойством (марковостью) r -го порядка ($r=0, 1, \dots$).

При $r=0$ марковская цепь называется последовательностью независимых испытаний (последовательностью Бернулли; иногда также такие последовательности называют марковскими нулевого порядка). При $r=1$ цепь называется простой цепью Маркова или цепью Маркова первого порядка. При $r > 1$ цепь называется сложной. Так, при $r=2$ говорят, что имеется сложная цепь Маркова второго порядка. Аналогично можно говорить о цепях третьего, четвертого и так далее порядков.

Из равенства (IV.2.6) очевидно, что цепь порядка r является также цепью любого порядка, большего, чем r . Действительно, если соотношение (IV.2.6) справедливо при некотором r , то оно будет справедливо при любом $m > r$, поскольку все испытания с номерами от $h-r-1$ до $h-m$ можно отнести к событию $A \in \mathfrak{A}_{h-r-1}$. Так, в частности, цепь нулевого порядка является также цепью первого порядка, цепь первого порядка — цепью второго, третьего и так далее порядков, однако цепь первого порядка может не быть цепью нулевого порядка. Следует подчеркнуть,

что указанная терминология во многих случаях оказывается неудобной в силу ее неопределенности. Поэтому часто полезно называть порядок марковости по наименьшему возможному порядку. Так, говоря о цепи Маркова первого порядка, можно учитьвать соглашение, что она не является последовательностью Бернулли. В будущем мы часто будем пользоваться именно такой терминологией.

Марковость r -го порядка означает, что знание исходов r случившихся подряд испытаний полностью определяет все дальнейшее предсказание вероятностных свойств случайной последовательности. Термин «предсказание» употребляется здесь не в том смысле, что знание настоящего обеспечивает детерминированное предсказание будущего, оно лишь позволяет дать настолько хорошее вероятностное предсказание будущего, что это предсказание не может быть улучшено никакой дополнительной информацией, исходящей от прошлого.

Детерминированное предсказание также относится к марковскому предсказанию как частный случай.

Формула (IV.2.6), выражающая марковское свойство, дана в таком виде, при котором «будущее» и «прошлое» входят в нее несимметрично. Однако этому соотношению можно придать симметричный вид:

$$\begin{aligned} p[BA | a_I(h-1), a_J(h-2), \dots, a_L(h-r)] = \\ = p[B | a_I(h-1), a_J(h-2), \dots, a_L(h-r)] \times \\ \times p[A | a_I(h-1), a_J(h-2), \dots, a_L(h-r)], \end{aligned} \quad (\text{IV. 2. 7})$$

где $B \in \mathfrak{B}_h$, $A \in \mathfrak{A}_{h-r-1}$.

Формула (IV.2.7) интерпретируется так: будущее и прошлое независимы при фиксированном настоящем. Подчеркнем, что (IV.2.7) не отражает каких-либо новых особенностей случайной последовательности, а является лишь легко выводимым следствием из (IV.2.6) и основных свойств безусловных и условных вероятностей.

Простейшими событиями, фигурирующими в (IV.2.7), являются события вида

$$\begin{aligned} \{a_M(h)\}; \{a_M(h) | a_I(h-1)\}; \{a_M(h) | a_I(h-1), a_J(h-2)\}; \dots \\ \dots; \{a_M(h) | a_I(h-1), \dots, a_K(h-r+1)\}; \\ \{a_M(h) | a_I(h-1), \dots, a_K(h-r+1), a_L(h-r)\}. \end{aligned} \quad (\text{IV. 2. 8})$$

События указанного вида получают, если в качестве будущего взять одно испытание и не рассматривать прошлого. Если задать вероятности всех событий из (IV.2.8) для любых $h \geq r$ и каких угодно $I, J, \dots, L, M \in \mathcal{S}$, т. е. числа

$$\begin{aligned} p_M(h); p_I; p_M(h-1, h), p_I, J; p_M(h-2, h-1, h), \dots \\ \dots, p_{\underbrace{L, K, \dots, J, I}_r}; p_M(h-r, h-r+1, \dots, h-1, h), \end{aligned} \quad (\text{IV. 2. 9})$$

где

$$p_{L, K, \dots, J, I; M}(h-r, h-r+1, \dots, h-1, h) = \\ = p(a_M(h) | a_I(h-1), \dots, a_L(h-r)),$$

то вероятностная мера, связанная с марковской цепью r -го порядка, будет определена единственным образом. Для того чтобы убедиться в этом, рассмотрим сначала конечные наборы состояний, отнесенные к смежным моментам времени. Тогда вероятностью реализации, состоящей из одного испытания в момент h , будет вероятность вида $p_M(h)$. Вероятность реализации, состоящей из двух последовательных испытаний, может быть вычислена с помощью чисел из (IV.2.9) как

$$p_I(h-1) p_{I; M}(h-1, h). \quad (\text{IV. 2.10})$$

Далее, вероятность реализации, состоящей из трех последовательных испытаний, будет рассчитываться как

$$p_J(h-2) p_{J; I}(h-2, h-1) p_{J, I; M}(h-2, h-1, h). \quad (\text{IV. 2.11})$$

Для реализации из $r+1$ последовательных испытаний соответствующую вероятность вычисляем как

$$p_L(h-r) p_{L; K}(h-r, h-r+1) \dots \underbrace{p_{L, K, \dots, J, I; M}}_r(h-r, \\ h-r+1, \dots, h-1, h), \quad (\text{IV. 2.12})$$

где все вероятности, входящие в (IV.2.12), берутся из (IV.2.9).

Формулы (IV.2.10)–(IV.2.12) дают вероятности наборов испытаний длины от 2 до $r+1$. Они построены путем учета свойств условных и безусловных вероятностей и не используют марковского свойства (IV.2.6). Для того чтобы выразить вероятность набора испытаний длины $r+2$ через вероятности (IV.2.9), уже можно использовать марковское свойство. Очевидно, что

$$p[a_K(h-r-1), a_L(h-r), \dots, a_J(h-2), a_I(h-1), a_M(h)] = \\ = p[a_M(h) | a_I(h-1) a_J(h-2), \dots, a_K(h-r-1)] \times \\ \times p[a_K(h-r-1), \dots, a_I(h-1)]. \quad (\text{IV. 2.13})$$

Из-за наличия марковского свойства порядка r испытание $a_K(h-r-1)$ в условии из (IV.2.13) можно отбросить. Что касается второго множителя в (IV.2.13) справа, представляющего вероятность набора длины $r+1$, то соответствующую вероятность мы уже изобразили формулой (IV.2.12).

Таким образом, получаем

$$p[a_K(h-r-1), a_L(h-r), \dots, a_J(h-2), a_I(h-1), a_M(h)] = \\ = p_K(h-r-1) p_{K; L}(h-r-1, h-r) \dots \underbrace{p_{K, L, \dots, J, I; M}}_r(h-r-1, \\ h-r, \dots, h-2, h-1, h). \quad (\text{IV. 2.14})$$

Марковская цепочка (IV.2.14) состоит только из вероятностей, взятых из (IV.2.9). Очевидно, что аналогично, через вероятности из (IV.2.9), можно выразить вероятность любого конечного набора последовательных испытаний. Произвольный конечный набор отличается от только что рассмотренных наборов тем, что t_1, t_2, \dots, t_k не обязаны представлять смежные моменты времени. Для вычисления вероятности такого набора в него включаются все пропущенные моменты времени и для каждого пропущенного момента времени осуществляется суммирование по всему множеству состояний. Так, например,

$$p [a_I(5) a_J(7) a_K(10)] = \\ = \sum_{s \in \{\mathcal{S}\}, t \in \{\mathcal{S}\}, w \in \{\mathcal{S}\}} p [a_I(5) a_S(6) a_J(7) a_T(8) a_W(9) a_K(10)].$$

В итоге мы видим, что задание вероятностей (IV.2.9) определяет вероятности любых конечных наборов для марковской цепи порядка r . Эти вероятности, как указывалось на с. 226, согласованы, так как они построены по формулам, связывающим условные вероятности с безусловными. Поэтому, согласно утверждению на с. 232, задание вероятностей (IV.2.9) единственным образом определяет вероятностную меру для марковской цепи порядка r .

Легко также показать, что при таком задании вероятностей имеют место (IV.2.6) и (IV.2.7).

Рассмотрим теперь, что означает стационарность для марковской цепи. Согласно определению стационарности все вероятности не должны зависеть от сдвига h . Поэтому индекс h в этом случае можно опустить. Длину звена r также можно не отмечать отдельно (она видна из нижней индексировки по количеству указанных там индексов).

Таким образом, в случае стационарной цепи Маркова вероятности (IV.2.9) могут быть представлены в виде

$$P_M; P_I; M; \dots; P_{L, K}, \dots, J, I; M.$$

Итак, *стационарная марковская цепь — это последовательность, у которой безусловные и переходные вероятности вида (IV.2.9) не зависят от времени h* . Нужно подчеркнуть, что вероятности $p_M(h)$ в стационарной последовательности не меняются при изменении времени h (являются стационарными). В частности, совпадающим со стационарным должно быть начальное распределение. Чтобы ослабить это очень стеснительное условие, для марковских цепей вводится обычно понятие однородности. *Марковская цепь r -го порядка называется однородной, если все переходные вероятности в (IV.2.9) не зависят от времени*. Начальное распределение при этом может быть произвольным распределением вероятностей. Безусловные вероятности $p_M(h)$ могут зависеть от начального распределения (см. распределение вероятностей исходов в начальный и последующие моменты времени в примере

IV.2). Приведенные определения стационарности и однородности различаются лишь по отношению к начальному распределению.

Наличие однородности упрощает запись формул. Так, марковская цепочка (IV.2.14) в случае стационарности приобретает вид $P_K P_{K;L} \dots P_{K, \dots, J, I; M}$, а в случае однородности выглядит как $P_K (h - r - 1), P_{K;L} \dots P_{K, \dots, J, I; M}$.

Всякую стационарную цепь порядка $r > 1$ можно свести к эквивалентной цепи порядка 1, увеличивая количество состояний. Если цепь является однородной, но не стационарной, то такое приведение сложной цепи к простой ведет к потере некоторой части информации, имеющейся в исходной сложной цепи. Причина этого будет разъяснена далее (на с. 277 и следующих). В то же время из этого разъяснения будет видно, что основные черты поведения сложной цепи отражаются простой цепью при увеличенном числе состояний независимо от того, каким типом однородности обладала исходная цепь.

Прием, с помощью которого сложную цепь сводят к простой, и вопрос о том, какая информация не охватывается вновь полученной простой цепью, будут рассмотрены в п. IV.3.6.

Приведенный в указанном пункте материал показывает, что принципиального различия между простой и сложной марковской цепями с точки зрения их математического содержания в принципе не существует. По этой причине большинство монографий ограничивается рассмотрением одних только простых цепей, которым противопоставляются не сложные цепи Маркова, а существенно немарковские последовательности.

Случайная последовательность называется существенно немарковской, если она не обладает марковским свойством никакого конечного порядка r . Таким образом, существенно немарковская последовательность имеет то свойство, что все прошлое, даже самое отдаленное, влияет в вероятностном смысле на будущее.

Как показывает опыт, представление геологических явлений в терминах марковских цепей (марковского последствия) приводит к открытию новых черт геологических объектов, порожаемых этими явлениями. При этом оказывается важным анализ деталей строения сложных цепей. В зависимости от продолжительности настоящего (r) выделяется ряд классов марковских цепей различного порядка.

IV.2.4. Последовательности независимых испытаний и простые цепи Маркова

Как отмечалось, при $r=0$ случайная последовательность называется последовательностью независимых испытаний или последовательностью Бернулли. Марковское свойство (IV.2.6) применительно к последовательности Бернулли имеет вид:

$$p(B|A) = p(B) \quad (B \in \mathfrak{B}_h, A \in \mathfrak{A}_{h-1}), \quad (\text{IV. 2. 15})$$

т. е. любое событие в будущем независимо от любого события в настоящем и прошлом.

Из вероятностей (IV.2.9), определяющих марковскую цепь для последовательности независимых испытаний, имеют значение лишь безусловные вероятности $p_M(h)$, $M \in \mathcal{S}$, остальные вероятности из (IV.2.9) совпадают с безусловными. Так, например, $p_{J, I; M}(h-2, h-1)$, $h = p_M(h)$, поскольку в силу (IV.2.15) вероятность последнего исхода в тройке последовательных испытаний не может зависеть от предыдущих исходов.

Вероятность марковской цепочки в случае последовательности Бернулли рассчитывается особенно просто. Так,

$$p[a_L(h-r), a_T(h-r+1), \dots, a_J(h-1), a_I(h)] = \\ = p_L(h-r) p_T(h-r+1) \dots p_I(h).$$

В случае последовательности независимых испытаний условные вероятности не зависят от условия. Поэтому понятие «однородная цепь» здесь оказывается бессодержательным. В то же время понятие «стационарность» означает, что безусловные вероятности $p_I(h)$ не должны зависеть от времени h . Иными словами, $p_I(h) = p_I$.

Термины «однородность» и «стационарность» не используются при характеристике последовательностей Бернулли. Вместо этого говорят о последовательностях Бернулли с постоянными вероятностями состояний (это отвечает стационарности) и о последовательностях Бернулли с вероятностями состояний, зависящими от времени (это отвечает нестационарному случаю).

Последовательность Бернулли с постоянными вероятностями всегда обратима, в то же время последовательность независимых испытаний с вероятностями, зависящими от времени, всегда необратима.

Последовательность независимых испытаний, более чем какой-либо другой объект, отвечает интуитивному представлению о чистой случайности. Характерной чертой такой последовательности является так называемая невозможность системы. Она заключается в следующем.

Рассмотрим бесконечную последовательность Бернулли с постоянными вероятностями состояний. Из этой исходной последовательности отберем элементы с некоторыми номерами и из них составим новую бесконечную подпоследовательность. Правила выбора элементов подпоследовательности таковы: до первого испытания нужно решить, включать его в подпоследовательность или нет. После того как сделано первое испытание, но еще не реализовано второе, нужно решить, включать ли второе испытание в последовательность или нет. Вообще, зная результат первых k испытаний, мы должны до реализации $k+1$ -го испытания решить, включать или нет это испытание в подпоследовательность.

Таким образом, решение должно приниматься лишь с учетом известного прошлого. В остальном правило отбора является произвольным и может меняться для различных номеров испытаний. «Невозможность системы» означает, что любая подпоследовательность, построенная по изложенному методу, является снова последовательностью Бернулли с теми же вероятностями состояний, что и в исходной последовательности (Феллер, 1964, с. 203—205).

Понятие об отсутствии системы в последовательности независимых испытаний используется в статистике. Оно весьма существенно также, скажем, в выработке глобальной стратегии поисков нефти, когда информация о перспективности подлежащего разбуриванию участка очень ограничена.

Как отмечалось в гл. II и III, независимость наблюдений (исходов испытаний) является исключением, так как обычно в геологии наблюдения зависимы. В то же время для большинства стандартных статистических гипотез (проверка равенства средних и т. п.) требуются именно независимые испытания. Это было подробно рассмотрено в гл. III. Таким образом, если бы были какие-то области геологии, в которых можно было бы опираться на независимые наблюдения, то это привело бы к корректному использованию в геологии весьма эффективных методов. К сожалению, вследствие неразработанности общих вопросов в геологических науках сейчас нет четко выделенных областей, где заведомо можно опираться на независимость наблюдений и соответственно на бернуллиевские последовательности. В то же время можно привести некоторые примеры, где такие последовательности наблюдались. Так, например, М. Е. Демина (1978), исследуя размещение песчнков разного минерального состава в песчаниках разного возраста в плоскости слоистости и перпендикулярно ей, установила, что бернуллиевские последовательности довольно часто встречаются на плоскостях слоистости.

После работ М. А. Романовой в северо-восточной Якутии было обращено внимание на то, что в гранитах, кристаллизовавшихся из магмы, обогащенной летучими компонентами, имеется повышенная вероятность встречи последовательностей зерен кварца, калиевого полевого шпата и плагиоклаза, не отличающихся от бернуллиевских (Вистелиус, Романова, 1976). Эти наблюдения были подтверждены затем специальными работами в Сьерра-Неваде (Калифорния), на Камчатке и на массиве Мальсбург в ФРГ. Интересные наблюдения в этом направлении были сделаны Д. Н. Ивановым на массиве Кызыл-Тас в Казахстане (Иванов, 1978) и в дальнейшем на Чалбинской интрузии в Приамурье и в массиве Безымянном в Якутии.

Специальный интерес представляет проблема существования связи между датой падения метеорита и его типом. Расчеты наблюдений, приведенных в примере IV.2, показывают, что никаких указаний на зависимость между датой падения и типом структуры железного метеорита, видимо, нет. В таком же аспекте следовало бы

изучить вопрос о падении метеоритов вообще. В частности, интересен вопрос о зависимости L -типов хондритов от даты их падения. Здесь, видимо, никакой зависимости нет, и это позволяет проверить реальность заключений о поведении, скажем, концентрации платиноидов в зависимости от петрографического типа хондрита.

Не останавливаясь на других примерах, отметим только, что с позиций гипотезы независимости интересно было бы проанализировать размещение изоморфных примесей рассеянных элементов в минералах. Впрочем, неясно, насколько для этого созрела методическая подготовленность геохимии.

Рассмотрим теперь случай, когда $r=1$, т. е. простую (первого порядка) марковскую цепь. Марковское свойство первого порядка (IV.2.6) в этом случае имеет вид:

$$p[B | a_I(h-1)A] = p[B | a_I(h-1)] \quad (B \in \mathfrak{B}_h, A \in \mathfrak{A}_{h-2}). \quad (\text{IV. 2. 16})$$

Случайная последовательность, которой отвечает (IV. 2. 16), будет полностью определяться вероятностями из (IV. 2. 9), т. е. в данном случае числами $p_I(h-1)$, $p_{I;J}(h-1, h)$, $I, J \in \mathcal{S}$, $h \in \{1, 2, \dots\}$.

Вероятность для цепочки $a_M(h-r)$, $a_N(h-r+1)$, $a_T(h-r+2)$, \dots , $a_J(h-1)$, $a_I(h)$ рассчитывается в случае простой цепи как

$$p_M(h-r) p_{M;N}(h-r, h-r+1) p_{N;T}(h-r+1, h-r+2), \dots \\ \dots, p_{J;I}(h-1, h).$$

Стационарность в случае простой цепи означает, что все вероятности $p_I(h)$, $p_{I;J}(h-1, h)$... не зависят от h . Поэтому здесь все определяется наборами чисел p_I и $p_{I;J}$. Однако стационарные вероятности p_I должны быть согласованы с вероятностями $p_{I;J}$. Как правило, они рассчитываются по этим вероятностям единственным образом, так что в случае стационарности вероятностная мера на простой цепи определяется одними числами¹ $p_{I;J}$. Если же имеется лишь марковская однородность, то переходные вероятности не зависят от h , а безусловные вероятности $p_I(h)$ могут зависеть от h (в конечном счете эти вероятности зависят от задания начального распределения $p_I^{(0)}$, $I \in \mathcal{S}$).

Итак, в случае однородной простой цепи вероятностная мера определяется заданием чисел $p_I^{(0)}$ и $p_{I;J}$. Однородная марковская цепь, у которой начальное распределение вероятностей совпадает со стационарным, эквивалентна, очевидно, стационарной цепи. Что же касается термина «однородная марковская цепь», то он применяется только так, как мы его определили, т. е. означает постоянство переходных вероятностей при произвольном начальном распределении.

Из формулы (IV.2.7), выражающей марковское свойство в симметричном виде относительно будущего и прошлого, ясно, что

¹ Это имеет место, если стационарная цепь обладает эргодическим свойством (понятие «эргодичность» подробно рассматривается далее).

простая марковская цепь, рассмотренная в обратном направлении (т. е. обращенная последовательность), также будет обладать простым марковским свойством. Однако эта *обращенная простая марковская цепь* уже не будет однородной, даже если исходная последовательность была однородной простой марковской цепью. Для того чтобы простая марковская цепь сохранила однородность после обращения, надо, чтобы исходная последовательность была стационарной марковской цепью. Поэтому, если нужно работать с однородными цепями и требуется обратимость, то на последовательность необходимо наложить требование о совпадении начального распределения со стационарным. Для выяснения этого явления возьмем в однородной последовательности пару смежных состояний $a_I(h-1); a_J(h)$. Вероятность этой пары в исходной последовательности $p(a_I(h-1), a_J(h))$. В обращенной последовательности вероятность той же пары $p(a_J(h-1), a_I(h))$. Переходная вероятность для прямой последовательности будет $p_{I;J}(h-1, h)$, а для обращенной последовательности переходная вероятность из I в J , которую мы будем отмечать знаком $(-)$, равна

$$p_{I;J}^{(-)} = \frac{p[a_J(h-1), a_I(h)]}{p[a_I(h)]} = \frac{p_J(h-1) p_{J;I}(h-1, h)}{p_I(h)}. \quad (\text{IV. 2. 17})$$

В случае однородной марковской цепи переходная вероятность в обратном направлении, из I в J , согласно (IV. 2. 17), будет

$$\frac{p_J(h-1)}{p_I(h)} p_{J;I}. \quad (\text{IV. 2. 18})$$

Как видим, (IV.2.18) зависит от h по определению. Таким образом, обращенная цепь не обладает марковской однородностью.

Если же исходная цепь имеет стационарное исходное распределение, то $p_J(h-1) = p_J$, $p_I(h) = p_I$, и (IV.2.18) дает

$$p_{I;J}^{(-)} = \frac{p_J}{p_I} p_{J;I}. \quad (\text{IV. 2. 19})$$

Здесь обращенная последовательность оказывается однородной простой марковской цепью.

Итак, если $p_{I;J}^{(-)} = p_{J;I}$ и безусловные вероятности не зависят от времени, то простая однородная марковская цепь является обратимой, т. е. обладающей свойством (IV.2.5). Если же безусловные вероятности состояний зависят от времени, то цепь необратима.

Простые марковские цепи классифицируются по вероятностным свойствам отдельных состояний и по свойствам множеств состояний. Так, множество состояний называется эргодическим, если из любого состояния, относящегося к этому множеству, за какое-то число шагов можно перейти в любое другое состояние (в частности, в то же самое) из этого множества. Если множество всех

состояний \mathcal{S} является эргодическим, то цепь называется эргодической.² Если же из состояния I нельзя попасть в состояние J ни за какое число шагов, то J называется состоянием, недостижимым из I .

Подмножество \mathcal{S}_1 множества \mathcal{S} называется замкнутым, если никакое состояние из $\mathcal{S} \setminus \mathcal{S}_1$ не достижимо из любого состояния в \mathcal{S}_1 . Если у множества всех состояний \mathcal{S} нет никакого замкнутого подмножества, кроме самого \mathcal{S} , то цепь называется неприводимой. Смысл терминов «приводимость» и «неприводимость» в том, что в случае приводимости имеется замкнутое подмножество \mathcal{S}_1 состояний, на котором определена отдельная цепь, не связанная с остальными состояниями. Такую цепь удобно изучать отдельно, т. е. «привести» общую цепь к нескольким цепям на разных подмножествах указанного типа.

Если цепь неприводима, то для двух любых состояний I и J можно указать такую цепочку переходов, которая начинается с состояния I , заканчивается состоянием J и внутри которой каждый переход осуществляется с положительной вероятностью.

Пример IV. 4. Рассмотрим простую однородную цепь с четырьмя состояниями $\{1, 2, 3, 4\}$ и переходными вероятностями $p_{11}=p_{12}=p_{13}=0$; $p_{14}=1$; $p_{21}=p_{22}=p_{23}=0$; $p_{24}=1$; $p_{31}=p_{32}=1/2$; $p_{33}=p_{34}=0$; $p_{41}=p_{42}=p_{44}=0$; $p_{43}=1$. Хотя среди переходных вероятностей много нулей, все же цепь является эргодической и неприводимой. Укажем соответствующие цепочки переходов:

1 → 1 путем 1 → 4 → 3 → 1; 1 → 2 путем 1 → 4 → 3 → 2;
1 → 3 путем 1 → 4 → 3; 1 → 4 путем 1 → 4;

2 → 1 путем 2 → 4 → 3 → 1; 2 → 2 путем 2 → 4 → 3 → 2;
2 → 3 путем 2 → 4 → 3; 2 → 4 путем 2 → 4;

3 → 1 путем 3 → 1; 3 → 2 путем 3 → 2; 3 → 3 путем 3 → 1 → 4 → 3; 3 → 4 путем 3 → 1 → 4;

4 → 1 путем 4 → 3 → 1; 4 → 2 путем 4 → 3 → 2; 4 → 3 путем 4 → 3; 4 → 4 путем 4 → 3 → 1 → 4.

Таким образом, цепь неприводима.

С другой стороны, рассмотрим простую однородную марковскую цепь на трех состояниях с вероятностями перехода $p_{11}=1/3$; $p_{12}=1/3$; $p_{13}=1/3$; $p_{21}=0$; $p_{22}=1$; $p_{23}=0$; $p_{31}=0$; $p_{32}=0$; $p_{33}=1$.

В данном случае из подмножества $\{2, 3\}$ нельзя выйти, т. е. попасть в состояние 1, поэтому подмножество $\{2, 3\}$ является замкнутым и цепь оказывается приводимой.

Если замкнутое подмножество состоит из одного состояния, то такое состояние называется поглощающим. В поглощающее состояние можно попасть, но из него нельзя перейти в какое-либо иное состояние. Состояние называется невозвратным, если в него

² В § IV.3 марковские цепи классифицируются с другой точки зрения. При этом понятия, являющиеся общими для обоих параграфов, хотя и выражены в разных терминах, но имеют один и тот же смысл. Так, эргодические цепи настоящего параграфа и § IV.3 являются одним и тем же объектом. То же относится и к понятиям «периодичность» и «приводимость».

можно однажды попасть, но, выйдя из него, вернуться в это же состояние нельзя.

Состояние I называется периодическим с периодом h , если из I в I можно вернуться только за число шагов, кратное h (т. е. за h , $2h$, $3h$. . .), причем h — наименьшее целое число, обладающее таким свойством. Тривиальный случай $h=1$, согласно принятому нами определению, мы не будем считать периодическим.

В неприводимой периодической цепи обязательно имеет место $p_{II}=0$, $I \in \mathcal{S}$. Действительно, если бы переход из I в I за один шаг имел положительную вероятность, то в силу простого марковского свойства переход из I в I за любое число шагов также имел бы положительную вероятность. Далее, для неприводимой цепи можно показать, что если хотя бы одно состояние из \mathcal{S} является периодическим с периодом h , то и любое другое состояние будет периодическим с тем же периодом. Иными словами, цепь будет периодична в целом.

Неприводимая периодическая цепь, являющаяся в то же время обратимой, может иметь лишь кратчайший возможный период $h=2$. Действительно, пусть неприводимая цепь обратима. Возьмем любое состояние I . В силу принятой неприводимости всегда найдется такое состояние $J \neq I$, что переход из I в J будет возможен за один шаг. Тогда в силу принятой обратимости будет возможен также переход за один шаг из J в I . Таким образом, возможна цепочка переходов с положительными вероятностями вида I, J, I . Это значит, что возвращение из I в I возможно за два шага. Поэтому либо период равен двум, либо периода нет совсем.

Дополнительные сведения о простых марковских цепях будут даны в § IV.3.

Рассмотрим некоторые геологические ситуации, в которых можно предполагать появление простых марковских цепей. Окатанность гальки обычно оценивается коэффициентом окатанности Ваделла — W . Он представляет собой среднее из радиусов кривизны выступов на проекции контура гальки, нормированное радиусом круга, вписанным в этот контур. Этот коэффициент характеризует не геометрические особенности контура, а лишь его гладкость. Арифметизируем значения коэффициентов окатанности, деля их на классы и нумеруя классы так, что большей окатанности отвечает больший номер класса.

В опытах Кьюнена измерялась окатанность кубов различных пород через равные расстояния (шаг), «пройденные» каждым кубом (Кюнен, 1956). В этих условиях весьма правдоподобно предположение, что до известного предела окатанность ведет себя следующим образом. 1. Окатанность в следующий момент ($W(h+1)$) зависит от окатанности в предыдущий момент ($W(h)$) хотя бы потому, что окатанность по каждой реализации практически всегда монотонно возрастает.³ 2. Дальнейшее изменение окатанности

³ Подразумевается, что раскалывания изучаемых частиц не происходит.

не зависит от того, через какие предварительные значения коэффициент окатанности прошел ранее.

Таким образом, последовательность коэффициентов окатанности одной и той же частицы, взятых через равные интервалы времени (или равные расстояния), оказывается простой цепью Маркова.

Цепь эта, естественно, всегда нестационарна, так как окатанность растет со временем (пройденным расстоянием). Но эта цепь может оказаться однородной, если изменение окатанности происходит за один шаг одинаково на любом участке пути.

Рассмотрим теперь кристаллизацию участка гомогенного расплава, близкого по составу к эвтектике и удаленного от контактов с вмещающей полостью. При этом пусть кристаллизация идет так, что выделяющиеся кристаллы не претерпевают взаимного смещения. Поскольку кристаллизация носит эвтектический характер, естественно допустить, что кристаллы всех состояний растут примерно одновременно. В этом случае соотношение «кристалл состава I и его сосед — кристалл состава J » таково, что при росте I происходит в окрестности I обеднение расплава веществом, формирующим I . Вследствие этого рядом с I скорее всего появится зерно минерала не I (при двух возможных состояниях — зерно J). В то же время при одновременной кристаллизации части расплава, заметно удаленные от данного зерна, не будут испытывать в связи с его ростом ни обогащения, ни обеднения веществом, идущим на его образование. В таких условиях естественно предположить возникновение простой марковской цепи, обладающей обратимостью и стационарным начальным распределением. Обратимость следует ожидать из-за гомогенности расплава. Стационарное распределение скорее всего появляется из-за того, что кристаллизация во всех точках, занятых расплавом, происходит одновременно, т. е. нет развития процесса во времени. Что же касается пространственного развития процесса, то это только форма описания.

С изложенной позиции анализ большого количества несомненно первично магматических, неизмененных вторичными преобразованиями гранитов показал, что породы с последовательностями зерен, имеющими свойства, указанные в предыдущем абзаце, действительно существуют. Они наблюдались, например, в аплитовых гранитах в Гатино (Квебек, Канада), а также в гранитах Мальсбурга (ФРГ) и Кызыл-Тас (Казахстан, СССР).

Соотношения между гранитами с последовательностями зерен, аналогичными простым цепям Маркова, и с другими типами последовательностей зерен показаны в табл. IV.1. Почти во всех дайках аплитов и гранодиоритпорфиров, от бассейна р. Колымы, п-ова Чукотки и до Югославии, ФРГ и Шотландии, встречаются породы, неотличимые от предсказанных только что схемой кристаллизации, приводящей к простой марковости. Наконец, опытные исследования габро-диабазов с габровой структурой в Гульшаде (Центральный Казахстан) и нефелиновых окситов в Ока

(Канада) показали, что и эти породы укладываются в рассмотренную схему. Все это проверено с помощью статистических критериев, описание которых читатель найдет в гл. VI этой книги.

Создается впечатление, что чередование биоценозов, поддающихся типизации в пределах отдельных экологических ниш, вследствие отравления окружающей среды продуктами жизнедеятельности данного биоценоза, также может приводить к марковскому чередованию фаций с разными биоценозами.

Число подобных примеров может быть легко увеличено. Это показывает, что марковские структуры очень распространены в природных явлениях.

IV.2.5. Сложные цепи Маркова и существенно немарковские последовательности

Перейдем к изучению случая, когда $r=2$ (марковская цепь второго порядка).

Марковское свойство (IV.2.6) у цепи Маркова второго порядка имеет вид

$$p[B | a_I (h-1) a_J (h-2) A] = p[B | a_I (h-1) a_J (h-2)], \quad (\text{IV.2.20})$$

где $B \in \mathfrak{B}_h$, $A \in \mathfrak{A}_{h-3}$.

Вероятностная мера для цепи со свойством (IV.2.20) полностью определяется вероятностями из (IV.2.9), т. е. в данном случае числами $p_I(h)$, $p_{I,J}(h, h+1)$, $p_{I,J;K}(h, h+1, h+2)$ (для всех целых $h \geq 0$ и всех $I, J, K \in \mathcal{S}$).

Вероятность цепочки

$$a_K(h-r), a_L(h-r+1), a_M(h-r+2), \dots, a_T(h-3), a_W(h-2), \\ a_J(h-1), a_I(h)$$

вычисляется как

$$p_K(h-r) \cdot p_{K,L}(h-r, h-r+1) \cdot p_{K,L;M}(h-r, h-r+1, h-r+2) \cdot \dots \\ \dots \cdot p_{T,W;J}(h-3, h-2, h-1) \cdot p_{W,J;I}(h-2, h-1, h).$$

Однородность и стационарность марковской цепи второго порядка существенно различаются. Чтобы задать однородную цепь второго порядка, надо указать все вероятности p_I , $p_{I,J}$, $p_{I,J;K}$; для задания стационарной цепи обычно⁴ достаточно задать $p_{I,J;K}$. Вероятности p_I и $p_{I,J}$ в этом случае единственным образом рассчитываются с помощью $p_{I,J;K}$ (см. пример в конце § IV.3).

Опыт показывает, что марковские цепи второго порядка чрезвычайно широко распространены в гранитах первично магматического происхождения, претерпевших некоторую вторичную переработку при образовании в них высокотемпературных гидротер-

⁴ См. сноску на с. 237.

мальных месторождений (Вистелиус, Романова, 1972; Иванов, 1978; Романова, 1978, и др.).

Распространенность в гранитоидах цепей различного, в том числе второго, порядка может быть проиллюстрирована данными табл. IV.1.

Таблица IV.1

Марковские свойства изученных последовательностей из зерен калиевого полевого шпата, кварца и плагиоклаза в рудоносных (показаны звездочкой) и безрудных гранитных массивах (данные на 1978 г.)

Массив (или дайка)	Число изученных образцов	Число образцов со следующими марковскими порядками последовательностей зерен		
		<2	2	>2
Гатино (Канада)	90	77	13	Нет
Мальсбург (ФРГ)	31	16	15	Нет
Кызыл-Тас (СССР)	214	102	101	11
Вост. Иультин* (СССР)	119	10	96	13
Омсукчан* (СССР)	305	15	239	51
Арга-Бных-Хая* (СССР)	140	8	112	20
Карменеллис* (Великобритания)	73	Нет	61	12

Пример IV.5. В районе Гульшад (Прибалхашье) имеются габбро-диабазы, состоящие из пироксена (Pug) и плагиоклаза (Pl) габбровой структуры несомненно первично магматического происхождения. Для части этих габбро-диабазов характерно наличие в породе хаотически разбросанных выделений биотита (Bt). Необходимо проверить происхождение биотита. Если биотит имеет первично магматический генезис, то последовательность индивидов Pug, Pl и Bt в габбро-диабазе будет простой цепью Маркова, как это вытекает из принятой схемы кристаллизации расплава, рассмотренной при описании простых цепей Маркова на с. 241. Если же биотит появился в габбро-диабазе под действием калиевого метасоматоза и располагается между первичными зернами Pug и Pl, то последовательность зерен Pug и Pl должна быть простой цепью Маркова, а последовательность на трех состояниях Pug, Pl и Bt должна иметь специфические черты. Выясним, какие это черты, и проверим их наличие путем исследования породы под микроскопом.

Начнем исследования с того, что исключим выделения биотита (не будем их учитывать) и рассмотрим последовательность на двух состояниях — Pug и Pl. Если схема, на которую мы ссылались, верна, то эта последовательность должна быть простой цепью Маркова. Выполнив подсчеты с помощью методов, излагаемых в гл. VI, мы убедились, что последовательность на двух состояниях (Pug и Pl) не отличима от простой однородной и обратимой цепи Маркова.

Рассмотрим теперь последовательность на трех состояниях — Pуг, P1 и Bt. При этом мы будем фиксировать каждое зерно Pуг и P1 отдельно, а лейсты биотита, находящиеся между ними, принимать за один индивид, который будем называть биотитовым пакетом. Это делается из-за того, что границы между лейстами биотита часто определить трудно. Предположим, что появление биотитового пакета может происходить с постоянной вероятностью π ($0 < \pi < 1$) независимо от места его внедрения.

Для выяснения специфики последовательности Pуг, P1 и Bt рассмотрим сначала вероятность

$$p [a_I(h) | a_{Bt}(h-1) a_J(h-2) A] = \pi p'_J; I \neq p [a_I(h) | a_{Bt}(h-1)] = p'_I, (A \in \mathfrak{A}_{h-3}), \quad (IV. 2. 21)$$

где штрихом отмечаются вероятности, относящиеся к последовательности на двух состояниях $I, J \in \{Pуг, P1\}$.

Далее,

$$p [a_I(h) | a_J(h-1) A] = (1 - \pi) p'_J; I. \quad (IV. 2. 22)$$

Таким образом, если включить в «настоящее» два последовательных испытания, то «прошлое» не будет иметь значения для вероятности событий в будущем, так что в силу (IV. 2. 21) и (IV. 2. 22) рассматриваемая последовательность оказывается марковской цепью второго порядка и не является простой цепью. Полученная цепь Маркова второго порядка однородна, так как первичная простая марковская цепь была однородной и вероятность внедрения биотитового лейста не зависела от места внедрения в цепь.

Расчеты с помощью методов, изложенных в гл. VI, показывают, что последовательности зерен Pуг, P1 и пакетов Bt в габбро-диабазе Гульшада действительно неотличимы от однородной цепи Маркова второго порядка. Это может рассматриваться как подтверждение вторичности (метасоматического происхождения) биотита.

Изучение сложных марковских цепей с $r > 2$ не вносит каких-либо существенно новых моментов по сравнению со случаем, когда $r=2$. Отметим только, что при $r > 2$ для задания однородной марковской цепи нужно значительно больше информации, чем для стационарной.

Для сложных марковских цепей, по-видимому, нет установившейся классификации. Наиболее часто они приводятся к простым цепям, а последние классифицируются так, как уже отмечалось.

Рассмотрим общую стохастическую схему, которая приводит к возникновению марковской цепи любого конечного порядка r ($r=1, 2, \dots, k$). Допустим, что имеется случайная последовательность на множестве из двух состояний α и β . При этом вслед за состоянием β обязательно наступает состояние α , за состоянием α может следовать α с вероятностью p и β с вероятностью $1-p$. Это происходит независимо от того, как вела себя случайная последовательность ранее.

Рассматриваемая последовательность, согласно определению, является однородной цепью Маркова первого порядка с переходными вероятностями $p_{\alpha\alpha}=p$, $p_{\alpha\beta}=1-p$; $p_{\beta\alpha}=1$ и $p_{\beta\beta}=0$. Для того чтобы распределение состояний с любым номером было одинаковым (стационарным), надо взять в качестве начального распределения $p_{\alpha}(0)=\frac{1}{2-p}$, $p_{\beta}(0)=\frac{1-p}{2-p}$.

Каждая реализация такой последовательности кодируется и пропускается через преобразователь со следующим свойством. Устройство не пропускает более чем r одинаковых символов подряд. Появление в $r+1$ -й раз подряд состояния α отключает устройство, и оно включается снова появлением очередного состояния β . Полученная на выходе случайная последовательность является однородной цепью Маркова порядка r .

Действительно,

$$p [a_{\beta}(h+1) | \underbrace{a_{\alpha}(h), \dots, a_{\alpha}(h-r+1)}_r] = 1$$

и

$$p [a_{\beta}(h+1) | a_{\alpha}(h), \dots, a_{\beta}(h-r+1)] < 1,$$

как это следует непосредственно из схемы.

Рассмотрим далее вероятности

$$p [a_{\beta}(h+1) | \underbrace{a(h), \dots, a(h-r+1)}_r A]; \quad A \in \mathfrak{A}_{h-r}, \quad (\text{IV. 2. 23})$$

где испытания $a(l)$ от $l=h$ до $l=h-r+1$ -го моментов реализовались как угодно, но так, чтобы среди этих испытаний хотя бы в одном реализовалось β . После появления β весь случайный процесс (как исходный, так и преобразованный) начинается в вероятностном смысле заново. Поэтому прошлое не влияет на будущее, и вероятность (IV. 2. 23) не зависит от A .

В качестве подтверждения этой схемы рассмотрим деятельность гейзера. Она состоит из чередования выбросов воды различной интенсивности, разделенных периодами покоя. В случае покоя фиксируется состояние C , в случае выброса — серия (см. гл. II) состояний B . Длина серии совпадает с отметкой уровня воды, накопленной в водомере, оценивающим дебит выброса. Так, если в результате выброса был отмечен объем воды (в условной шкале) 2, то фиксируется серия BB , за которым обязательно следует состояние C . Допустим, что водомер может вмещать количество воды, поступившее при любом выбросе. Предположим также, что фиксируемая на таком водомере последовательность из C и B образует простую стационарную марковскую цепь (последнее допущение произвольно и преследует чисто иллюстративные цели). Пусть теперь реальный водомер имеет предельную вместимость с максимальной отметкой r . Таким образом, при особенно интенсивных выбросах вода выше r выливается и количество ее сверх r не фиксируется, а берется отметка r . Тогда наблюдения

над последовательностью B и C , полученные с помощью реального водомера, дадут марковскую цепь порядка r . Итак, структура цепи будет зависеть от вместимости водомера.

Рассмотрим теперь простейшую схему возникновения существенно немарковской последовательности. Такая последовательность может появиться, например, следующим путем. Имеется случайная последовательность на двух состояниях α и β . Последовательность обладает тем свойством, что, попав первый раз в состояние β , система может перейти из него в состояние α . При втором попадании в β это состояние становится поглощающим. В случае такой последовательности, зная, что в настоящий момент реализовалось состояние β , мы можем утверждать, что переход в β будет обязательным, если известно, что в прошлом, хотя бы самом отдаленном, уже наблюдалось состояние β . Если же состояние β ни разу до этого не встречалось, то переход в β не будет достоверным. Таким образом, для прогноза необходимо знание всего предыдущего.

Как будет показано в гл. V, существенно немарковские последовательности возникают также при объединении состояний в обратимых стационарных цепях. Поэтому следует ожидать, что существенно немарковские последовательности могут достаточно часто встречаться в природных явлениях. К сожалению, опознание с помощью статистических тестов таких последовательностей связано с большими трудностями. Некоторые данные по этому вопросу мы приведем в конце гл. VI.

IV.2.6. Распределения серий в марковских цепях

В известной мере структуру марковских цепей можно также характеризовать распределениями серий. Напомним, что I -серий длины l называется сочетание элементов I и \bar{I} (не I), расположенных в последовательности так, что h -е и $h+l+1$ -е места заняты состоянием \bar{I} , а места от $h+1$ до $h+l$ заняты состоянием I (это определение отнесено к любым целым h и $l > 0$). Если \bar{I} — любые состояния из множества состояний \mathcal{S} , не совпадающие с I , то мы говорим о серии без фиксированных концов. Если же в качестве \bar{I} на h -м и $h+l+1$ -м местах располагаются два определенных состояния J и K , где $J, K \in \mathcal{S} \setminus \{I\}$, то говорят о серии с фиксированными концами.

Дискретное распределение вероятностей длины серии приписывает каждой I -серии длины l ($l = 1, 2, \dots$) вероятность p_l ; $\sum_{l=1}^{\infty} p_l = 1$.

Распределения серий по длине могут использоваться для проверки гипотез о структуре случайных последовательностей.

Предположим, например, что имеется последовательность независимых испытаний с постоянными вероятностями. Для нее легко указать распределение серий по длине. Если это распределение сильно отличается от эмпирического, то гипотезу о незави-

симых испытаниях с постоянной вероятностью следует отбросить. Сложнее вопрос о том, можно ли принять эту гипотезу в случае достаточной близости теоретического и эмпирического распределений (как отмечалось в гл. III, в этом случае необходима оценка мощности статистического критерия). Действительно, такое же распределение серий по длине могут иметь некоторые последовательности с зависимыми испытаниями. Однако в классе стационарных марковских последовательностей разным порядкам марковости всегда отвечают разные распределения I -серий по длине. Это следует из приводимой ниже теоремы, в которой рассматривается распределение серий по длине. При этом мы ограничиваемся лишь изучением стационарных цепей на трех состояниях, т. е. при $\mathcal{S} = \{I, J, K\}$, с марковостью порядков $r=0, 1$ и 2 .

Т е о р е м а IV.1. Вероятности p_i ($i=1, 2, \dots$) в распределении I -серий по длине, отвечающие стационарной марковской цепи порядка r , образуют, начиная с p_r , геометрическую прогрессию со знаменателем

$$\frac{p_{l+1}}{p_l} = p [a_I(h) | a_I(h-1), \dots, a_I(h-r)] \quad (l=r, r+1, \dots). \quad (IV.2.24)$$

Для простой марковской цепи распределение по длине серий с фиксированными концами не зависит от того, какие состояния зафиксированы в концах серии.

Для марковских цепей порядка $r > 1$ выбор состояния, фиксированного в конце серий, влияет на распределение серий по длине.

Для серий с нефиксированными концами имеем:

а) последовательность Бернулли ($r=0$)

$$p_1 = p_I^r; \quad p_{1+s} = p_1 \cdot p_I^s \quad (s=0, 1, 2, \dots), \quad (IV.2.25)$$

б) простую цепь Маркова ($r=1$)

$$p_1 = p [a_I(h) | a_I(h-1)], \\ p_{1+s} = p_1 \{p [a_I(h) | a_I(h-1)]\}^s \quad (s=0, 1, 2, \dots). \quad (IV.2.26)$$

в) цепь Маркова второго порядка

$$p_1 = \frac{p_J p_J; I p [a_I(h)/a_I(h-1), a_J(h-2)] + p_K p_K; I p [a_I(h)/a_I(h-1), a_K(h-2)]}{p_J p_J; I + p_K p_K; I}, \\ p_2 = \frac{\{p_J p_J; I p [a_I(h)/a_I(h-1), a_J(h-2)] + p_K p_K; I p [a_I(h)/a_I(h-1), a_K(h-2)]\} \times p [a_I(h)/a_I(h-1), a_I(h-2)]}{p_J p_J; I + p_K p_K; I}, \quad (IV.2.27)$$

$$p_{2+s} = p_2 \{p [a_I(h)/a_I(h-1), a_I(h-2)]\}^s \quad (s=0, 1, 2, \dots)$$

Каждое из распределений (IV.2.25) и (IV.2.26) зависит от одного параметра (в качестве такого параметра можно взять,

например, p_I для последовательности Бернулли и $p_{I;I}$ — для простой цепи Маркова). Распределение (IV. 2. 27) зависит от семи независимых параметров (в качестве таких параметров можно взять, например, $P_J, P_K, P_{J;I}, P_{K;I}, P_{J,I;I}, P_{K,I;I}$ и $P_{I,I;I}$). Для случая с фиксированными концами (для определенности берется серия из состояний I ; перед серией располагается состояние J , а после серии — состояние K) при последовательности Бернулли и простой цепи Маркова распределения, как отмечалось, совпадают соответственно с (IV. 2. 25) и (IV. 2. 26). Для цепи Маркова второго порядка

$$\begin{aligned}
 p_1 &= \frac{p[a_I(h)/a_I(h-1), a_I(h-2)] p[a_K(h)/a_I(h-1), a_I(h-2)]}{p[a_K(h)/a_I(h-1), a_I(h-2)] p[a_I(h)/a_I(h-1), a_I(h-2)] +} \\
 &\quad + p[a_I(h)/a_I(h-1), a_I(h-2)] p[a_K(h)/a_I(h-1), a_I(h-2)] \\
 &\quad \times p[a_I(h)/a_I(h-1), a_I(h-2)] \times \\
 &\quad \times p[a_I(h)/a_I(h-1), a_I(h-2)] \times \\
 &\quad \times p[a_K(h)/a_I(h-1), a_I(h-2)] \\
 p_2 &= \frac{p[a_K(h)/a_I(h-1), a_I(h-2)] \times}{p[a_K(h)/a_I(h-1), a_I(h-2)] \times} \\
 &\quad \times p[a_I(h)/a_I(h-1), a_I(h-2)] + \\
 &\quad + p[a_I(h)/a_I(h-1), a_I(h-2)] \times \\
 &\quad \times p[a_K(h)/a_I(h-1), a_I(h-2)] \\
 p_{2+s} &= p_2 \{p[a_I(h)/a_I(h-1), a_I(h-2)]\}^s \quad (s=0, 1, 2, \dots).
 \end{aligned} \tag{IV. 2. 28}$$

Распределения для последовательности Бернулли и простой цепи Маркова зависят каждое от одного параметра (в качестве независимого параметра в последовательности Бернулли можно принять p_I , а в простой цепи Маркова — $p_{I;I}$). В случае цепи Маркова второго порядка распределение определяется четырьмя независимыми параметрами (в качестве таких параметров можно взять $P_{J,I;K}, P_{J,I;I}, P_{I,I;I}$ и $P_{I,I;K}$).

Итак, для проверки гипотезы о порядке марковости по распределению длины серий с фиксированными концами при $r=0$ и $r=1$ необходимо использовать не менее трех разрядов. При $r=2$ число разрядов должно быть не менее шести. При работе с сериями с нефиксированными концами для случаев $r=0$ и $r=1$ нужно иметь не менее трех разрядов, а для случая $r=2$ — не менее девяти.

Утверждения теоремы IV.1 непосредственно следуют из формул (IV. 2. 25)—(IV. 2. 28), определяющих вероятности длин серий и аналогичных им при $r > 2$, которые здесь не приводятся. Вывод указанных формул основывается на подсчете специальных переходных вероятностей. Так, например, независимо от того, какого порядка r будет цепь, вероятность иметь I -серию длины l , перед которой стоит состояние J и вслед за которой стоит состояние K , представляет условную вероятность

$$\begin{aligned}
 p_l &= p[a_K(h+l+1), a_I(h+l), \dots, a_I(h+3), \\
 &\quad a_I(h+2) | a_J(h), a_I(h+1)].
 \end{aligned} \tag{IV. 2. 29}$$

Для получения распределения серий по длине выражение (IV.2.29) необходимо разделить на сумму всех выражений такого вида, взятую по l , от 1 до ∞ .

Действительно, для того чтобы осуществилось рассматриваемое событие (серия с фиксированными концами), необходимо, чтобы условие $a_J(h)$, $a_I(h+1)$ было выполнено. Появление $a_J(h)$ необходимо потому, что рассматриваются только серии, перед которыми встречено состояние J , а дальнейшее появление I необходимо потому, что мы не включаем в множество серий пустые (длины 0). Если указанное условие реализовалось, то для получения рассматриваемой серии нужно, чтобы далее еще $l-1$ раз подряд следовало состояние I , которое затем сменялось бы состоянием K . Итак, нужно рассчитать вероятность (IV.2.29). Вычисление (IV, 2.29) производится по-разному — в зависимости от порядка марковости r .

Так, например, для стационарной цепи Маркова первого порядка (IV.2.29) вычисляется как

$$\frac{p[a_J(h) a_I(h+1) a_I(h+2), \dots, a_I(h+l), a_K(h+l+1)]}{p[a_J(h) a_I(h+1)]} = \\ = \frac{p_J p_{J; I} (p_{I; I})^{l-1} p_{I; K}}{p_J p_{J; I}} = p_{I; K} (p_{I; I})^{l-1}.$$

Далее полученное выражение делится на

$$\sum_{i=1}^{\infty} p_{I; K} (p_{I; I})^{i-1} = \frac{p_{I; K}}{p_{I; I}},$$

откуда

$$p_i = p_{I; I} (p_{I; I})^{i-1},$$

т. е. получается распределение серий по длине с фиксированными концами. Как видно, эта формула совпадает с (IV.2.26), т. е. со случаем с нефиксированными концами.

Как отмечалось в теореме, фактическая зависимость функции распределения от состояний, находящихся на концах серий, возникает, начиная с цепей, обладающих марковостью второго порядка.

Рассмотрим приложение выведенных формул для простого случая.

Пример IV.6. В аплитовидном граните (обр. 0-51) близ с. Омсукчан в бассейне Колымы была зафиксирована последовательность зерен кварца, калиевого полевого шпата и плагиоклаза (Ab), состоявшая из 967 зерен. В этой последовательности было встречено распределение серий из зерен плагиоклаза по длине при нефиксированных концах, приведенное в табл. IV.2.

Необходимо сравнить эмпирическое распределение серий по длине с теоретическими распределениями для $r=0$ и $r=1$ для того, чтобы проверить гипотезу, чем является последователь-

ность — последовательностью независимых испытаний ($r=0$) или простой цепью Маркова ($r=1$). Известно, что изучаемая последовательность является однородной (Вистелиус, Романова, 1972).

Путем расчета по формуле (IV.2.25) получаем $p_1 \approx 0.582$, в то время как наблюдаемое значение для $p_1=0.787$ при 233 наблюдаемых сериях длины 1. Это дает настолько большое слагаемое входящее в статистику χ^2 , что бернуллиевская гипотеза бракуется без дальнейших расчетов.

В табл. IV.2 даны результаты подсчета по формуле (IV.2.26). В эту формулу подставлено значение оценки $P_{\overline{A}B, \overline{A}B}$, которому отвечает наблюдаемое значение $\tilde{p}_1=0.787$. Из подсчета видно, что гипотеза о том, что изученная последовательность — простая однородная цепь Маркова, не противоречит наблюдаемому распределению серий. Отметим, что применение специального статистического критерия для проверки этой гипотезы (Вистелиус, Романова, 1972) привело к такому же заключению.

IV.3. МАТРИЧНЫЕ МЕТОДЫ ИЗУЧЕНИЯ ЦЕПЕЙ МАРКОВА

Матричные методы являются весьма важным инструментом для изучения цепей Маркова. Ниже дается систематическое изложение их применения.

IV.3.1. Марковские матрицы и стохастические векторы

Как отмечалось, задание переходных вероятностей на один шаг и фиксация начального распределения полностью определяют все случайное явление, связанное с соответствующей простой однородной цепью Маркова с конечным числом состояний. С позиций линейной алгебры (теории матриц) это означает, что случайное явление, связанное с однородной простой марковской цепью, полностью определяется заданием квадратной матрицы $P_{I;J}^{(1)}$ (в будущем для простоты мы ее часто будем обозначать $P^{(1)}$ или P) и вектором p_0 (здесь $P_{I;J}^{(1)}$ означает, что происходит переход с I непосредственно на J , при этом в изучаемой цепи эти состояния расположены рядом, что отмечено верхним индексом 1; если бы они были расположены на расстоянии в h испытаний, мы имели бы соответственно индекс h , а соответствующую матрицу переходов обозначили бы $P_{I;J}^{(h)}$). Это позволяет рассматривать однородную марковскую цепь как объект, исследуемый теорией матриц. Такой подход удобен при работе с наблюдениями, когда приходится широко пользоваться ЭВМ и стандартными программами.

Поскольку сложная марковская цепь (как отмечалось, у такой цепи порядок марковости $m > 1$) приводится к простой с помощью увеличения числа состояний, что отмечалось на с. 234 и

Проверка марковской гипотезы по распределению серий

Длина серии	1	2	3	4
Наблюдённое количество серий	233	49	10	4
Количество серий, вычисленное по (IV. 2. 26)	233.0	49.6	10.5	2.2

$$0.50 > p [\chi(\frac{3}{2}) \geq 1.50] > 0.25$$

будет рассмотрено в IV. 3. 6, здесь мы остановимся только на вопросах, связанных с однородной простой марковской цепью.

Всякой однородной простой марковской цепи отвечает стохастический вектор p и квадратная марковская матрица P размера $s \times s$, где s — число состояний в марковской цепи. Стохастическому вектору соответствует точка на поверхности s -мерного тетраэдра. Этот тетраэдр задается с помощью соотношений

$$\sum_{i=1}^s p_i = 1, \quad 0 \leq p_i \leq 1 \quad (i=1, 2, \dots, s)$$

в системе координат (p_1, p_2, \dots, p_s) .

Элементы марковской матрицы — переходные вероятности, т. е. неотрицательные числа $0 \leq p_{ij} \leq 1$; сумма элементов в каждой строке этой матрицы равна единице. Марковские матрицы обладают следующими свойствами.

1. Произведение двух марковских матриц есть снова марковская матрица. Это свойство обобщается на произведение любого числа произвольных марковских матриц одного порядка, которое всегда является марковской матрицей.

2. Среди характеристических чисел марковской матрицы всегда имеется характеристическое число $\lambda_1=1$; любое характеристическое число марковской матрицы по модулю не превышает единицы, т. е.

$$|\lambda_i| \leq 1 \quad (i=1, 2, \dots, s).$$

Среди характеристических чисел марковской матрицы могут быть собственные числа, по модулю равные единице, как вещественные, так и комплексные (т. е. некоторые характеристические числа могут быть -1 , 1 или $a+bi$ с $a^2+b^2=1$).

3. Собственному числу $\lambda_1=1$ отвечает для матрицы P собственный стохастический вектор

$$\begin{pmatrix} \frac{1}{s} \\ \vdots \\ \frac{1}{s} \end{pmatrix}.$$

4. Транспонированная марковская матрица P^T имеет, как и любая матрица, те же характеристические числа, что и P . При этом характеристическому числу $\lambda_1=1$ транспонированной марковской матрицы отвечает собственный вектор, являющийся стохастическим вектором p_{st} . Этот вектор, по определению, отвечает уравнению

$$P^T p_{st} = p_{st} \quad (\text{IV. 3. 1})$$

Распределение вероятностей состояний, задаваемое указанным вектором, называется стационарным распределением состояний, а вектор p_{st} — стационарным стохастическим вектором.

Важным частным случаем стохастической матрицы является матрица, элементы которой p_{ij} не зависят от индекса j (одинаковые вероятности в пределах одного столбца, меняющиеся, вообще говоря, от столбца к столбцу). Для такой матрицы стационарное распределение задается ее строкой. Матрицы такого типа отвечают последовательностям независимых испытаний с постоянными вероятностями (последовательности Бернулли). Последовательности Бернулли, как мы отмечали, рассматриваются здесь как марковские цепи нулевого порядка.

Другим важным типом марковской матрицы является так называемая *дважды стохастическая матрица*. Под этим названием известна стохастическая матрица, у которой сумма по любому столбцу равна единице. Очевидно, что всякая симметрическая марковская матрица является в то же время дважды стохастической, но не всякая дважды стохастическая матрица симметрична. Дважды стохастическая матрица обладает стационарным распределением вероятностей p_{st} , приписываемым всем состояниям одинаковую вероятность. Таким образом, для дважды стохастической матрицы

$$p_{st} = \begin{pmatrix} \frac{1}{s} \\ \vdots \\ \frac{1}{s} \end{pmatrix}.$$

Начальному распределению состояний отвечает стохастический вектор p_0 , соответствующий начальному состоянию системы и представляющий, как отмечалось, некую точку на поверхности s -мерного тетраэдра. Действие линейного оператора, отвечающего матрице $P_{I;J}^{(1)}$, переводит точку p_0 на поверхности тетраэдра в новое положение p_1 , где p_1 — распределение вероятностей состояний после первого испытания, т. е. в момент времени $t=1$. Этому действию матрицы на вектор отвечает уравнение

$$P^T p_0 = p_1 \quad (\text{IV. 3. 2})$$

(подчеркнем, что вероятностный смысл здесь имеет действие P^T на p , а не действие P). Далее,

$$P^2 p_1 = (P^2)^2 p_0 = p_2$$

(т. е. точка p_0 переводится во второе положение p_2) и вообще

$$P^k p_{k-1} = (P^k)^k p_0 = p_k, \quad (IV.3.3)$$

где p_k и p_{k-1} соответствуют распределениям состояний после k -го и $k-1$ -го испытаний.

Поскольку после любого испытания распределение вероятностей состояний есть стохастический вектор, мы видим, что действие марковского оператора (матрицы) перемещает точку (распределение вероятностей состояний) по поверхности тетраэдра.

Следует отметить, что действие матрицы P^k , не имеющей нулевых элементов, на стохастический вектор, некоторые компоненты которого равны нулю, приводит к стохастическому вектору без нулей.

Если мы имеем дело с однородной простой цепью Маркова, то общий элемент матрицы переходных вероятностей на k шагов, который мы будем обозначать $p_{I;J}^{(k)}$, совпадает с соответствующим элементом k -той степени матрицы переходных вероятностей на один шаг, иными словами,

$$p_{I;J}^{(k)} = \| P^k \|_{I,J}. \quad (IV.3.4)$$

Докажем это с помощью математической индукции. Для $k=1$ соотношение, очевидно, выполняется. Пусть далее (IV.3.4) выполняется для $k=n-1$, т. е.

$$p_{I;J}^{(n-1)} = \| P^{(n-1)} \|_{I,J}.$$

Докажем, что соотношение (IV.3.4) справедливо и при $k=n$. Чтобы перейти из состояния I в состояние J за n шагов, нужно перейти из состояния I в произвольное состояние L за $n-1$ шаг, а затем из состояния L за один шаг попасть в J . Для однородной простой марковской цепи вероятность таких переходов вычисляется по формуле

$$p_{I;J}^{(n)} = \sum_{L \in \mathcal{S}} p_{I;L}^{(n-1)} p_{L;J} = \sum_{L \in \mathcal{S}} \| P^{n-1} \|_{I,L} p_{L;J} = \| P^n \|_{I,J}.$$

IV.3.2. Предельные распределения в марковских цепях

Классифицируем теперь марковские матрицы с точки зрения предельного поведения точки в процессе ее последовательного перемещения по грани тетраэдра.

Наиболее распространенным в геологических приложениях является случай *эргодической марковской цепи*. В этом случае последовательно перемещающаяся точка p_0, p_1, \dots, p_k стремится занять предельное положение, из которого она больше не выходит. Таким образом, здесь существует предельное распределение вероятностей состояний

$$p_f = \lim_{k \rightarrow \infty} p_k.$$

Характеристическая черта эргодической цепи заключается в том, что одна и та же предельная точка p_f достигается независимо от того, из какого исходного положения p_0 вышла марковская цепь (независимость p_f от p_0). Распределение p_f в эргодической цепи всегда является стационарным, а предельная матрица переходных вероятностей за k шагов, т. е. $\lim_{k \rightarrow \infty} P^{(k)}$ существует и является матрицей, отвечающей последовательности независимых испытаний с постоянными вероятностями, т. е. матрицей, у которой элементы в пределах каждого столбца не меняются (см. с. 252).

Достаточно часто может встречаться случай, когда предельного положения точки, отвечающей распределению вероятностей состояний, не существует. Если p_f не существует, то цепь обязана быть *периодической*. Как отмечалось, это означает, что из любого состояния I можно вернуться в I только за количество шагов, кратное периоду h , в то время как $P_{II}^{(s)} = 0$ при $s \neq rh$ (r и h — целые). Наконец, возможен промежуточный случай — предельная точка существует, но она зависит от начального положения (p_f зависит от p_0). Такие цепи называют *квазиэргодическими* (Гантмахер, 1967).

Пример IV.7. Рассмотрим вопрос о существовании в природе аналогов эргодических марковских цепей. Для этого изучим последовательность зерен кварца (Q), калиевого полевого шпата (Or) и плагиоклаза (Ab) в шлифе аляскита с р. Каракульджур в Центральном Тянь-Шане, неотличимую от простой однородной марковской цепи, с матрицей переходных вероятностей, округленных до десятых (Вистелиус, 1967):

$$P = \begin{matrix} & \begin{matrix} \text{Or} & \text{Q} & \text{Ab} \end{matrix} \\ \begin{matrix} \text{Or} \\ \text{Q} \\ \text{Ab} \end{matrix} & \begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.6 & 0.0 \end{pmatrix} \end{matrix}. \quad (\text{IV. 3. 5})$$

Начальное распределение состояний нам не известно. Возьмем за начальное распределение

$$p_0 = \begin{pmatrix} 0.0 \\ 0.5 \\ 0.5 \end{pmatrix}.$$

Тогда поведение случайной системы можно иллюстрировать графически как последовательное перемещение точки по грани трехмерного тетраэдра. Действительно, распределения состояний после первых трех испытаний

$$p_1 = P^1 p_0 = \begin{pmatrix} 0.50 \\ 0.35 \\ 0.15 \end{pmatrix},$$

$$p_2 = P^T p_1 = (P^T)^2 p_0 = \begin{pmatrix} 0.315 \\ 0.465 \\ 0.220 \end{pmatrix}, \quad (\text{IV. 3. 6})$$

$$p_3 = P^T p_2 = (P^T)^3 p_0 = \begin{pmatrix} 0.3220 \\ 0.4755 \\ 0.2025 \end{pmatrix}.$$

Вычислять предельное распределение точки, повторяя снова и снова расчеты типа (IV. 3. 6), нерационально. Практически (Феллер, 1964), для того чтобы разыскать предельное распределение, нужно сначала убедиться, что оно существует, а затем определять его компоненты из уравнения (IV. 3. 1). В данном случае для этого нужно решить систему линейных уравнений

$$P^T y = y,$$

где P^T — транспонированная матрица (IV. 3. 5), а y — ее собственный вектор, отвечающий собственному числу $\lambda_1 = 1$.

Отсюда получим систему:

$$\begin{aligned} 0.3y_1 + 0.3y_2 + 0.4y_3 &= y_1, \\ 0.5y_1 + 0.4y_2 + 0.6y_3 &= y_2, \\ 0.2y_1 + 0.3y_2 &= y_3. \end{aligned}$$

Так как сумма компонент вектора y приводится в дальнейшем к 1, то можно сначала выбрать одну из компонент произвольной. Возьмем $y_1 = 1$. Тогда из третьего уравнения найдем

$$y_3 = 0.2 + 0.3y_2.$$

Подставляя выражение для y_3 в первое уравнение, получим

$$y_2 = 1.4762, \quad y_3 = 1.6429.$$

Таким образом

$$y = \begin{pmatrix} 1.0000 \\ 1.4762 \\ 0.6429 \end{pmatrix}.$$

Приведем сумму компонент к единице и получим окончательно

$$p_{st} = \begin{pmatrix} 0.3206 \\ 0.4733 \\ 0.2061 \end{pmatrix}.$$

Если бы мы неограниченно повторяли операции типа (IV. 3. 6), то пришли бы именно к этой точке.

Возьмем теперь другое начальное распределение, например

$$p'_0 = \begin{pmatrix} 0.6 \\ 0.4 \\ 0.0 \end{pmatrix}.$$

Тогда будем иметь

$$p_1' = \begin{pmatrix} 0.30 \\ 0.46 \\ 0.24 \end{pmatrix}, \quad p_2' = \begin{pmatrix} 0.3240 \\ 0.4760 \\ 0.1980 \end{pmatrix}, \quad p_3' = \begin{pmatrix} 0.3198 \\ 0.4666 \\ 0.2082 \end{pmatrix}, \text{ и т. д.}$$

Приведенные данные показывают, что для последовательности зерен изученных аляскинтов существует предельная точка и что в эту предельную точку мы приходим независимо от того, какой была начальная. Таким образом, исследованная последователь-

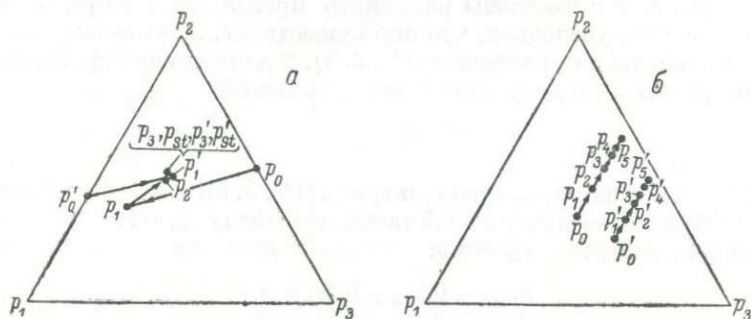


Рис. IV.4. Сходимость начального вектора к вектору стационарного распределения.

а — эргодический, *б* — квазиэргодический случаи.

ность действительно ведет себя как простая однородная эргодическая цепь Маркова.

Из приведенных вычислений и рис. IV. 4 видно, что предельная точка достигается практически после второго шага и не зависит от положения исходной. Если подобного рода быстрая сходимость имеет место, то говорят, что цепь быстро устанавливается (далее мы увидим, что скорость сходимости зависит от модулей характеристических чисел).

При геологических исследованиях мы, как правило, не знаем начального распределения, так как оперируем обычно с единичной реализацией. При этом исследуемая единичная реализация начинается не с того места, где начинает осуществляться механизм, порождающий последовательность, а на некотором расстоянии от него (например, на некотором расстоянии от контакта интрузии или подошвы толщи). Можно предположить, что на этом расстоянии цепь успевает установиться, т. е. распределение вероятностей исходов становится очень близким к стационарному распределению. В тех случаях, когда есть основания полагать, что цепь устанавливается к месту наблюдения (скажем, к изучаемому разрезу) исследуемой части последовательности, принимается, что начальное и предельное распределения совпадают.

Очевидно также, что имеется ряд случаев, когда исход начального испытания имеет другую вероятность, чем исходы последующих. Это отражено в геологии в таких понятиях, как «зальбанд» жилы или «краевая часть интрузии». В последнем случае, в краевой части интрузии, вероятность встречи, скажем, темноватого минерала много выше, чем в главной, более внутренней, части интрузивного тела гранита.

Таким образом, изучая гранит от контакта внутрь массива, нужно считаться с возможными различиями между вероятностями появления первого от контакта зерна определенного состава и вероятностью появления зерна того же состава на некото-

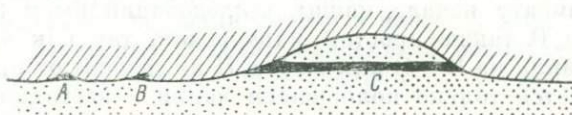


Рис. IV.5. Схема формирования нефтяной залежи.

Носая штриховка — экраняющий горизонт; *точки* — пористый горизонт, по которому мигрируют воды, несущие капельную нефть; *А и В* — каверны, заполненные нефтью; *С* — растущая нефтяная залежь.

ром расстоянии от контакта. Каждый отдельный случай требует индивидуального рассмотрения с учетом несоразмерности размера шлифа, в котором ведется исследование, и всего массива в целом. Аналогичные соображения должны учитываться и при исследовании других объектов — рудных тел, осадочных толщ, распределения фауны в экологической нише и т. п.

Пример IV.8. Рассмотрим теперь пример, показывающий, что не во всех случаях при геологических испытаниях различные начальные распределения приводят к одному и тому же предельному распределению. Иногда предельные распределения бывают различными, что видно из следующего. Пусть в мигрирующих по пласту водах имеются капли нефти молекулярного размера. При этом допустим, что процесс уже шел достаточно длительное время и соотношения между водой, нефтью и покрывающим, экраняющим, слоем уже установились. При этом примем также, что пласт имеет небольшие, редко встречающиеся куполовидные поднятия, площадь которых мала по сравнению со всей площадью, на которой распространен пласт. Экраняющий пласт непроницаем для воды и нефти, но в нем имеются мелкие шероховатости (поры, трещины) на нижней поверхности, которые уже были заполнены нефтью к тому моменту, когда мы начали изучать систему. В противоположность этому в пределах куполовидных поднятий нефть еще накапливается, вырастая в залежь за счет аккумуляции отмеченных молекулярных капель нефти в ловушке. Эти соотношения показаны на рис. IV.5.

Опишем охарактеризованную систему в терминах переходных вероятностей при трех состояниях: капля нефти в мигрирующих водах (В), капля нефти в трещине, заполненной нефтью (П), и капля нефти в растущей нефтяной залежи (Н). Рассмотрим состояния капли в последовательные моменты времени в предположении, что положение капли в данный момент известно. Очевидно, что переход из состояний Н и П простой марковский, так как независимо от предыстории можно предсказать, что далее будет либо Н, если было Н, либо П, если ранее было П. Из Н нет перехода ни в В, ни в П, так же как нет перехода из П ни в В, ни в Н. Предположим также, что переход капли из состояния В возможен либо в В, либо в Н и невозможен в П, так как трещины и поры к моменту начала наших исследований были заполнены. Переход из В также простой марковский, так как вероятность перехода капли из В в В или из В в Н не зависит от предыстории и определяется только положением капли в настоящий момент. Будем проводить наблюдения через достаточно длительные промежутки времени, определяя положение капли в момент $h-1$ и предсказывая поведение капли в момент h .

Если испытания проходят по изложенной схеме, то явление может быть охарактеризовано матрицей переходных вероятностей ⁵

$$P = \begin{matrix} & \begin{matrix} В & Н & П \end{matrix} \\ \begin{matrix} В \\ Н \\ П \end{matrix} & \begin{pmatrix} p & 1-p & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{matrix}. \quad (IV. 3. 7)$$

Начальное распределение нам не известно. В качестве такового выберем, например, вектор с компонентами $p(В) = p(Н) = p(П) = 1/3$. Тогда, пользуясь формулами (IV. 3. 6), получим

$$p_1 = \begin{pmatrix} 0.2333 \\ 0.4334 \\ 0.3333 \end{pmatrix}, \quad p_2 = \begin{pmatrix} 0.1633 \\ 0.5034 \\ 0.3333 \end{pmatrix}, \quad p_3 = \begin{pmatrix} 0.1143 \\ 0.5524 \\ 0.3333 \end{pmatrix},$$

$$p_4 = \begin{pmatrix} 0.0800 \\ 0.5867 \\ 0.3333 \end{pmatrix}, \quad p_5 = \begin{pmatrix} 0.0560 \\ 0.6107 \\ 0.3333 \end{pmatrix}.$$

Если в качестве начального распределения принять $p(В) = 0.25$, $p(Н) = 0.25$ и $p(П) = 0.50$, то последовательность векторов окажется следующей

$$p_1 = \begin{pmatrix} 0.1750 \\ 0.3250 \\ 0.5000 \end{pmatrix}, \quad p_2 = \begin{pmatrix} 0.1225 \\ 0.3775 \\ 0.5000 \end{pmatrix}, \quad p_3 = \begin{pmatrix} 0.0858 \\ 0.4143 \\ 0.5000 \end{pmatrix},$$

⁵ Приводимые значения вероятностей условны, использование значений, ожидаемых в результате естественных наблюдений, привело бы к громоздким расчетам.

$$p_4 = \begin{pmatrix} 0.0601 \\ 0.4399 \\ 0.5000 \end{pmatrix}, \quad p_5 = \begin{pmatrix} 0.0421 \\ 0.4579 \\ 0.5000 \end{pmatrix}.$$

Как видим, две изученные последовательности точек на грани тетраэдра не стремятся к одной и той же точке, поскольку последняя компонента в этих точках разная; это происходит не из-за того, что взято мало итераций. Подобное поведение системы может вызываться одной из двух причин: либо предельной точки не существует, либо своя предельная точка существует для каждого начального распределения. Для решения этого вопроса нет необходимости находить стационарное распределение, как мы это делали в предыдущем примере (как выяснится в дальнейшем, стационарное распределение не эквивалентно в настоящем случае предельному распределению). Найдем предельное распределение, пользуясь непосредственно его определением, т. е. формулой (IV.3.2) как

$$\lim_{k \rightarrow \infty} (P^T)^k p_0.$$

Для этого найдем

$$\lim_{k \rightarrow \infty} (P^T)^k = \lim_{k \rightarrow \infty} \begin{pmatrix} p & 0 & 0 \\ \dots & \dots & \dots \\ 1-p & 1 & 0 \\ \dots & \dots & \dots \\ 0 & 0 & 1 \end{pmatrix}^k.$$

Разбивая P^T на клетки пунктиром, как показано в формуле, по известному правилу (Фаддеев, Фаддеева, 1960, с. 12), получим

$$(P^T)^k = \begin{pmatrix} p^k & 0 & 0 \\ x & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Элемент, обозначенный x , легко определить, учитывая, что матрица должна быть транспонированной марковской, т. е. сумма элементов по столбцу должна быть равна единице. Таким образом,

$$(P^T)^k = \begin{pmatrix} p^k & 0 & 0 \\ 1-p^k & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

и в силу того, что $0 < p < 1$,

$$\lim_{k \rightarrow \infty} (P^T)^k = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Пусть теперь начальному распределению отвечает вектор с компонентами $p_0(B)$, $p_0(H)$ и $p_0(\Pi)$. Тогда предельным распределением будет

$$p_f = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_0(B) \\ p_0(H) \\ p_0(\Pi) \end{pmatrix} = \begin{pmatrix} 0 \\ p_0(B) + p_0(H) \\ p_0(\Pi) \end{pmatrix}.$$

Как видим, предельное распределение существует, зависит от выбора начальных вероятностей и, таким образом, будет разным в зависимости от того, как начинается процесс. Следовательно, в рассматриваемом явлении цепь ведет себя как квазиэргодическая.

Вычислим теперь стационарное распределение. Пользуясь формулой (IV.3.3), получим

$$\begin{pmatrix} p & 0 & 0 \\ 1-p & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

$$(0 < p < 1),$$

откуда $a = 0$, b — произвольное число $0 < b < 1$, $c = 1 - b$.

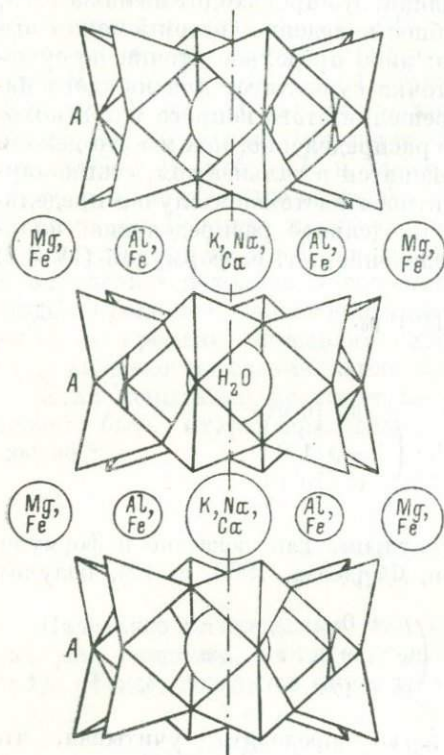


Рис. IV.6. Структура осунилита в проекции на плоскость (0110).

A^+ — кольца, связываемые изоморфно замещающимися ионами. Рассматривается последовательность колец (A) и случайно чередующихся ионов K, Na и Ca вдоль показанной на рисунке оси.

Таким образом,

$$p_{st} = \begin{pmatrix} 0 \\ b \\ 1 - b \end{pmatrix}.$$

Мы видим, стационарное распределение, т. е. распределение, отвечающее вектору из (IV.3.1), не обязано быть предельным, однако предельное распределение всегда является стационарным.

Пример IV.9. Для многих силикатов характерно чередование слоев или колец определенного состава, соединенных друг с другом катионами или водой (гидроксилем), при этом катионы изоморфно замещают друг друга. В итоге создается большое разно-

образии стохастических структур, которое рано или поздно должно привлечь внимание минералогов и кристаллохимиков.

Ниже в качестве примера периодической цепи рассматривается чередование структурных элементов в осумилите. Основным мотивом минерала является кольцо, состоящее из комплексов $(Si_5Al)O_{18}$, чередующихся по два. Эти кольца, которые мы будем обозначать буквой А, связываются катионами К, Na и Са. Рассмотрим наиболее удобный случай, полагая, что содержания К, Na и Са в минерале одного порядка, что подтверждается анализом (Дир и др., 1965, с. 320). Представление о структуре дает рис. IV.6.

Итак, рассмотрим чередование А, К, Na и Са вдоль прямой, проходящей через центральную часть кольца. Как видно из рис. IV. 6, кольцо А занимает в структуре детерминированное положение. Между каждыми соседними А случайно может располагаться либо К с постоянной вероятностью p_K , либо Na с вероятностью p_{Na} , либо Са с вероятностью p_{Ca} ($p_K + p_{Na} + p_{Ca} = 1$). Рассматриваемая последовательность представляет собой однородную (по предположению) простую марковскую цепь.

Действительно, рассмотрим переход через А. Какой бы ни была предыстория, после А с вероятностями $p_I, I \in \{K, Na, Ca\}$ могут появиться либо К, либо Na, либо Са. Если же взять переход через К (Na или Са), то независимо от предыстории далее, с вероятностью единица, следует А. Так как при четырех состояниях трудно дать графическую интерпретацию явления, то мы не будем рассматривать перемещение точки по грани тетраэдра, как это делалось до сих пор. Вместо этого рассмотрим поведение последовательности матриц P, P^2, P^3, \dots

Как уже отмечалось ранее, цепь называется периодической, если из произвольного состояния I можно вернуться в I лишь за $h, 2h, 3h, \dots$ шагов. Период h , если он имеется, должен быть один и тот же для всех состояний. В данном случае очевидно, что из А в А можно попасть за произвольное четное число шагов. Это же справедливо для К, Na и Са.

Таким образом, мы имеем пример простой периодической цепи Маркова с периодом $h=2$. Положим для простоты, что число ионов К, Na и Са одинаково. Тогда $p_K = p_{Na} = p_{Ca} = 1/3$.

Отсюда

$$P = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}, \quad P^3 = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Таким образом, для рассматриваемой последовательности имеем

$$P^3 = P, \quad P^4 = P^2, \quad P^5 = P^3 = P.$$

Итак, последовательность матриц не стремится ни к какому пределу, поскольку в этой последовательности имеются две ма-

трицы P и P^2 , чередующиеся неограниченное число раз, что исключает возможность существования предела. Точно так же, каким бы мы ни взяли начальное распределение p_0 для последовательности состояний, мы получим

$$p_0; p_1 = P^r p_0; p_2 = (P^r)^2 p_0; p_3 = (P^r)^3 p_0 = P p_0 = p_1; p_4 = (P^r)^2 p_0 = p_2, \dots$$

т. е. имеются две предельные точки — p_1 и p_2 .

В настоящем примере мы имели точную повторяемость матрицы и вектора. Это явление связано со специфическими значениями собственных чисел в данном случае и не обязательно для периодических цепей вообще. Характерна же для периодических цепей асимптотическая периодическая повторяемость. Это понятие будет рассмотрено далее.

До сих пор мы рассматривали поведение последовательности векторов p_0, p_1, \dots, p_k . Рассмотрим теперь поведение последовательностей матриц

$$P, P^2, \dots, P^k \text{ и } P^r, (P^r)^2, \dots, (P^r)^k.$$

Естественно, что поведение последовательностей векторов и последовательностей матриц тесно связано.

Если цепь эргодическая, то $\lim_{k \rightarrow \infty} P^k$ и $\lim_{k \rightarrow \infty} (P^r)^k = (\lim_{k \rightarrow \infty} P^k)^r$ существуют и, как отмечалось, $\lim_{k \rightarrow \infty} P^k$ представляет собой матрицу, отвечающую последовательности независимых испытаний (в будущем мы ее будем обозначать буквой B), т. е. матрицу вида

$$B = \begin{pmatrix} p_1 & p_2 & \dots & p_s \\ p_1 & p_2 & \dots & p_s \\ \dots & \dots & \dots & \dots \\ p_1 & p_2 & \dots & p_s \end{pmatrix}.$$

Но для матриц вида B результат операции $B^r p$ не зависит от того, какой стохастический вектор p был взят (в результате всегда получается вектор

$$\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_s \end{pmatrix},$$

т. е. транспонированная строка матрицы B). Этим объясняется то обстоятельство, что в эргодической цепи начальное и предельное распределения не связаны друг с другом.

Если цепь является квазиэргодической, то $\lim_{k \rightarrow \infty} P^k$ существует, но не является матрицей, отвечающей последовательности независимых испытаний. Поэтому результат действия $P^r p$ зависит от выбора p . Таким образом, предельное распределение в квазиэргодической цепи связано с начальным.

Наконец, если цепь периодическая, то $\lim_{k \rightarrow \infty} P^k$ существовать не может. Действительно, диагональные элементы степеней матрицы $P^h, P^{2h}, \dots, P^{rh}$, т. е. $p_{ii}^{(h)}, p_{ii}^{(2h)}, \dots, p_{ii}^{(rh)}$, больше нуля, а при других показателях степени, некратных h , эти же элементы равны нулю. В последовательности P, P^2, \dots на местах диагональных элементов периодически повторяются нули и положительные числа. Следовательно, никакой единый предел не устанавливается.

Мы уже отмечали ранее, как классифицируются по вероятностным свойствам марковские цепи. При этом мы выделяли эргодические, периодические, приводимые и неприводимые цепи. Такую же классификацию можно провести по характеристическим числам матрицы P . Для этого используется критерий классификации цепей.

Цепь является периодической в том и только в том случае, когда среди характеристических чисел матрицы P имеется число λ , не совпадающее с единицей, но по модулю равное ей. Если таких чисел несколько, то все они являются корнями из единицы одной и той же степени h , которая представляет период цепи. Для квазиэргодичности цепи необходимо и достаточно, чтобы единица была корнем характеристического полинома, кратности большей, чем первая. Наконец, если цепь не является ни периодической, ни квазиэргодической, то она является эргодической.

Доказательство опускаем.

В практических приложениях приведенное негативное определение эргодичности не удобно и обычно заменяется следующим условием. *Если существует целое положительное k , такое, что все элементы матрицы P^k положительны, то цепь эргодична.*

Приводимость цепи устанавливается на основании следующего правила: *цепь приводима в том и только в том случае, когда она является квазиэргодической.*

Применим изложенные критерии к матрицам из примеров IV.7—IV.9.

Для матрицы, отвечающей последовательности зерен в аляските с р. Каракульджур (IV.3.5), характеристическое уравнение имеет вид:

$$\begin{vmatrix} 0.3 - \lambda & 0.5 & 0.2 \\ 0.3 & 0.4 - \lambda & 0.3 \\ 0.4 & 0.6 & -\lambda \end{vmatrix} = 0,$$

т. е.

$$\lambda^3 - 0.70\lambda^2 - 0.29\lambda - 0.01 = 0.$$

Поскольку $\lambda = 1$ всегда является характеристическим числом стохастической матрицы, то характеристический полином можно сократить на $\lambda - 1$. В результате получим

$$\lambda^2 + 0.30\lambda + 0.01 = 0,$$

откуда

$$\lambda_1 = 1, \lambda_2 = -0.2618 \text{ и } \lambda_3 = -0.0383.$$

Поскольку корней с единичным модулем, кроме λ_1 , нет, матрица (IV.3.5) изученных аляскистов отвечает эргодической марковской цепи.

Еще проще мы могли бы убедиться в этом, возводя в степень матрицу P . Действительно, уже возведение ее в квадрат дает

$$P^2 = \begin{pmatrix} 0.32 & 0.47 & 0.21 \\ 0.33 & 0.49 & 0.18 \\ 0.30 & 0.44 & 0.26 \end{pmatrix}.$$

Поскольку P^2 не содержит нулевых элементов, P отвечает эргодической цепи Маркова.

Рассмотрим теперь пример IV.8, посвященный формированию нефтяной залежи. Матрица (IV.3.7) имеет характеристическое уравнение

$$(p - \lambda)(1 - \lambda)(1 - \lambda) = 0,$$

откуда

$$\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = p.$$

Как видим, цепь квазиэргодическая, так как единица является двукратным корнем характеристического полинома. Цепь является также приводимой, поскольку она имеет замкнутое подмножество $\{H, \Pi\}$, попав в которое система выйти из него уже не может.

В примере IV.9, посвященном свойствам осумилита, в качестве характеристического уравнения мы имеем

$$\begin{vmatrix} -\lambda & 1/3 & 1/3 & 1/3 \\ 1 & -\lambda & 0 & 0 \\ 1 & 0 & -\lambda & 0 \\ 1 & 0 & 0 & -\lambda \end{vmatrix} = 0,$$

откуда

$$\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = 0, \lambda_4 = 0.$$

Как видим, среди характеристических чисел имеется λ_2 , по модулю равное единице, но не совпадающее с ней. Наименьшая целая степень λ_2 , дающая единицу, есть $h=2$. Таким образом, в примере IV.9 цепь является периодической с периодом $h=2$.

IV.3.3. Некоторые расчеты, связанные с ценами Маркова

В приложениях, особенно при работе с эмпирическими матрицами переходных вероятностей, почти всегда имеет место случай, когда все характеристические числа матрицы P различны. Для этого случая большое значение имеет следующая формула,

разлагающая степени матрицы P^k на слагаемые, содержащие не зависящие от k постоянные матрицы и характеристические числа. Пусть $\lambda_1, \lambda_2, \dots, \lambda_s$ — характеристические числа матрицы P . Согласно предположению, они различны. Будем обозначать характеристический вектор-столбец матрицы P , отвечающий характеристическому числу λ_i , как x_i , а характеристический вектор-столбец матрицы P^T , отвечающий тому же λ_i , как y_i . Произведение $x_i y_i^T$ представляет собой матрицу размера $s \times s$. В то же время произведение $x_i^T y_i$ представляет собой число (скалярное произведение векторов x_i, y_i).

Тогда

$$P^k = \sum_{i=1}^s \lambda_i^k \frac{k_i y_i^T}{x_i^T y_i}. \quad (\text{IV. 3. 8})$$

Поскольку $\lambda_1 = 1$, $x_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ и $y_1 = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_s \end{pmatrix}$, то первое слагаемое в (IV. 3. 8) есть

$$\begin{pmatrix} \frac{p_1}{\sum_{i=1}^s p_i} & \dots & \frac{p_s}{\sum_{i=1}^s p_i} \\ \dots & \dots & \dots \\ \frac{p_1}{\sum_{i=1}^s p_i} & \dots & \frac{p_s}{\sum_{i=1}^s p_i} \end{pmatrix},$$

т. е. первое слагаемое в (IV. 3. 8) является матрицей последовательности независимых испытаний вида B , строка которой представляет собой стационарное распределение. Учитывая далее, что каждая матрица $x_i y_i^T / x_i^T y_i$ не зависит от k , мы можем представить формулу (IV. 3. 8) в виде

$$P^k = B + \sum_{i=2}^s C_i \lambda_i^k, \quad (\text{IV. 3. 8a})$$

где B — матрица для последовательности Бернулли, C_i — некоторые постоянные матрицы, а λ_i — характеристические числа матрицы P .

Формулы (IV.3.8) и (IV.3.8a) удобны как для изучения структуры марковской цепи, так и для практических расчетов с марковскими цепями. Так, при вычислении высоких степеней P^k рационально использовать именно эту формулу, а не последователь-

ное возведение в соответствующие степени, при котором накапливаются ошибки. По формуле (IV.3.8) также легко определить, насколько быстро цепь устанавливается. Следует подчеркнуть, что (IV.3.8) не применима к квазиэргодическим цепям, потому что в ней все характеристические числа должны быть различны по условию. Рассмотрим некоторые случаи.

В эргодической цепи характеристические числа $\lambda_2, \dots, \lambda_s$ по модулю все меньше единицы. Поэтому $\lim_{k \rightarrow \infty} \lambda_i^k = 0$ ($i=2, 3, \dots, s$). Таким образом, в (IV.3.8) при $k \rightarrow \infty$ остается только предельный член B , представляющий, как отмечалось, матрицу последовательности Бернулли.

При исследовании периодической цепи возможны два варианта. В первом все характеристические числа $\lambda_2, \dots, \lambda_s$ при возведении в степень h не меняются, т. е. $\lambda_i^h = \lambda_i$. Тогда выражение для P^k в точности периодически повторяется через h шагов, т. е. $P^{h+s} = P^s$. Этот случай чистой периодичности фигурирует в примере IV.9, посвященном осумилиту. В нем $\lambda_1=1, \lambda_2=-1, \lambda_3=\lambda_4=0$, т. е. $\lambda_i^2 = \lambda_i$.

Более распространенным является случай, когда одни характеристические числа имеют период h , а другие не имеют. Так, например, в цепи на четырех состояниях мы могли бы иметь $\lambda_1=1, \lambda_2=0, \lambda_3=-1, 0 < \lambda_4=\alpha < 1$.

Тогда, согласно (IV.3.8),

$$P^k = B + (-1)^k C_3 + \alpha^k C_4.$$

Здесь слагаемые B и $(-1)^k C_3$ будут повторяться через каждые два шага, а слагаемое $\alpha^k C_4$ непериодично; в целом, матрица P^k не будет в точности воспроизводиться через каждые два шага. Слагаемое $\alpha^k C_4$ затухает с ростом k из-за того, что $0 < \alpha < 1$. В итоге степени P^k , отстоящие друг от друга на два шага, сначала будут отличаться более заметно, чем в дальнейшем — при больших значениях k . Таким образом, степени P^k будут асимптотически воспроизводиться через каждые два шага. В этом заключается общее поведение периодической цепи, для которого характерна асимптотическая повторяемость P^k через h шагов. Периодичность цепи иногда неверно понимают как точную повторяемость матрицы P через h шагов, что совершенно необязательно. Еще менее свойственна для цепи детерминированная периодичность реализации. Оба эти явления возможны, но представляют весьма частные случаи периодической марковской цепи.

Приведем расчеты по формуле (IV.3.8), исследуя последовательность зерен в аляските с р. Каракульджур, приведенную в примере IV.7. Используем для этого характеристические числа матрицы указанной последовательности $\lambda_1=1, \lambda_2=-0.262$ и $\lambda_3=-0.038$. Решая уравнения

$$P x_i = \lambda_i x_i \quad (i=1, 2, 3),$$

получаем

$$x_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1.000 \\ -6.124 \\ 12.497 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 1.000 \\ -0.665 \\ -0.028 \end{pmatrix}.$$

При вычислении собственных векторов для подстановки в формулу (IV.3.8) нет необходимости нормировать векторы, так как нормировка автоматически осуществится при делении на скалярное произведение $x_i^T y_i$ (см. (IV.3.8)).

Далее, из уравнения

$$P^T y_i = \lambda_i y_i \quad (i = 1, 2, 3)$$

имеем

$$y_1 = \begin{pmatrix} 1.000 \\ 1.476 \\ 0.643 \end{pmatrix}, \quad y_2 = \begin{pmatrix} 1.000 \\ 1.623 \\ -2.623 \end{pmatrix}, \quad y_3 = \begin{pmatrix} 1.000 \\ -0.618 \\ -0.0382 \end{pmatrix}.$$

Согласно формуле (IV.3.8),

$$B = \begin{pmatrix} 0.321 & 0.474 & 0.205 \\ 0.321 & 0.474 & 0.205 \\ 0.321 & 0.474 & 0.205 \end{pmatrix}, \quad C_2 = \begin{pmatrix} -0.024 & -0.039 & 0.063 \\ 0.147 & 0.238 & -0.385 \\ -0.299 & -0.486 & 0.785 \end{pmatrix},$$

$$C_3 = \begin{pmatrix} 0.706 & -0.435 & -0.269 \\ -0.468 & 0.289 & 0.179 \\ -0.020 & 0.012 & 0.008 \end{pmatrix}.$$

Таким образом, в численном виде формула (IV.3.8) для последовательности зерен из аляскита с р. Каракульджур имеет вид:

$$P^k = \begin{pmatrix} 0.321 & 0.474 & 0.205 \\ 0.321 & 0.474 & 0.205 \\ 0.321 & 0.474 & 0.205 \end{pmatrix} + (-0.262)^k \begin{pmatrix} -0.024 & -0.039 & 0.063 \\ 0.147 & 0.238 & -0.385 \\ -0.299 & -0.486 & 0.785 \end{pmatrix} +$$

$$+ (-0.038)^k \begin{pmatrix} 0.706 & -0.435 & -0.269 \\ -0.468 & 0.289 & 0.179 \\ -0.020 & 0.012 & 0.008 \end{pmatrix}.$$

При $k=2$ (и округлении до второго знака) формула (IV.3.8) дает

$$P^2 = \begin{pmatrix} 0.32 & 0.47 & 0.21 \\ 0.33 & 0.49 & 0.18 \\ 0.30 & 0.44 & 0.26 \end{pmatrix},$$

т. е. то, что было получено ранее прямым возведением P во вторую степень.

Пример IV.10. О структуре
ковариационных последовательностей
для чередования мощностей
слоев некоторых осадочных толщ

Одной из важнейших проблем седиментологии является вопрос связи между мощностями следующих друг за другом слоев. При этом возможны различные гипотезы. Можно предполагать, что мощность слоев не зависит от их состава и чередование мощностей слоев не отражает чередования слоев по составу. Можно допустить, что существует специальный механизм, определяющий взаимоотношения между мощностями слоев. Можно, наконец, считать, что никакого специального механизма нет, а мощности слоев зависят от состава слоев, и чередование составов слоев целиком определяет соотношение между их мощностями. Вопрос этот может быть уточнен с помощью исследований, опирающихся на формулу (IV.3.8), позволяющую изучить связь между последовательностью составов слоев и отвечающих им мощностей.

Ниже с иллюстративной целью рассматривается упрощенная ситуация, когда слои a_i могут состоять только из двух типов осадков — слой состоит либо из песка (π), либо из глинистого осадка (γ), $i \in \{\pi, \gamma\}$. При этом чередование составов слоев представляет простую однородную цепь Маркова. Этой последовательности слоев отвечает последовательность их мощностей (каждому слою отвечает своя мощность)

$$x_1, x_2, \dots, x_n,$$

которую мы считаем стационарной (см. IV.2 и главу II).

Связь между мощностями слоев, отстоящими друг от друга на r слоев (например, первый и третий слой отстоят на 2), измеряется с помощью общепринятого коэффициента, называемого ковариацией,

$$K(r) = E(x_s x_{s+r}) - E x_s E x_{s+r},$$

где E — оператор математического ожидания.

Последовательность чисел

$$K(1), K(2), \dots, K(r) \tag{IV.3.9}$$

называется ковариационной последовательностью и может быть использована для исследования соотношений между мощностями слоев (в литературе график, отвечающий ковариационной последовательности, часто называют коррелограммой, особенно в тех случаях, когда значения $K(r)$ нормированы путем деления на $K(0)$; иногда график нормированных значений $K(r)$ называют графиком автокорреляционной или просто корреляционной функции), что было объяснено в главе II.

Выдвигая гипотезу о том, что наличие связи между мощностями слоев вызвано исключительно тем, что составы слоев связаны в простую однородную цепь Маркова, мы тем самым предполагаем,

что при фиксации составов слоев и расстояния между ними ковариации мощностей оказываются равными нулю.

Иными словами, мы выдвигаем гипотезу о том, что

$$E(x_s x_{s+r} | a_s = I; a_{s+r} = J) - E(x_s | a_s = I; a_{s+r} = J) E(x_{s+r} | a_s = I; a_{s+r} = J) = 0, \quad (\text{IV. 3. 10})$$

где a — состав слоя вообще, нижний индекс при a — номер слоя; $I, J \in \{\pi, \gamma\}$.

Гипотезу, выраженную (IV.3.10), можно проверить по данным послойного описания разрезов. Приняв (IV.3.10), мы можем определить общий вид ковариационной последовательности (IV.3.9), не зная ни конкретных значений математических ожиданий мощностей слоев фиксированного состава, ни матрицы переходных вероятностей для последовательности составов слоев. Вид корреляционной последовательности также не зависит от одномерных распределений вероятностей мощностей.

Теорема IV.2. Если последовательность составов слоев образует простую однородную цепь Маркова с двумя состояниями при начальном распределении, совпадающем со стационарным, и если условные мощности слоев при фиксации их составов (в смысле (IV.3.10)) независимы друг от друга, то возможно возникновение только двух типов корреляционных последовательностей.

Первый тип, который мы назовем в соответствии с классификацией, данной в работе автора (Вистеллиус, 1949), типом Б, характеризуется тем, что все коэффициенты в (IV.3.9) одного знака и монотонно убывают по абсолютной величине с ростом r . Второго типа, который мы называем типом Ф, характеризуется тем, что коэффициенты $K(1), K(3), K(5), \dots$ имеют один знак, а коэффициенты $K(2), K(4), K(6) \dots$ — другой, причем с ростом r абсолютные значения ковариаций в каждом ряду монотонно затухают.

Тип Б возникает в том и только в том случае, когда сумма диагональных элементов в матрице переходных вероятностей для составов слоев больше единицы. Тип Ф возникает в том и только в том случае, когда сумма элементов по главной диагонали меньше единицы. В случае равенства этой суммы единице корреляционная последовательность состоит из одних нулей.

Доказательство. Введем обозначения: $\mathbf{m} = \begin{pmatrix} m_\pi \\ m_\gamma \end{pmatrix}$ — вектор математических ожиданий мощностей песчаного и глинистого слоя; $\mathbf{P} = \begin{pmatrix} P_{\pi\pi} & P_{\pi\gamma} \\ P_{\gamma\pi} & P_{\gamma\gamma} \end{pmatrix}$; $\mathbf{P}^{(r)}$ — матрица переходных вероятностей на r шагов; $p_{ij}^{(r)}$ — общий элемент этой матрицы; \mathbf{p} — вектор стационарного распределения, а p_i — его компонента. Вычислим теперь

$$K(r) = E(x_s x_{s+r}) - E(x_s) E(x_{s+r}). \quad (\text{IV. 3. 11})$$

В силу стационарности последовательности мощностей слоев (IV.3.11) можно представить в виде

$$E(x_1 x_{1+r}) - (E x_1)^2,$$

а так как начальное и стационарное распределения совпадают, то

$$E x_1 = \sum_{i \in \{\pi, \gamma\}} p_i m_i.$$

Тогда $E^2 x_1$ можно представить как

$$\sum_{i, j \in \{\pi, \gamma\}} p_i p_j m_i m_j.$$

В то же время из-за независимости мощностей друг от друга при фиксации составов слоев и расстояния

$$E(x_1 x_{1+r}) = \sum_{i, j \in \{\pi, \gamma\}} p_i m_i p_j^{(r)} m_j. \quad (\text{IV. 3. 12})$$

Обращаясь к формуле (IV. 3. 8), находим, что

$$P^{(r)} = B + \lambda_2^r C_2, \quad (\text{IV. 3. 13})$$

где B — матрица с общим элементом $b_{ij} = p_j$; C_2 — некоторая матрица, не зависящая от r ; характеристические числа λ_i могут принимать значения только $\lambda_1 = 1$ и λ_2 , причем $-1 < \lambda_2 < 1$, так как последовательность составов слоев обладает эргодическим свойством.

Из (IV.3.13) следует, что $p_j^{(r)}$ разлагается на два слагаемых:

$$p_j^{(r)} = p_j + \lambda_2^r c_{ij}, \quad (\text{IV. 3. 14})$$

где c_{ij} — общий элемент матрицы C_2 .

Подставляя (IV.3.14) в (IV.3.8), получаем

$$E(x_1 x_{1+r}) = \sum_{i, j \in \{\pi, \gamma\}} p_i p_j m_i m_j + \sum_{i, j \in \{\pi, \gamma\}} p_i m_i c_{ij} \lambda_2^r m_j,$$

откуда

$$K(r) = \lambda_2^r \sum_{i, j \in \{\pi, \gamma\}} p_i m_i c_{ij} m_j. \quad (\text{IV. 3. 15})$$

Суммирование в (IV.3.15) дает некоторую постоянную c , не зависящую от r .

Таким образом,

$$K(r) = c \lambda_2^r.$$

Рассмотрим теперь матрицу переходных вероятностей

$$P = \begin{pmatrix} P_{\pi\pi} & P_{\pi\gamma} \\ P_{\gamma\pi} & P_{\gamma\gamma} \end{pmatrix}.$$

Равенство $P_{\pi\pi} + P_{\gamma\gamma} = 1$ невозможно, так как из этого равенства следовало бы, что $\lambda_2 = 0$. Тогда формула (IV.3.13) привела бы к $P =$

$= B$, откуда следовало бы, что изучаемая последовательность является последовательностью Бернулли, в то время как по предположению она является простой цепью Маркова. Поэтому либо $p_{\pi\pi} + p_{\gamma\gamma} > 1$, либо $p_{\pi\pi} + p_{\gamma\gamma} < 1$. Поскольку $p_{\pi\pi} + p_{\gamma\gamma} = \lambda_1 + \lambda_2 = 1 + \lambda_2$, в первом случае λ_2 оказывается положительным числом < 1 , во втором случае λ_2 — отрицательное число по модулю < 1 . Из этого следует, что при $p_{\pi\pi} + p_{\gamma\gamma} > 1$ формула (IV.3.15) приводит к ковариационной последовательности типа

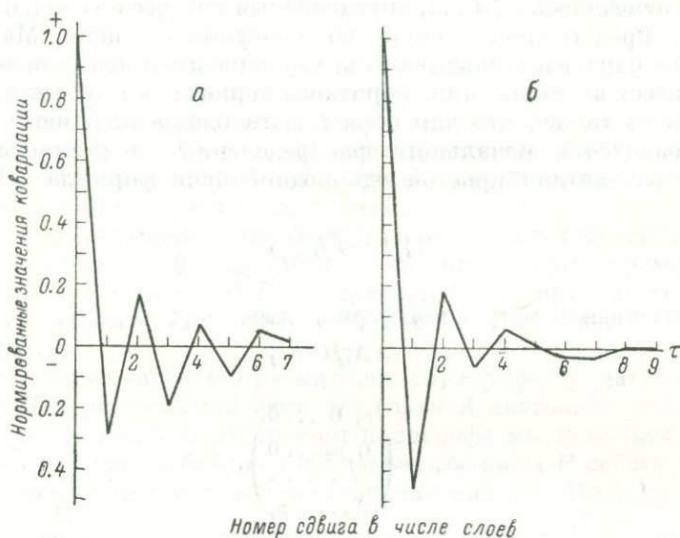


Рис. IV.7. Наблюдаемые ковариационные последовательности для флишевых отложений.

Отложения: а — каменноугольного возраста Южного Урала, б — верхнемелового возраста северо-западного Кавказа (Новороссийск).

Б. При $p_{\pi\pi} + p_{\gamma\gamma} < 1$ эта формула дает корреляционную последовательность типа Ф. Итак, теорема доказана.

Как правило, при изучении разрезов мы выделяем не два, а три типа составов слоев. Для трех состояний в марковской цепи кроме типов Б и Ф могут возникать также другие типы корреляционных последовательностей. Однако типы Б и Ф могут иметь место также и при трех состояниях. Действительно, это имеет место, если из трех математических ожиданий мощностей слоев два совпадают или если последовательность составов слоев остается простой марковской, когда два из трех состояний кодируются одним признаком (что будет подробно рассмотрено в гл. V). Так или иначе, но формула (IV.3.8) дает удобный инструмент для изучения связей между мощностями в случаях, когда эта связь индуцирована через свойства последовательностей составов слоев.

Изучение конкретных разрезов с выделением трех состояний (песок, алевроит, глина) показало, что эмпирический тип Б, выделенный в 1949 г., не отвечает теоретической схеме, разобранный здесь, и заметно отличается от предсказанной последовательности ковариаций. Тип Ф, выделенный тогда же, очень близок к тому, что получается теоретически; рис. IV.7 это подтверждает.

IV.3.4. Обратимость, периодичность и число состояний

Как отмечалось в § IV.2, определенный интерес в геологических задачах представляет вопрос об обратимости цепи Маркова. В том же параграфе указывалось, что обратимая цепь может быть периодической лишь при коротком периоде — $h=1$ или $h=2$. Отмечалось также, что для обратимости однородной цепи требуется совпадение начального распределения со стационарным. В этом случае для простой однородной цепи формула (IV.2.19)

$$P_{ij}^{(h)} = P_{ji} \frac{P_i}{P_j},$$

представленная в матричной форме, имеет вид

$$P^{(-)} = \Lambda_{st}^{-1} P^* \Lambda_{st}, \quad (IV.3.16)$$

где

$$\Lambda_{st} = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & p_s \end{pmatrix},$$

а P_i — компоненты вектора стационарного распределения.

Обратимость однородной марковской цепи имеет место в том и только в том случае, когда $P^{(-)} = P$. Последнее равенство с учетом (IV.3.16) выражается в виде

$$\Lambda_{st} P = P^* \Lambda_{st}. \quad (IV.3.17)$$

Формула (IV.3.17) является необходимым и достаточным условием обратимости однородной эргодической цепи Маркова.

Подтвердим приложение (IV.3.17) следующим доказательством. Эргодическая цепь на двух состояниях всегда является обратимой.

Действительно, пусть

$$P = \begin{pmatrix} \pi_1 & 1 - \pi_1 \\ \pi_2 & 1 - \pi_2 \end{pmatrix}.$$

тогда

$$P_{st} = \begin{pmatrix} \frac{\pi_2}{1 - \pi_1 + \pi_2} & \\ & 1 - \pi_1 \\ \frac{1 - \pi_1}{1 - \pi_1 + \pi_2} & \end{pmatrix}, \quad \Lambda_{st} = \begin{pmatrix} \frac{\pi_2}{1 - \pi_1 + \pi_2} & 0 \\ 0 & \frac{1 - \pi_1}{1 - \pi_1 + \pi_2} \end{pmatrix}.$$

Уравнение (IV.3.17) в данном случае принимает вид

$$\begin{pmatrix} \frac{\pi_2}{1-\pi_1+\pi_2} & 0 \\ 0 & \frac{1-\pi_1}{1-\pi_1+\pi_2} \end{pmatrix} \begin{pmatrix} \pi_1 & 1-\pi_1 \\ \pi_2 & 1-\pi_2 \end{pmatrix} = \\ = \begin{pmatrix} \pi_1 & \pi_2 \\ 1-\pi_1 & 1-\pi_2 \end{pmatrix} \begin{pmatrix} \frac{\pi_2}{1-\pi_1+\pi_2} & 0 \\ 0 & \frac{1-\pi_1}{1-\pi_1+\pi_2} \end{pmatrix}$$

и, очевидно, выполняется.

Таким образом, при двух состояниях обратимость имеет место независимо от конкретного вида матрицы переходных вероятностей.

Приведенный пример показывает, что некоторые особенности марковской цепи (в данном случае обратимость) могут быть связаны лишь с числом состояний.

Число состояний также существенно и для периодичности цепи. Так, для того чтобы однородная цепь без детерминированных переходов (т. е. с матрицей P без единиц) была периодической, необходимо, чтобы число состояний в ней было не менее четырех. Докажем это.

Как отмечалось в предыдущем параграфе, у матрицы, отвечающей периодической цепи, на главной диагонали должны располагаться нули. В настоящем параграфе мы показали, что для периодичности необходимо наличие у матрицы P собственных чисел с единичным модулем, не совпадающих с 1. Поэтому при $s=2$ периодическая цепь может иметь лишь матрицу $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, но эта матрица имеет только детерминированные переходы. При $s=3$ единственными возможными наборами характеристических чисел для периодической цепи будут либо $\lambda_1=1, \lambda_2=-\frac{1}{2}+\frac{\sqrt{3}}{4}i, \lambda_3=-\frac{1}{2}-\frac{\sqrt{3}}{4}i$, либо $\lambda_1=1, \lambda_2=-1, \lambda_3=0$, что вытекает из указанных свойств характеристических чисел матрицы периодической цепи. $\lambda_2=-\frac{1}{2}+\frac{\sqrt{3}}{4}i$, а также $\lambda_3=-\frac{1}{2}-\frac{\sqrt{3}}{4}i$ являются корнями четвертой степени из единицы. Этому случаю отвечает период $h=4$. Однако цепь на трех состояниях не может иметь период, равный четырем. Действительно, допустим, что мы имеем три состояния — I, J и K — и что осуществляется состояние I , а затем J . Тогда в следующем испытании обязано появиться K (с вероятностью 1), а затем I (с вероятностью 1). Если бы после K могло появиться J , то период был бы $h \leq 2$, так как мы попали бы из J в J за два шага. Точно так же не может осуществиться на четвертом шаге переход в K , так как тогда мы имели бы переход из K в K за один шаг.

Таким образом, на четвертом испытании должно осуществиться I . Поскольку мы перешли из I в I за три шага, то период оказывается не более трех.

Набор характеристических чисел $\lambda_1=1$, $\lambda_2=-1$ и $\lambda_3=0$ отвечает цепи с периодом $h=2$. Чтобы в такой цепи было невозможно возвращение в исходное состояние за три шага, необходимо, чтобы для некоторых $I \neq J$ были $p_{IJ}=1$. Но такая цепь имеет детерминированный переход. При четырех состояниях, как можно убедиться на примерах, периодичность уже может иметь место.

IV.3.5. Степени свободы

Рассмотрим теперь вопрос о количестве независимых параметров, которые нужно задать, чтобы полностью определить марковскую цепь того или иного вида. Число таких независимых параметров в статистике называется числом степеней свободы. Это число нужно знать при применении статистических тестов (см. гл. VI).

Простая однородная цепь полностью определяется начальным распределением \mathbf{p}_0 и матрицей \mathbf{P} . Поэтому в общем случае в простой однородной марковской цепи количество степеней свободы

$$\nu = s - 1 + s(s - 1) = (s + 1)(s - 1) = s^2 - 1,$$

где S — число состояний.

В геологии, как правило, работают с установившимися эргодическими цепями, у которых за начальное распределение принимается стационарное. Стационарное распределение, как известно, однозначно определяется матрицей \mathbf{P} . Таким образом, установившаяся однородная простая марковская цепь имеет количество степеней свободы

$$\nu = s(s - 1).$$

На практике также часто встречается класс обратимых эргодических стационарных цепей. Поскольку связи, налагаемые обратимостью, зависят от связей, налагаемых марковской структурой матрицы \mathbf{P} , подсчет ν здесь требует специального рассмотрения.

Т е о р е м а IV.3. Для обратимых эргодических стационарных простых цепей Маркова число степеней свободы

$$\nu = \frac{s(s + 1)}{2} - 1.$$

Д о к а з а т е л ь с т в о. Очевидно, что параметрами являются элементы матрицы \mathbf{P} , так как \mathbf{P} определяет стационарный вектор \mathbf{p}_{st} . Рассмотрим сначала произвольную матрицу \mathbf{M} размера $s \times s$ с неотрицательными элементами. Такая матрица имеет s^2 -степеней свободы, поскольку требование неотрицательности

элементов задается неравенствами, что не снижает числа степеней свободы. Допустим также, что M имеет одно из характеристических чисел, равное единице. Тогда

$$M^s p_{st} = p_{st} \quad (IV. 3. 18)$$

и

$$\Lambda_{st} e = p_{st}, \quad (IV. 3. 19)$$

где e — вектор-столбец, все s компонент которого — единицы.

Пусть, наконец, для матрицы M выполняется уравнение обратимости (IV. 3. 17), т. е.

$$\Lambda_{st} M = M^s \Lambda_{st}. \quad (IV. 3. 20)$$

Тогда матрица M обязана быть марковской. Покажем это.

Для квадратной матрицы с неотрицательными элементами принадлежность матрицы к классу марковских эквивалентна выполнению уравнения $Me = e$. Умножая (IV. 3. 20) справа на e , получаем

$$\Lambda_{st} Me = M^s \Lambda_{st} e$$

или, в силу (IV. 3. 17),

$$M^s p_{st} = \Lambda_{st} Me,$$

откуда вследствие (IV. 3. 1)

$$p_{st} = \Lambda_{st} Me$$

или

$$\Lambda_{st}^{-1} p_{st} = Me.$$

Поскольку

$$\Lambda_{st}^{-1} p_{st} = e,$$

получаем

$$Me = e.$$

Таким образом, ограничения, налагаемые уравнениями обратимости (IV. 3. 17) вместе с требованием $\lambda = 1$, уже содержат условие того, что сумма по строкам будет давать единицу. Условие $\lambda = 1$ дает одно уравнение

$$\begin{vmatrix} p_{11} - 1 & p_{12} & \dots & p_{1s} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ p_{s1} & \dots & \dots & p_{ss} - 1 \end{vmatrix} = 0.$$

Если представить уравнение обратимости (IV. 3. 17) в виде

$$p_{ij} = p_{ji} \frac{p_j}{p_i}, \quad (IV. 3. 21)$$

то станет ясно, что (IV. 3. 21) содержит s тождеств и $\frac{s^2 - s}{2}$ уравнений. Действительно, при $i = j$ в (IV. 3. 21) получаем тождество, а для пары i, j ($i \neq j$) — уравнение, такое же, как для пары j, i .

Таким образом, матрица M имеет s^2 параметров, для которых должно выполняться $\frac{s^2 - s}{2} + 1$ уравнений. Все эти уравнения независимы, т. е. всегда можно построить матрицу, удовлетворяющую любой части этих уравнений, но не всем уравнениям.

В итоге имеем

$$\nu = s^2 - \frac{s^2 - s}{2} - 1 = \frac{s^2 + s}{2} - 1, \quad (\text{IV. 3. 22})$$

чем и завершается доказательство.

Пример IV.11. Рассмотрим установившуюся однородную обратимую простую цепь Маркова на двух состояниях. Формула (IV.3.22) дает для этого случая $\nu=2$, что может показаться странным, так как любая установившаяся (даже необратимая) цепь определяется двумя параметрами. Это объясняется тем, что требование обратимости автоматически удовлетворяется в цепи на двух состояниях и, таким образом, обратимость не отнимает степеней свободы.

Иначе обстоит дело при трех состояниях. Тогда, как показывает (IV.3.22), $\nu=5$, в то время как для необратимой цепи марковская матрица имеет шесть степеней свободы. В итоге при трех состояниях обратимость приводит к потере одной степени свободы.

IV.3.6. Вычисления со сложными цепями Маркова

Специфика цепей второго порядка достаточно полно отражает специфику сложных цепей Маркова вообще, поэтому далее мы рассмотрим только однородные цепи Маркова второго порядка. Более того, мы рассмотрим только эргодические цепи — такие, у которых каждое состояние является достижимым из любого другого.

Изучение сложных цепей в основных чертах может быть сведено к изучению простой цепи. Для этого используют искусственный прием, связанный с переходом к большему числу состояний. Так, для однородной цепи второго порядка на s -состояниях образуют всевозможные пары состояний $(a_i^{(h)}, a_j^{(h+1)})$, где первый элемент пары отвечает появлению состояния I в h -й момент, а второй — J — в $h+1$ -й момент. Таких пар будет s^2 . Каждую из этих пар мы определим как некоторое новое состояние. Эти состояния образуют множество состояний Σ .

Построим теперь новую случайную последовательность с переходом на один шаг от пары $(a_i^{(h)}, a_j^{(h+1)})$ к паре $(a_L^{(h+1)}, a_K^{(h+2)})$ ($I, J, K, L \in \mathcal{S}$). Подчеркнем, что при этом последний элемент предыдущей пары и начальный элемент последующей пары отвечают испытанию, происшедшему в один и тот же момент времени. Тогда переходная вероятность на один шаг будет $P_{I, J; L, K}^{(h), (h+1), (h+2)}$. В силу того что последовательность предполагается однородной, верхние индексы могут быть опущены. Поэтому переходная вероятность далее будет обозначаться $P_{I, J; L, K}$. Положим, согласно определению,

что

$$P_{I, J; L, K} = 0, \text{ если } J \neq L,$$

и

(IV. 3. 23)

$$P_{I, J; L, K} = P_{I, J; K}, \text{ если } J = L.$$

Вероятности $P_{I, J; K}$ можно собрать в прямоугольную матрицу размером $s^2 \times s$. В то же время случайной последовательности пар на множестве Σ отвечает квадратная матрица P_{Σ} с общим элементом $P_{I, J; L, K}$, соответствующим (IV.3.23).

Образованная указанным способом случайная последовательность пар из Σ с матрицей P_{Σ} называется двузвенной цепью. По построению очевидно, что указанная последовательность будет простой цепью в том и только в том случае, когда исходная цепь на множестве $\{\mathcal{S}\}$ имела порядок марковости не выше второго (нулевой, первый или второй).

Построив квадратную матрицу P_{Σ} , получим матрицу простой цепи, изученную нами в предыдущих пунктах § IV.3. С ней можно работать так, как это было указано в § IV.3. Таким образом, изучение цепи второго порядка в значительной мере может быть сведено к изучению цепи первого порядка.

Очевидно, что в общем случае однородной марковской цепи второго порядка матрица P_{Σ} не содержит всей информации относительно цепи на \mathcal{S} , она дает лишь вероятности перехода с пары последовательных испытаний на третье. Однако как начальное распределение для цепи на множестве \mathcal{S} , так и переходные вероятности внутри пары состояний не определяются этой матрицей. Таким образом, в общем случае для определения однородной марковской цепи второго порядка надо еще задать p_0 и $P_{I, J}$ на \mathcal{S} .

В случае стационарной цепи с $p_0 = p_{st}$ матрица P_{Σ} содержит всю информацию относительно случайной последовательности на \mathcal{S} . Вектор p_0 и матрицу $P_{I, J}$ мы получим из P_{Σ} путем соответствующих суммирований. Делается это так. Найдем сначала вектор стационарного распределения π_{st} , отвечающий P_{Σ} . Его компоненты будем обозначать π_{IJ} (это вектор с s^2 -компонентами соответственно числу пар состояний). Тогда I -тая компонента стационарного вектора исходной цепи

$$p_I = \sum_{J \in \mathcal{S}} \pi_{IJ} \text{ или } p_I = \sum_{J \in \mathcal{S}} \pi_{JI}. \quad (\text{IV. 3. 24})$$

Совпадение вычислений по указанным формулам можно использовать для контроля расчета компонент характеристического вектора π_{IJ} . Элементы матрицы переходных вероятностей $P_{I, J}$ вычисляются как

$$P_{I; J} = \frac{\pi_{IJ}}{p_I}. \quad (\text{IV. 3. 25})$$

Напомним, что деление в последней формуле возможно, так как мы приняли исходную цепь эргодической, что приводит к $p_I > 0$.

Таким образом, построив по матрице P_Σ или по содержащей эквивалентную ей информацию матрице $P_{I,J}$, K матрицу P_{IJ} и p_{st} , мы получим всю информацию об исходной цепи на множестве \mathcal{S} . Распределение вероятностей исходов после каждого испытания вычислять здесь не нужно, так как это распределение все время будет совпадать со стационарным.

В случае неустановившейся однородной цепи (начальное распределение не совпадает со стационарным) после каждого испытания будет существовать свое распределение вероятностей исходов до тех пор, пока цепь не стабилизируется. В этом случае распределение вероятностей состояний после k шагов, т. е. $p^{(k)}$, вычисляется для марковской цепи второго порядка следующим образом. Предполагается, что исходные параметры $p^{(0)}$, $P_{I,J}$ и P_Σ даны. Тогда после первого испытания мы будем иметь распределение $p^{(0)}$, а после второго

$$p^{(1)} = P_{I,J}^T p^{(0)}. \quad (\text{IV. 3. 26})$$

Далее, образуем вектор $\pi^{(2)}$, содержащий s^2 -компонент, представляющих распределение вероятностей всех пар в первом и втором испытаниях. Компоненты этого вектора подсчитываются по формуле

$$\pi_{IJ}^{(2)} = p_{IJ}^{(0)} p_{IJ}. \quad (\text{IV. 3. 27})$$

С помощью вектора $\pi^{(2)}$ можно найти распределение вероятностей состояний после третьего шага. Для этого находим вектор (с s^2 -компонентами)

$$\pi^{(3)} = P_\Sigma^T \pi^{(2)}. \quad (\text{IV. 3. 28})$$

По вектору $\pi^{(3)}$ определяем распределение вероятностей состояний в соответствии с формулой

$$p_J^{(3)} = \sum_{I \in \mathcal{S}} \pi_{IJ}^{(3)}. \quad (\text{IV. 3. 29})$$

Аналогично

$$\pi^{(4)} = (P_\Sigma^T)^2 \pi^{(2)} = P_\Sigma^T \pi^{(3)}, \quad (\text{IV. 3. 30})$$

и распределение состояний после четвертого шага будет иметь общую компоненту

$$p_J^{(3)} = \sum_{I \in \mathcal{S}} \pi_{IJ}^{(4)}, \quad (\text{IV. 3. 31})$$

и вообще

$$p_J^{(k)} = \sum_{I \in \mathcal{S}} \pi_{IJ}^{(k+1)}. \quad (\text{IV. 3. 32})$$

Вектор $\pi^{(k+1)}$ определяется из уравнения

$$\pi^{(k+1)} = (P_\Sigma^T)^{k-1} \pi^{(2)}, \quad (\text{IV. 3. 33})$$

где $\pi^{(k)}$ предполагается вычисленным ранее с помощью указанного рекуррентного правила.

Пример IV.12. Вычисления, связанные со сложной марковской цепью, мы проиллюстрируем на примере последовательности зерен биотита (Bt), пироксена (Pyr) и плагиоклаза (Pl) из габбро-диабазы района Гульшад. На с. 243, 244 (пример IV.5) мы установили структуру соответствующей случайной последовательности и показали, что это однородная марковская цепь второго порядка. Рассмотрим далее два случая.

В первом случае предположим, что цепь уже установилась. Во втором случае, где мы этого предполагать не будем, начальное распределение не будет совпадать со стационарным.

Как отмечалось, вычисления в этих случаях производятся по-разному.

Рассмотрим сначала случай, когда предполагается стационарность. Наблюдения дали следующие матрицы переходных частот:

$$N_{I, J; K} = \begin{matrix} & & \begin{matrix} h \\ h-2 & h-1 & \text{Bt} & \text{Pyr} & \text{Pl} \end{matrix} \\ \begin{matrix} \text{Bt} \\ \text{Bt} \\ \text{Bt} \\ \text{Pyr} \\ \text{Pyr} \\ \text{Pl} \\ \text{Pl} \\ \text{Pl} \end{matrix} & \begin{matrix} \text{Bt} \\ \text{Pyr} \\ \text{Pl} \\ \text{Bt} \\ \text{Pyr} \\ \text{Pl} \\ \text{Pl} \end{matrix} & \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 8 \\ 2 & 5 & 34 \\ 0 & 0 & 9 \\ 0 & 4 & 10 \\ 8 & 25 & 60 \\ 0 & 10 & 33 \\ 7 & 9 & 75 \\ 31 & 62 & 129 \end{bmatrix} \end{matrix}, \quad P_{I, J; K} = \begin{matrix} & & \begin{matrix} h \\ h-2 & h-1 & \text{Bt} & \text{Pyr} & \text{Pl} \end{matrix} \\ \begin{matrix} \text{Bt} \\ \text{Bt} \\ \text{Bt} \\ \text{Pyr} \\ \text{Pyr} \\ \text{Pl} \\ \text{Pl} \\ \text{Pl} \end{matrix} & \begin{matrix} \text{Bt} \\ \text{Pyr} \\ \text{Pl} \\ \text{Bt} \\ \text{Pyr} \\ \text{Pl} \\ \text{Pl} \end{matrix} & \begin{bmatrix} 0 & 0 & 0 \\ 0.1000 & 0.1000 & 0.8000 \\ 0.0488 & 0.1220 & 0.8293 \\ 0.0000 & 0.0000 & 1.0000 \\ 0.0000 & 0.2857 & 0.7143 \\ 0.0860 & 0.2688 & 0.6452 \\ 0.0000 & 0.2326 & 0.7674 \\ 0.0769 & 0.0989 & 0.8242 \\ 0.1396 & 0.2793 & 0.5811 \end{bmatrix} \end{matrix} \quad (6)$$

Прежде всего построим матрицу⁷ P_{Σ} для двухзвенной цепи в соответствии с формулой (IV.3.23):

$$P_{\Sigma} = \begin{matrix} & & \begin{matrix} h \\ h-2; & h-1 & \text{BtBt} & \text{BtPyr} & \text{BtPl} & \text{PyrBt} & \text{PyrPyr} & \text{PyrPl} & \text{PlBt} & \text{PlPyr} & \text{PlPl} \end{matrix} \\ \begin{matrix} \text{Bt} \\ \text{Bt} \\ \text{Bt} \\ \text{Pyr} \\ \text{Pyr} \\ \text{Pl} \\ \text{Pl} \\ \text{Pl} \end{matrix} & \begin{matrix} \text{Bt} \\ \text{Pyr} \\ \text{Pl} \\ \text{Bt} \\ \text{Pyr} \\ \text{Pl} \\ \text{Pl} \end{matrix} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.10000 & 0.10000 & 0.8000 & 0.0000 & 0.0000 & 0.0000 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0488 & 0.1220 & 0.8293 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2857 & 0.7143 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0860 & 0.2688 & 0.6452 \\ 0 & 0.2326 & 0.7674 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0769 & 0.0989 & 0.8242 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1396 & 0.2793 & 0.5811 \end{bmatrix} \end{matrix} \quad (IV.3.34)$$

⁶ Поскольку вероятность условия в первой строке равна нулю, условные вероятности в этой строке не определены. Мы примем соглашение о замене их в этом случае нулями.

⁷ Строку и столбец с тривиальными нулевыми элементами в матрице P_{Σ} мы не вычеркиваем, чтобы формулы на с. 277, 278 имели место во всех случаях. Возникающее в силу этого собственное число $\lambda=0$ и неопределенный собственный вектор не вносят осложнений, поскольку в силу эргодического характера цепи в формулах (IV.3.25)–(IV.3.33) деление на нуль не встречается.

В качестве второго шага из уравнения $P_{\Sigma}^T \pi_{st} = \pi_{st}$, т. е. из системы линейных уравнений (IV.3.35), найдем вектор стационарного распределения π_{st} :

$$\begin{array}{rcccccccc}
 0\pi_{11} + 0\pi_{12} & + 0\pi_{13} & + 0\pi_{21} + 0\pi_{22} & + 0\pi_{23} & + 0\pi_{31} & + 0\pi_{32} & + 0\pi_{33} & = \pi_{11}, \\
 0\pi_{11} + 0\pi_{12} & + 0\pi_{13} & + 0\pi_{21} + 0\pi_{22} & + 0\pi_{23} & + 0.2326\pi_{31} + 0\pi_{32} & + 0\pi_{33} & & = \pi_{12}, \\
 0\pi_{11} + 0\pi_{12} & + 0\pi_{13} & + 0\pi_{21} + 0\pi_{22} & + 0\pi_{23} & + 0.7674\pi_{31} + 0\pi_{32} & + 0\pi_{33} & & = \pi_{13}, \\
 0\pi_{11} + 0.1000\pi_{12} + 0\pi_{13} & & + 0\pi_{21} + 0\pi_{22} & + 0\pi_{23} & + 0\pi_{31} & + 0.0769\pi_{32} + 0\pi_{33} & & = \pi_{21}, \\
 0\pi_{11} + 0.1000\pi_{12} + 0\pi_{13} & & + 0\pi_{21} + 0.2857\pi_{22} + 0\pi_{23} & & + 0\pi_{31} & + 0.0939\pi_{32} + 0\pi_{33} & & = \pi_{22}, \\
 0\pi_{11} + 0.8000\pi_{12} + 0\pi_{13} & & + 0\pi_{21} + 0.7143\pi_{22} + 0\pi_{23} & & + 0\pi_{31} & + 0.8242\pi_{32} + 0\pi_{33} & & = \pi_{23}, \quad (IV.3.35) \\
 0\pi_{11} + 0\pi_{12} & + 0.0488\pi_{13} + 0\pi_{21} + 0\pi_{22} & & + 0.0360\pi_{23} + 0\pi_{31} & & + 0\pi_{32} & + 0.1396\pi_{33} & = \pi_{31}, \\
 0\pi_{11} + 0\pi_{12} & + 0.1220\pi_{13} + 0\pi_{21} + 0\pi_{22} & & + 0.2383\pi_{23} + 0\pi_{31} & & + 0\pi_{32} & + 0.2793\pi_{33} & = \pi_{32}, \\
 0\pi_{11} + 0\pi_{12} & + 0.8293\pi_{13} + 0\pi_{21} + 0\pi_{22} & & + 0.6452\pi_{23} + 0\pi_{31} & & + 0\pi_{32} & + 0.5811\pi_{33} & = \pi_{33}.
 \end{array}$$

Система (IV.3.35) имеет бесчисленное множество решений, но среди них имеется лишь один стохастический вектор

$$\pi_{st} = \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{13} \\ \pi_{21} \\ \pi_{22} \\ \pi_{23} \\ \pi_{31} \\ \pi_{32} \\ \pi_{33} \end{bmatrix} = \begin{bmatrix} 0.0000 \\ 0.0183 \\ 0.0760 \\ 0.0154 \\ 0.0271 \\ 0.1799 \\ 0.0789 \\ 0.1770 \\ 0.4275 \end{bmatrix}. \quad (IV.3.36)$$

Найдем теперь стационарное распределение исходов состояний по первой формуле (IV.3.24). Подсчеты дают

$$\begin{aligned}
 P_{Bt} &= 0.0000 + 0.0183 + 0.0760 = 0.0943, \\
 P_{PyT} &= 0.0154 + 0.0271 + 0.1799 = 0.2224, \\
 P_{P1} &= 0.0789 + 0.1770 + 0.4375 = 0.6834.
 \end{aligned}$$

Контроль по второй формуле (IV.3.24) дает

$$\begin{aligned}
 P_{Bt} &= 0.0000 + 0.0154 + 0.0789 = 0.0943, \\
 P_{PyT} &= 0.0183 + 0.0271 + 0.1770 = 0.2224, \\
 P_{P1} &= 0.0760 + 0.1799 + 0.4275 = 0.6834.
 \end{aligned}$$

Таким образом,

$$P_{st} = \begin{pmatrix} P_{Bt} \\ P_{PyT} \\ P_{P1} \end{pmatrix} = \begin{pmatrix} 0.0943 \\ 0.2224 \\ 0.6834 \end{pmatrix}. \quad (IV.3.37)$$

Найдем теперь матрицу $P_{I; J}$. Эта матрица в рассматриваемом случае стационарной цепи должна быть согласована с P_{Σ} . Вычисляя ее по формуле (IV.3.25), получим

$$P_{I; J} = \begin{matrix} & \begin{matrix} \overbrace{\text{Bt} & \text{Pyr} & \text{PI}}^h \\ \text{Bt} & 0.0000 & 0.1941 & 0.8059 \\ \text{Pyr} & 0.0692 & 0.1219 & 0.8089 \\ \text{PI} & 0.1155 & 0.2590 & 0.6255 \end{matrix} \end{matrix} \quad (\text{IV. 3. 38})$$

Выражения (IV. 3. 34), (IV. 3. 37) и (IV. 3. 38) дают всю информацию о цепи, поскольку распределение вероятностей исходов испытаний одинаково в любой момент (стационарное распределение (IV. 3. 36)). Так, в частности,

$$p^{(1)} = P_{I; J}^T p^{(0)} \quad (\text{IV. 3. 39})$$

Формулу (IV. 3. 39) используем для контроля вычислений:

$$p^{(1)} = \begin{pmatrix} 0.0000 & 0.0692 & 0.1155 \\ 0.1941 & 0.1219 & 0.2590 \\ 0.8059 & 0.8089 & 0.6255 \end{pmatrix} \begin{pmatrix} 0.0943 \\ 0.2224 \\ 0.6834 \end{pmatrix} = \begin{pmatrix} 0.0943 \\ 0.2224 \\ 0.6834 \end{pmatrix}.$$

Обратимся теперь ко второму случаю, когда стационарность не предполагается, но мы допускаем, что цепь обладает марковской однородностью. Рассмотрим этот случай на том же материале габбро-диабазов Гульшада. При этом очевидно, что P_{Σ} остается той же. В качестве $P_{I; J}$ мы возьмем эмпирическую матрицу частот, соответствующую $N_{I; J}$ (эта матрица получена с учетом того, что изучаемые на практике последовательности состоят из отдельных кусков, как это объяснялось ранее и будет подробно рассмотрено в гл. VI). Учитывая произвольность задания $p^{(0)}$, возьмем его, не обосновывая, равным

$$\begin{pmatrix} 0.1 \\ 0.3 \\ 0.6 \end{pmatrix}.$$

Этих данных, т. е. P_{Σ} , $P_{I; J}$ и $p^{(0)}$, достаточно, чтобы вычислить распределение вероятностей исходов в каждый момент времени. Подсчитаем сначала $p^{(1)}$ по формуле (IV. 3. 26), откуда

$$p^{(1)} = \begin{pmatrix} 0.0000 & 0.0750 & 0.1156 \\ 0.1887 & 0.1167 & 0.2554 \\ 0.8113 & 0.8083 & 0.6290 \end{pmatrix} \begin{pmatrix} 0.1 \\ 0.3 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 0.0919 \\ 0.2071 \\ 0.7010 \end{pmatrix}.$$

Вычислим далее $p^{(2)}$. Для этого сначала вычислим вектор $\pi^{(2)}$, компоненты которого разыскиваются по формуле (IV.3.27). Получим

$$\pi^{(2)} = \begin{bmatrix} \pi_{11}^{(2)} & 0.0000 \\ \pi_{12}^{(2)} & 0.0189 \\ \pi_{13}^{(2)} & 0.0811 \\ \pi_{21}^{(2)} & 0.0225 \\ \pi_{22}^{(2)} & 0.0350 \\ \pi_{23}^{(2)} & 0.2425 \\ \pi_{31}^{(2)} & 0.0694 \\ \pi_{32}^{(2)} & 0.1532 \\ \pi_{33}^{(2)} & 0.3774 \end{bmatrix}.$$

Затем по формуле (IV.3.28) рассчитаем

$$\pi^{(3)} = \begin{bmatrix} 0.0 & 0 & 0.0 & 0 & 0 & 0 & 0 \\ 0.0 & 0 & 0.0 & 0 & 0.2326 & 0 & 0 \\ 0.0 & 0 & 1.0 & 0 & 0.7674 & 0 & 0 \\ 0.0 & 0.1000 & 0 & 0 & 0 & 0.0769 & 0 \\ 0.0 & 0.1000 & 0 & 0.02857 & 0 & 0.0989 & 0 \\ 0.0 & 0.8000 & 0 & 0.07143 & 0 & 0.8242 & 0 \\ 0.0 & 0.0488 & 0.0 & 0.0860 & 0 & 0 & 0.1396 \\ 0.0 & 0.1220 & 0.0 & 0.2688 & 0 & 0 & 0.2793 \\ 0.0 & 0.8293 & 0.0 & 0.6452 & 0 & 0 & 0.5811 \end{bmatrix} \begin{bmatrix} 0.0000 \\ 0.0189 \\ 0.0811 \\ 0.0225 \\ 0.0350 \\ 0.2450 \\ 0.0694 \\ 0.1532 \\ 0.3774 \end{bmatrix} = \begin{bmatrix} 0.0000 \\ 0.0161 \\ 0.0758 \\ 0.0137 \\ 0.0270 \\ 0.1664 \\ 0.0755 \\ 0.1805 \\ 0.4430 \end{bmatrix}. \quad (\text{IV.3.40})$$

Наконец, найдем распределение исходов во второй момент времени с помощью суммирования по формуле (IV.3.29).

Так, например,

$$p_{\text{вг}}^{(2)} = \pi_{12}^{(3)} + \pi_{22}^{(3)} + \pi_{32}^{(3)} = 0.0161 + 0.0270 + 0.1805 = 0.2236.$$

В итоге

$$p^{(2)} = \begin{pmatrix} 0.0912 \\ 0.2236 \\ 0.6852 \end{pmatrix}.$$

Для того чтобы найти распределение после третьего момента времени, необходимо повторить все вычисления, произведенные при нахождении $p^{(2)}$. Для этого сначала находим $\pi^{(4)}$ по правой части формулы (IV.3.30), т. е. подставляем в левую часть (IV.3.40) вместо вектора $\pi^{(2)}$ вектор $\pi^{(3)}$. В итоге получаем

$$\pi^{(4)} = \begin{bmatrix} \pi_{11} & 0.0000 \\ \pi_{12} & 0.0180 \\ \pi_{13} & 0.0732 \\ \pi_{21} & 0.0155 \\ \pi_{22} & 0.0272 \\ \pi_{23} & 0.1809 \\ \pi_{31} & 0.0799 \\ \pi_{32} & 0.1777 \\ \pi_{33} & 0.4276 \end{bmatrix}.$$

Суммируя компоненты вектора $\pi^{(4)}$ по формулам (IV.3.31), находим

$$p^{(3)} = \begin{pmatrix} 0.0954 \\ 0.2229 \\ 0.6817 \end{pmatrix}.$$

Сравнивая $p^{(3)}$ с p_{sl} из (IV.3.37), убеждаемся, что после четвертого шага цепь практически устанавливается. Дальнейшие распределения $p^{(k)}$ ($k > 3$) можно считать совпадающими с p_{sl} .

IV.4. НЕКОТОРЫЕ ОБОБЩЕНИЯ МАРКОВСКОГО СВОЙСТВА

Здесь рассматриваются вероятностные соотношения, которые всегда выполняются в марковских цепях, но могут иметь место и в более широком классе случайных последовательностей. Такие соотношения связаны с понятиями восстанавливающего события, частного марковского перехода, ограниченно марковского перехода и марковского признака. Все эти понятия широко используются в геологических задачах.

IV.4.1. Восстанавливающие события

Рассмотрим бесконечную случайную последовательность и некоторое отнесенное к ней событие \mathcal{A} . Предположим, что событие \mathcal{A} зависит лишь от конечного числа испытаний и может повторяться неограниченно в бесконечной реализации. К событиям указанного типа относятся, например, появление подряд трех зерен кварца в последовательности зерен породообразующих минералов в граните. Такое событие принадлежит к типу \mathcal{A} , так как оно, очевидно, зависит от конечного (три) числа испытаний и может осуществляться бесконечное число раз в бесконечной реализации. В то же время событие, заключающееся в том, что более половины всех зерен в последовательности зерен образовано кварцем, не может быть отнесено к указанному типу \mathcal{A} , так как зависит от исхода бесконечного числа испытаний.

Разделим теперь бесконечную случайную последовательность на куски. При этом пусть каждый кусок начнется с испытания, следующего непосредственно за испытанием, завершающим осуществление очередного события \mathcal{A} . Заканчивается этот кусок тем испытанием, которым заканчивается осуществление следующего \mathcal{A} . В только что приведенном примере, где \mathcal{A} — три последовательных появления зерна кварца, реализации

$$\dots \text{Or, Ab, Q, Q, Q} | \text{Ab, Q, Q, Or,} \\ \text{Q Q Q} | \text{Ab Ab Ab Q Q Q} | \text{Q Ab Or Ab Or Ab} \dots \quad (\text{IV. 4. 1})$$

соответствовали бы куски, отделенные вертикальными прямыми.

Если условная вероятностная мера при условии, что осуществилось \mathcal{A} , определена на каждом куске одинаково и независимо от ее

задания на остальных кусках, то событие \mathcal{A} называется восстанавливающим для соответствующей случайной последовательности.

Моменты времени, отвечающие испытаниям, завершающим отдельные куски, называются моментами регенерации (восстановления). В приведенном примере моментами регенерации являются номера испытаний, отвечающие последнему появлению зерна Q в тройке QQQ .

Восстановление описывает стохастический механизм цикличности — каждый отдельный кусок является вероятностным аналогом цикла, а наступление очередного восстанавливающего события \mathcal{A} означает завершение очередного цикла. Нужно иметь в виду, что в случайной последовательности разные циклы могут реализовываться по-разному. Общим в них обязано быть лишь одно — каждый цикл завершается событием \mathcal{A} . Этим стохастические циклы отличаются от циклов, где реализация каждого куска является точной копией любого другого куска. В случайной последовательности каждый цикл начинается в одних и тех же начальных условиях (в примере (IV.4.1) он начинается после трех зерен кварца, если появление этих зерен было восстанавливающим событием).

Сведения о течении процесса в более ранних кусках, очевидно, ничего не дают для прогноза в данном куске, случайная последовательность каждый раз начинается как бы заново — после появления очередного \mathcal{A} .

Рассмотрим теперь бесконечную последовательность, состоящую из отдельных кусков. Пусть \mathcal{A} — восстанавливающее событие, встретившееся в каком угодно месте бесконечной последовательности. Обозначим через $\mathfrak{A}_{(-)}$ σ -алгебру событий, зависящих только от исходов испытаний, случившихся до \mathcal{A} . Пусть также $\mathfrak{A}_{(+)}$ означает σ -алгебру событий, зависящих лишь от исходов испытаний, случившихся после появления \mathcal{A} . Тогда очевидно, что в силу восстанавливающего свойства события \mathcal{A} имеет место равенство

$$p(C | \mathcal{A}; B) = p(C | \mathcal{A}), \quad C \in \mathfrak{A}_{(+)}, \quad B \in \mathfrak{A}_{(-)}. \quad (IV.4.2)$$

Формула (IV.4.2) представляет собой очевидную аналогию с марковским свойством (IV.2.6). В роли $\mathfrak{A}_{(+)}$ в (IV.2.6) фигурирует \mathfrak{B}_h , в роли \mathcal{A} — событие $\{a_L(h-r), a_R(h-r+1), \dots, a_J(h-2), a_I(h-1)\}$, наконец, $\mathfrak{A}_{(-)}$ соответствует \mathfrak{A}_{h-r-1} .

Представление о восстанавливающем событии дает пример IV.3. Действительно, если мы будем рассматривать последовательность АВАЭ₁, АВАЭ₁, АВАЭ₂, . . . , то событие АВА является восстанавливающим. Появление слоев А, затем В, а за ними А завершает построение пакета. При этом после завершения пакета возникает всегда одна и та же вероятностная ситуация — с постоянными вероятностями появляется либо К, либо тот или иной замещающий его элемент. С другой стороны, само по себе появление А не явля-

ется восстанавливающим событием, так как поведение случайной последовательности после появления одного А и после ВА различно. Легко заметить, что появление В также служит восстанавливающим событием в данном примере.

Явление восстановления привлекает к себе внимание в связи с концепцией цикличности осадконакопления. Эта концепция развивается на основе чисто детерминистического представления о механизме, вызывающем появление повторяемости в составе осадков при смене их во времени. При обращении к конкретным разрезам для сохранения детерминированного характера повторяемости прибегают к так называемой типизации состава слоев. Кроме того очень часто при весьма слабой документации явления допускается размыв тех или иных слоев, наличие которых необходимо для получения детерминированного цикла. Несмотря на все эти усилия, в седиментологии речь всегда идет лишь о закономерностях, т. е. о некоторых тенденциях, но отнюдь не о жестко выполняемых законах. Между тем эти широкие закономерности, установленные в седиментологии, представляют не что иное, как реализацию некоторых случайных процессов. При этом явление цикличности, по-видимому, во многих случаях может быть отражением процесса, проходящего с восстановлением. На важность восстанавливающего свойства в геологической литературе обращалось внимание неоднократно (Вихерт, 1970; Wickman, 1966; Schwarzscher, 1975).

Вопрос о том, что представляют собой так называемые ритмы и циклы осадконакопления (или чередования слоев), может рассматриваться с разных точек зрения. При этом прежде всего следовало бы проверить альтернативу о существовании детерминированных циклов, регулярность которых нарушена уничтожением отдельных слоев размывом, против гипотезы о том, что чередование слоев с самого начала представляло случайную последовательность с восстановлением. Есть основания полагать, что гипотеза и альтернатива приводят к разному строению возникающих случайных последовательностей. Такую работу следовало бы начать с изучения флишевых толщ.

Рассмотрим теперь понятие «восстанавливающее событие» в связи с представлением об однородной марковской цепи. Очевидно, что в простой однородной марковской цепи в качестве восстанавливающего события можно взять исход любого единичного испытания. Действительно, если наступило $a_1(h)$, то этим определяется все дальнейшее вероятностное течение процесса. После очередного появления $a_1(h+t)$ вероятностное течение процесса полностью совпадает с его течением после $a_1(h)$.

Справедливо и обратное: если в некоторой случайной последовательности наступление каждого состояния $I \in \{\mathcal{S}\}$ есть восстанавливающее событие, то эта случайная последовательность является простой марковской цепью. Аналогичное положение наблюдается в сложных цепях. В однородной марковской цепи

r -го порядка каждое r -звено (т. е. [появление любого сочетания состояний в r последовательных испытаниях] является восстанавливающим событием, и, наоборот, если в некоторой случайной последовательности каждое r -звено есть восстанавливающее событие, то эта случайная последовательность является марковской цепью r -го порядка.

Появление некоторого состояния в отдельном испытании может быть восстанавливающим событием и в немарковской последовательности. То же относится и к r -звеньям.

IV.4.2. Типы марковских переходов

Пусть имеется случайная стационарная последовательность. Рассмотрим в ней сочетание, состоящее из исходов r смежных последовательных испытаний, т. е. событие вида

$$\{a_I(h-1), a_J(h-2), \dots, a_L(h-r)\}. \quad (IV.4.3)$$

Такое событие будем называть *переходом r -го порядка*. Все вероятностные соотношения, связанные с (IV.4.3), не зависят от h в силу принятой стационарности. Мы используем индекс h только для того, чтобы пояснить очередность испытаний. Если состояния I, J, \dots, L выбраны в этом переходе конкретно, фиксированным образом, то мы говорим о частном переходе r -го порядка.

Частный переход r -го порядка называется марковским, если соотношение

$$\begin{aligned} p[B | a_I(h-1), a_J(h-2), \dots, a_L(h-r); A] = \\ = p[B | a_L(h-1), \dots, a_L(h-r)] \end{aligned} \quad (IV.4.4)$$

выполняется при фиксированных I, J, \dots, L для всех событий $B \in \mathfrak{B}_h$ и всех $A \in \mathfrak{A}_{h-r-1}$ ($h \geq r$).

Частный переход r -го порядка называется бернуллиевским, если

$$p[B | a_L(h-1), a_J(h-2), \dots, a_L(h-r); A] = p(B), \quad (IV.4.5)$$

где $B \in \mathfrak{B}_h$ и $A \in \mathfrak{A}_{h-r-1}$.

Понятие частного марковского (бернуллиевского) перехода полезно в двух отношениях. Прежде всего оно подчеркивает то обстоятельство, что марковские переходы могут иметь место и в существенно немарковских последовательностях. Кроме того, с помощью этого понятия можно произвести более детальную классификацию самих марковских цепей. Так, в марковской цепи r -го порядка каждый частный переход порядка $l \geq r$ является марковским по определению. Кроме того в этой цепи могут существовать марковские (бернуллиевские) переходы любых меньших порядков. Обратное утверждение справедливо в следующей формулировке: *если в случайной стационарной последовательности все частные переходы r -го порядка марковские, то эта последовательность есть марковская цепь r -го порядка.*

Рассмотрим некоторые ситуации.

Пусть имеется существенно немарковская последовательность, в которой появление состояния I является восстанавливающим событием. Тогда, согласно определению восстанавливающего события, переход первого порядка через I — марковский. В то же время в этой последовательности всегда найдется такое состояние J , что переход через него не будет марковским.

Рассмотрим далее исходную последовательность зерен Og , Q и Ab в граните, являющуюся простой марковской цепью. Предположим, что между каждой парой смежных неплаггиоклазовых зерен может возникнуть одно вторичное зерно плаггиоклаза с постоянной вероятностью π . Если реализуется такой процесс, то мы говорим, что произошел метасоматоз по Ab .

Как будет показано в следующей главе, в результате реализации рассматриваемой схемы возникает марковская цепь второго порядка. При этом в ней имеют место марковские переходы второго порядка через Ab ($h-1$) Ab ($h-2$); Ab ($h-1$) Q ($h-2$) и Ab ($h-1$) Og ($h-2$). В то же время переходы через Q и Og сохраняют такую же структуру, что и в исходной последовательности; иными словами, эти переходы первого порядка являются марковскими или бернуллиевскими.

Указанная стохастическая схема подтверждена большим числом наблюдений (Романова, 1977). Как правило, в цепях второго порядка, на состояниях Og , Q , Ab , отвечающих последовательностям зерен в метасоматически измененных гранитах, мы находим один или два перехода первого порядка, являющихся марковскими. Если бы образовывались вторичные зерна Q , а не Ab , то переход первого порядка через Q не был бы марковским. Таким путем мы можем выяснить многие детали процесса, метасоматически перерабатывающего гранит.

Пример IV.13. В результате исследования разреза ЧЗ красноцветной толщи Челекена (Вистелиус, Романова, 1962), состоящего из чередования песчаных (π) и глинистых (γ) слоев, которое, как можно предполагать, является стационарной случайной последовательностью, возникло предположение, что частные переходы второго порядка через π ($h-2$) π ($h-1$) и γ ($h-2$) π ($h-1$) являются марковскими. В то же время в той же последовательности имеется переход первого порядка через γ ($h-1$). Проверим эту гипотезу, вычислив теоретическое распределение по длине π -серий и сравнив его с наблюдаемым распределением. Согласно (IV.2.29), для π -серии длины $l \geq 2$ вероятность p_l находится как

$$p_l [\underbrace{a_\gamma (h+l) a_\pi (h+l-1), \dots, a_\pi (h+1)}_{l-1} / a_\pi (h) a_\gamma (h-1)]. \quad (IV. 4. 6)$$

Опираясь на предполагаемые свойства частных марковских переходов, рассчитаем

$$\begin{aligned}
 p_l &= \frac{p[a_\gamma(h-1), a_\pi(h), a_\pi(h+1), \dots, a_\pi(h+l-1), a_\gamma(h+l)]}{p[a_\gamma(h-1), a_\pi(h)]} = \\
 &= \frac{p[a_\gamma(h-1)] p[a_\pi(h)/a_\gamma(h-1)] p[a_\pi(h+1)/a_\pi(h), a_\gamma(h-1)]}{p[a_\gamma(h-1)] p[a_\pi(h)/a_\gamma(h-1)]} \times \\
 &\quad \times \{p[a_\pi(h+2)/a_\pi(h+1), a_\pi(h)]\}^{l-2} \times \\
 &\quad \times p[a_\gamma(h+l)/a_\pi(h+l-1), a_\pi(h+l-2)]. \quad (IV.4.7)
 \end{aligned}$$

Используя стационарность и произведя сокращения в (IV.4.7), получим

$$p_l = (p_{\pi, \pi; \pi})^{l-2} p_{\pi, \pi; \gamma} p_{\gamma, \pi; \pi} \quad (l \geq 2). \quad (IV.4.8)$$

Для серий длины 1 будем иметь

$$p_1 = p_{\gamma, \pi; \gamma}. \quad (IV.4.9)$$

Распределение серий по длине (IV.4.8) и (IV.4.9) определяется двумя независимыми параметрами — $p_{\pi, \pi; \pi}$ и $p_{\gamma, \pi; \pi}$. При сравнении с помощью статистики χ^2 теоретического распределения с эмпирическим, сведенным в n разрядов, имеем $n-3$ степени свободы. В качестве оценки параметров используем соответствующие эмпирические переходные вероятности, полученные в одной длинной реализации: $p_{\pi, \pi; \pi} = 0.2647$ и $p_{\gamma, \pi; \pi} = 0.1276$. Расчеты приведены в табл. IV.3.

Т а б л и ц а IV.3

Наблюдаемые и вычисленные частоты для серий песчаных слоев в разрезе ЧЗ (Челекен)

Длина серии из песчаных слоев	Наблюдаемое число серии, m	Вычисленное число серий, n	$\frac{(m-n)^2}{n}$
1	171	171.0	0.00
2	19	18.4	0.03
3	4	4.9	0.17
4	1	1.3	0.07
5	1	0.40	0.90
Σ	196	196	1.17

$$0.50 < p[(\chi_{(2)}^2 \geq 1.17)] < 0.75.$$

Итак, гипотеза о том, что частные переходы через π второго порядка, а через γ первого порядка являются марковскими, не опровергается. Она показывает, что изученная последовательность слоев имеет как бы промежуточную по сложности структуру между марковскими цепями первого и второго порядка. Предположение, что это цепь первого порядка, — слишком грубо, а предположение, что она отражает процесс, обладающий всеми чертами цепи второго порядка, — излишне сложно.

Введем теперь понятие *ограниченно марковского перехода*. Как только что указывалось, для того чтобы случайная стационарная последовательность была марковской цепью r -го порядка, необходимо и достаточно, чтобы всякий частный переход r -го порядка был марковским. Иными словами, критерий заключается в том, чтобы равенство

$$\begin{aligned} & p [B | a_I(h-1), \dots, a_L(h-r) A] = \\ & = p [B | a_I(h-1), \dots, a_L(h-r)] \quad (B \in \mathfrak{B}_h, A \in \mathfrak{A}_{h-r-1}) \end{aligned}$$

выполнялось для любого набора I, J, \dots, L . Этот критерий марковости можно заменить другим необходимым и достаточным условием, заключающимся в следующем: *для того чтобы случайная стационарная последовательность была марковской цепью r -го порядка, необходимо и достаточно, чтобы для любого частного перехода r -го порядка выполнялось равенство*

$$\begin{aligned} & p [a_K(h) | a_I(h-1), \dots, a_L(h-r) A] = \\ & = p [a_K(h) | a_I(h-1), \dots, a_L(h-1)] \end{aligned} \quad (\text{IV. 4. 10})$$

при всех $K \in \mathcal{S}$ и произвольном событии $A \in \mathfrak{A}_{h-r-1}$ из прошлого. Произвольное событие из будущего $B \in \mathfrak{B}_h$ мы здесь заменили на простейший частный вид подобного события $a_K(h)$, т. е. на будущее, ограниченное исходом следующего испытания. Если некоторый частный переход через фиксированный набор I, \dots, L

обладает свойством (IV.4.10), то мы говорим, что этот частный переход r -го порядка является *ограниченно марковским*. Подчеркнем еще раз, что наличие ограниченной марковости у любого частного перехода r -го порядка достаточно и необходимо для того, чтобы стационарная случайная последовательность была марковской цепью r -го порядка. Очевидно, что всякий частный марковский переход является в то же время ограничено марковским. В то же время ограничено марковский переход не всегда является частным марковским. Таким образом, класс случайных последовательностей, в которых имеются ограничено марковские переходы, шире класса последовательностей с частными марковскими переходами.

Ограниченно марковские переходы встречаются достаточно часто в геологических явлениях, что будет показано ниже.

Пример IV.14. Допустим, что имеются два пакета из слоев

$$A_1 = \{\pi(h-1), \gamma(h)\} \text{ и } A_2 = \{\pi(h-1), \pi(h), \gamma(h+1)\},$$

где π — песчанистый, а γ — глинистый слой.

Предположим далее, что случайная последовательность со значениями из $\{A_1, A_2\}$ представляет стационарную и простую марковскую цепь с матрицей переходных вероятностей

$$P = \begin{matrix} & A_1 & A_2 \\ \begin{matrix} A_1 \\ A_2 \end{matrix} & \begin{pmatrix} 0 & 1.0 \\ 0.9 & 0.1 \end{pmatrix} \end{matrix}.$$

Предположим далее, как это часто бывает во время полевых работ, что положение границы между пакетами нельзя определить. Таким образом, фактически мы наблюдаем последовательность на $\{\pi, \gamma\}$. Рассмотрим в этой последовательности переход через γ на ближайшее следующее состояние. Очевидно, что

$$p[a_\gamma(h) | a_\gamma(h-1) A] = 0, \quad p[a_\pi(h) | a_\gamma(h-1) A] = 1, \quad A \in \mathfrak{A}_{h-2}.$$

Таким образом, согласно определению, переход через γ является ограниченно марковским. Покажем, что этот переход не будет частным марковским. Для этого рассмотрим вероятности

$$\begin{aligned} p[a_\pi(h) | a_\gamma(h-1), a_\pi(h-2), a_\gamma(h-3)] &= 1, \\ p[a_\pi(h) | a_\gamma(h-1), a_\pi(h-2), a_\pi(h-3)] &= 1, \\ p[a_\gamma(h+1), a_\pi(h) | a_\gamma(h-1), a_\pi(h-2), a_\gamma(h-3)] &= 0, \\ p[a_\gamma(h+1), a_\pi(h) | a_\gamma(h-1), a_\pi(h-2), a_\pi(h-3)] &= 0.9. \end{aligned}$$

Вероятности указанных событий легко подсчитать, помня строение пакетов A_1 и A_2 и учитывая переходные вероятности в матрице P . Так, например, допустим, что реализовалось условие $a_\gamma(h-3)$, $a_\pi(h-2)$ $a_\gamma(h-1)$. Очевидно, что слой $a_\pi(h-2)$ может быть лишь из пакета A_1 , а поэтому слой $a_\gamma(h-1)$ должен быть из того же пакета A_1 , и переход на $a_\pi(h)$ невозможен.

Как видим, если рассматривать событие в будущем, зависящее от исходов двух последовательных испытаний, то прошлое уже оказывает влияние на вероятности таких событий. Если прошлое представлено сочетанием $a_\gamma(h-3)$, $a_\pi(h-2)$ при настоящем $a_\gamma(h-1)$, то вероятность события $a_\pi(h)$, $a_\gamma(h+1)$ равна нулю. Если же взять прошлое, составленное из $a_\pi(h-3)$, $a_\pi(h-2)$ при настоящем $a_\gamma(h-1)$, то вероятность того же события в будущем будет равна 0.9. Таким образом, переход через γ ограниченно марковский, но не частный марковский.

Аналогичные ситуации могут возникать при изучении метасоматических явлений. Пусть, например, имеется первично магматический гранит, последовательность зерен Og , Q и Ab в котором является цепью Маркова первого порядка. В результате вторичной переработки породы в ней развиваются вторичные зерна кварца. Эти зерна прирастают только к первичным зернам кварца. Механизм этого явления с точки зрения описания окончательной последовательности эквивалентен тому, что между каждой парой смежных первичных зерен QQ , $Q Ab$, $Ab Q$, $Q Og$, $Og Q$ может добавляться серия из вторичных зерен Q . Вероятность $p(l)$ того, что появится серия ровно из l вторичных зерен, может зависеть только от составов смежных первичных зерен и от длины серии.

Иными словами, предполагается, что вероятность $p_{I;J}(l)$ не меняется от того, что происходит в других смежных парах.

Рассмотрим окончательную последовательность. Переходы через Ab и Og в ней будут ограниченно марковскими. Что касается перехода через Q , то его вероятностные свойства зависят от распределения вероятностей $p_{I;J}(l)$. Если предположить, что серии более некоторой предельной длины возникать не могут, то переход через Q будет частным марковским высокого порядка. При неограниченных сериях может возникнуть существенно немарковский переход через Q . В следующей главе этот вопрос будет рассмотрен подробно.

Введем теперь понятие *марковского признака*. Подобно тому как мы ограничили будущее исходом лишь одного непосредственно следующего за настоящим испытанием, точно так же прошлое можно ограничить исходом лишь одного непосредственно предшествовавшего настоящему испытанию. Если оба таких ограничения мы сделаем одновременно, то получим следующее понятие.

Если частный переход r -го порядка $a_I(h-1)$, $a_J(h-2)$, ..., $a_L(h-r)$ обладает свойством

$$p[a_K(h) | a_I(h-1), a_J(h-2), \dots, a_L(h-r), a_R(h-r-1)] = \\ = p[a_K(h) | a_I(h-1), a_J(h-2), \dots, a_L(h-r)] \quad (\text{IV. 4. 11})$$

для фиксированного набора I, J, \dots, L при всех $R, K \in \mathcal{S}$, ($h \geq r+1$), то мы говорим, что частный переход r -го порядка обладает марковским признаком. Поскольку наличие частного марковского признака связано с конечным числом испытаний, в отличие от частного марковского или ограниченно марковского перехода, то наличие марковского признака может быть проверено по наблюдениям. Следует, однако, отметить, что марковский признак не дает критерия марковости. Даже в том случае, когда любой частный переход r -го порядка обладает марковским признаком, это еще не означает, что последовательность является цепью r -го порядка. Марковский признак является необходимым, но не достаточным условием марковости. Поэтому марковский признак следует использовать лишь для браковки гипотез о порядке марковости.

Пусть заранее известно, что случайная последовательность является марковской цепью некоторого неизвестного порядка, предположим также, что относительно порядка этой цепи может быть выдвинуто ограниченное количество гипотез (скажем, r может быть равно 0 или 1, или 5, но не может быть другим числом). В этом случае проверка марковского признака по наблюдениям позволяет установить порядок марковости \hat{r} . Если возможные значения r исчерпываются величинами 0, 1 и 5, как мы только что предположили, то для установления порядка r мы сначала проверяем по наблюдениям равенство

$$p[a_I(h) | a_J(h-1)] = p(a_J) \quad (\text{IV. 4. 12})$$

при всех $I, J \in \mathcal{S}$, т. е. марковский признак нулевого порядка.

Если (IV.4.12) не отвергается, то мы принимаем $r=0$, если (IV.4.12) отвергается, то мы испытываем по наблюдениям равенство

$$p[a_I(h) | a_J(h-1), a_K(h-2)] = p[a_I(h) | a_J(h-1)], \quad (\text{IV.4.13})$$

т. е. проверяем наличие марковского признака первого порядка.

Если наблюдения не отвергают (IV.4.13), то принимается $r=1$ (следует помнить, что при $r=1$ выполняется также $r=5$). Наконец, если наблюдения отвергают $r=1$, то мы принимаем $r=5$. Однако если у нас нет предварительной информации о том, что последовательность представляет марковскую цепь, то такой подход к выбору порядка марковости не приемлем. Кроме того, фактически мы можем не знать заранее, какому конечному набору принадлежит r . В этом случае использование марковского признака также не обеспечивает определения правильного r .

В реальной работе у нас часто есть основания предполагать, что последовательность является марковской цепью. Кроме того, более уверенно можно ожидать, что r не превосходит, скажем, трех. В этой ситуации марковский признак практически является достаточным для установления истинного порядка цепи.

Перекодировка состояний в марковской цепи, производимая так, что некоторым ранее различавшимся состояниям приписывается один и тот же кодовый номер, часто приводит к потере последовательностью марковского свойства. Для того чтобы оно сохранилось, необходимо выполнение весьма ограничительных условий. В то же время для сохранения при перекодировке марковского признака подобные условия могут быть значительно ослаблены. Здесь понятие марковского признака приобретает самостоятельный смысл (см. гл. V).

IV.5. ТРЕХМЕРНЫЕ УПАКОВКИ И МАРКОВСКИЕ ПОСЛЕДОВАТЕЛЬНОСТИ

Среди геологических объектов имеются такие, для которых понятие случайной, в частности марковской, последовательности является естественным. К ним относятся, скажем, последовательности слоев в осадочных толщах, последовательности зон разного состава у плагноклазов или турмалинов, последовательности расстояний между днищами в колонии фавозитов и т. п. Однако имеются такие области, с которыми понятие марковской последовательности связать труднее. Это существенно трехмерные области, состоящие из набора небольшого числа структурных единиц. К этим областям относятся горные породы. Однако описание соотношений между минеральными индивидуумами в горной породе (гранитах, песчаниках) в марковских терминах привело к высшей степени важным и нетривиальным результатам при решении геологических задач. Отсюда появилась необходимость со-

гласования структуры трехмерной упаковки со структурой характеризующих ее марковских последовательностей. Такие последовательности по предположению возникают на прямых, произвольно проходящих через трехмерную упаковку. Разработка этого вопроса представляет очень большой интерес для петрографии. Ниже приводится пример, показывающий, что трехмерная упаковка с простыми марковскими последовательностями на каждой проходящей через нее прямой возможна, причем у всех таких марковских цепей параметры одинаковы.

Пример IV.15. Пусть имеется плотнейшая упаковка из шаров одного размера, как это показано для плоского слоя на рис. IV.8 (Vistelius, 1976).

Пусть при этом каждый шар окрашен внутри только в один цвет I ($I=1, 2, \dots, k$). Цвет I для данного шара выбирается с по-

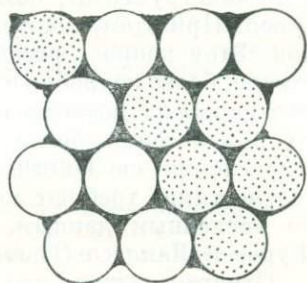


Рис. IV.8. Плоская плотнейшая упаковка шаров двух типов — светлых и покрытых точками.

Чередование типов шаров — случайное (определено по таблице случайных чисел). Последовательность отрезков на прямой, проходящей через такую упаковку, была бы представлена отрезками, покрытыми точками, светлыми и залитыми черным. Она представляла бы простую марковскую цепь.

стоянной вероятностью независимо от цветов всех уже выбранных шаров (с помощью таблицы случайных чисел). Вероятность исхода каждого испытания постоянна и равна p_I для I -того цвета. Поверхности шаров покрыты $k+1$ -м цветом. Этим же цветом закрашены промежутки между шарами, как это показано на рис. IV.8. Смежные шары всегда соприкасаются друг с другом, но только через слой краски. Проведем через охарактеризованную трехмерную упаковку случайную прямую. Зафиксируем на этой прямой чередование $k+1$ -го цвета (k цветов отвечает цвету шаров и один цвет ($k+1$ -й) — окраске поверхности шара и промежутков между ними). Таким образом, на любой прямой, проходящей через упаковку, мы получим случайную последовательность цветов на $k+1$ состоянии. Убедимся, что такая последовательность представляет простую марковскую цепь. Действительно, выбор цвета для настоящего испытания влияет на появление цвета в испытании со следующим номером. Так, если мы находимся на I -том цвете, то $p_I; k+1=1$. В то же время $p_I; J=0$, $I, J \in \{1, 2, \dots, k\}$. Исходы более ранних испытаний не меняют переходных вероятностей. Вдоль каждой случайной прямой, пересекающей изучаемую упаковку, переходные вероятности на один шаг одни и те же. Этими вероятностями — в силу стационарности и простой марко-

ности — полностью определяются соответствующие случайные последовательности. Таким образом, вдоль каждой прямой мы наблюдаем одну и ту же простую марковскую цепь.

Более специфический случай рассмотрен в работе Вистелиуса и Харламова (1979).

Итак, мы сравнительно детально изучили марковские цепи и аппарат для работы с ними. В то же время понятие о восстанавливаемом событии дано нами без соответствующего математического аппарата. Этот аппарат может быть найден в доступном изложении в главах 13 и 11 книг Феллера (1964, 1967).

Марковские цепи с частными марковскими переходами различных порядков не требуют привлечения какого-либо специального «инструмента», помимо используемого в теории марковских цепей. При этом, конечно, следует учитывать специфику переходов. Этот вопрос, рассмотренный в настоящей главе лишь частично, более подробно изложен в гл. VI. Что касается существенно немарковских последовательностей, то их изучение основывается на привлечении более сложных математических конструкций. Вычисления, связанные с существенно немарковскими последовательностями, требуют мощной вычислительной базы: приходится по численным данным повторно осуществлять преобразования Фурье и Лапласа (Harris, 1955).

Отметим также, что для немарковских последовательностей с ограниченно марковскими переходами литература отсутствует, так как они, по-видимому, появились только в связи с анализом геологических объектов.

В заключение следует отметить, что в примерах были рассмотрены вероятностные схемы, касающиеся структуры некоторых силикатов, и одна из схем, связанная с формированием нефтяных залежей. Хотелось бы, чтобы на вопросы такого типа обратили внимание соответствующие специалисты.

Л и т е р а т у р а

- В и с т е л и у с А. Б. К вопросу о механизме связи при слоеобразовании. — Докл. АН СССР, 1949, т. 65, № 7, с. 535—538.
- В и с т е л и у с А. Б. О кристаллизации аляскитов с р. Каракульджур (Центральный Тянь-Шань). — Докл. АН СССР, 1967, т. 172, № 1, с. 165—168.
- В и с т е л и у с А. Б., Р о м а н о в а М. А. Красноцветные отложения полуострова Челекен. М.—Л., Изд-во АН СССР, 1962. 227 с.
- В и с т е л и у с А. Б., Р о м а н о в а М. А. Концепция идеальных гранитов и ее использование при съемочных, петрографических и поисковых работах. — В кн.: Идеальные граниты, вып. 1. Л., «Наука», 1972, с. 4—47.
- В и с т е л и у с А. Б., Р о м а н о в а М. А. О вырожденном случае модели кристаллизации идеальных гранитов. — Докл. АН СССР, 1976, т. 228, № 1, с. 170—173.

- Вистелиус А. Б., Харламов Б. П. О трехмерной упаковке с марковскими линейными сечениями. — Докл. АН СССР, 1979, т. 245, № 2, с. 431—434.
- Вихерт А. В. Слоеобразование как пуассоновский случайный процесс с переменным во времени параметром. — В кн.: Математические методы в геологии. Сб. НСО, вып. 7. М., изд-во МГУ, 1970, с. 135—140.
- Гантмахер Ф. Р. Теория матриц. 2-е изд., М., «Наука», 1967. 575 с.
- Демина М. Е. О механизме фиксации зерен при формировании песчаных толщ. — В кн.: Исследования по математической геологии. Л., «Наука», 1978, с. 62—69.
- Дир У. А., Хаун Р. А., Зусман Дж. Породообразующие минералы. Т. I. М., «Мир», 1965. 310 с.
- Дир У. А., Хаун Р. А., Зусман Дж. Породообразующие минералы. Т. III. М., «Мир», 1966. 317 с.
- Иванов Д. Н. О природных аналогах идеальных гранитов (на примере массива Кызыл-Тас в Центральном Казахстане). — В кн.: Исследования по математической геологии. Л., «Наука», 1978, с. 9—84.
- Кемени Дж., Снелл Дж. Конечные цепи Маркова. М., «Наука», 1970. 271 с.
- Кринов Е. Л. Метеориты. М.—Л., Изд-во АН СССР, 1948. 335 с.
- Романова М. А. Белые граниты Арга-Ынных-Хая (Якутия) и механизм их образования. — В кн.: Исследования по математической геологии. Л., «Наука», 1978, с. 25—39.
- Статистическая идентификация идеальных гранитов и продуктов их преобразования. — В кн.: Идеальные граниты, вып. 2. Л., «Наука», 1972, с. 3—47. Авторы: А. Б. Вистелиус, Г. С. Лельчук, Г. А. Талмуд, А. В. Фаас.
- Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры. М., Изд-во физ.-мат. лит-ры, 1960. 656 с.
- Феллер В. Введение в теорию вероятностей и ее приложения. Т. I. 2-е изд. М., «Мир», 1964. 498 с.
- Феллер В. Введение в теорию вероятностей и ее приложения. Т. II. М., «Мир», 1967. 752 с.
- Boyd F. R., Smith D. Compositional Zoning in Pyroxenes from Lunar Rock 12021. Oceanus Procellarum. — J. Petrol., 1971, v. 12, No. 3, p. 439—464.
- Harris T. E. On chains of infinite order. — Pac. J. Math., 1955, suppl. 1, v. 5, p. 707—725.
- Kuennen Ph. H. Experimental abrasion of pebbles. I. Wet sandblasting. — In: Leidse Geol. Mede, 1956, v. 20, p. 131—137.
- Schwarzacher W. Sedimentation models and quantitative stratigraphy. Amsterdam, Elsevier, 1975. 382 p.
- Vistelius A. B. Mathematical Geology and the progress of geological sciences. — J. Geol., 1976, v. 84, No. 6, p. 629—653.
- Wickman F. E. Repose period patterns of volcanoes. V. General Discussion and tentative stochastic model. — Arkiv Miner. Geol., 1966, Bd 4, Hft 5, No. 11, S. 351—367.

Г л а в а V

ПРЕОБРАЗОВАНИЯ МАРКОВСКИХ ЦЕПЕЙ

Модели геологических процессов, приводящие к простым цепям Маркова. Теория укрупнения состояний. Какого типа случайной последовательностью окажется простая цепь Маркова, после того как она будет преобразована по той или иной схеме (модели вторичного процесса). Различные типы схем преобразования. Примеры из петрографии и седиментологии. Математические задачи, требующие разрешения.

Ключевые слова:

сильное укрупнение, слабое укрупнение, сгущение, разрежение, замещение, пакеты, слияние серий, удлинение серий.

V.1. ВВЕДЕНИЕ

Технический аппарат, рассмотренный в гл. IV, позволяет представить результаты многих геологических явлений в виде цепей Маркова. Это дает возможность на реалистической основе проверять гипотезы о ходе явлений. Однако наиболее удобные для работы простые цепи реализуются относительно редко. Обычно они отвечают весьма специфическим обстановкам. Гораздо чаще геолог имеет результаты, охарактеризованные разрезом или профилем, в котором на основной процесс наложены искажения. Если эти искажения реализовались по той или иной устойчивой схеме и систематически деформировали результаты первичного процесса, то во многих случаях можно выяснить схему, по которой шел как сам процесс, так и наложенные на него явления. Зная, что процесс в чистом виде порождал простую цепь Маркова, и построив схему искажения, можно предсказать, какого типа будет наблюдаемая последовательность. Практически такой метод работы оказывается очень продуктивным. Поэтому необходимо знать, как происходят преобразования простых цепей Маркова под влиянием различных процессов. Этому вопросу посвящена настоящая глава.

Прежде чем перейти к изложению, рассмотрим некоторые случаи, реально имевшие место в геологической работе и потребовавшие изучения вопроса о преобразовании цепи.

Схема образования последовательности слоев, в которой состояниями являются составы слоев, в некоторых случаях порождает простую марковскую цепь. Однако если наблюдается чередование глинистых и песчаных слоев, то это значит, что иногда мог происходить полный размыв глинистого слоя перед отложением песчаного. Это может привести к преобразованию простой цепи Маркова в последовательность, резко отличающуюся от нее.

Многие горные породы естественно рассматривать как результат кристаллизации тройной системы с компонентами a, b и c , дающими тройную эвтектику. Если ввести стадии кристаллизации, выделив индексом 1 стадию докотектическую, индексом 2 — котектическую и индексом 3 — эвтектическую, то последовательность зерен при некоторых условиях окажется простой цепью Маркова на шести состояниях — скажем, a_1, a_2, a_3, b_2, b_3 и c_3 . Однако петрограф при изучении породы под микроскопом не может для каждого зерна определить стадию, на которой оно образовалось. Он фиксирует только состав зерен. Таким образом, простая цепь Маркова на шести состояниях преобразуется в случайную последовательность на трех состояниях (a, b и c). Каковы свойства этой последовательности, можно сказать только после изучения теории преобразования цепей.

Аналогичных примеров можно подобрать множество, из любой области геологических наук. По мере разбора теории преобразований мы будем рассматривать также и наиболее характерные примеры.

В качестве исходной случайной последовательности в этой главе будет рассматриваться, если не сделано специальной оговорки, стационарная и эргодическая простая марковская цепь. Далее задается некоторое правило, по которому преобразуется каждая реализация этой цепи. В результате мы получаем новую случайную последовательность, свойства которой изучаем. При этом рассматриваются следующие преобразования: укрупнение по множеству состояний, укрупнение по множеству моментов времени, сгущение и разрежение цепи, замещение состояний и, наконец, стирание границ в последовательности пакетов.

Допустим, что множество состояний Σ исходной цепи разбито на непересекающиеся подмножества, т. е.

$$\Sigma = \Sigma_I \cup \Sigma_J \cup \dots \cup \Sigma_L \cup \Sigma_K,$$

причем

$$\Sigma_I \cap \Sigma_J = \emptyset, \text{ если } I \neq J.$$

Затем производится перекодировка реализаций исходной цепи так, что любому состоянию из Σ_I приписывается один и тот же символ I, \dots , любому состоянию из Σ_K — символ K . Моменты времени, отвечающие состояниям исходной последовательности, в перекодированной последовательности сохраняются. В резуль-

тате получается новая случайная последовательность на множестве состояний $\mathcal{S} = \{I, J, \dots, L, K\}$.

Полученная указанным способом последовательность построена из исходной цепи с помощью операции, называемой укрупнением по множеству состояний.

Другое возможное преобразование заключается в перекодировке только номеров исходов испытаний в реализациях цепи. Эта перекодировка производится так, что некоторым ранее различавшимся номерам испытаний приписывается один и тот же номер. Возникающая при этом последовательность получается с помощью операции, называемой укрупнением по времени.

Предположим далее, что в любой реализации исходной цепи между некоторыми парами исходов смежных испытаний могут размещаться исходы новых испытаний (отнесенные к тому же множеству состояний Σ или к какому-либо другому множеству состояний). Если при таком преобразовании взаимное упорядочение всех прежних исходов испытаний не меняется, то мы говорим, что происходит сгущение исходной цепи.

Преобразование, заключающееся в вычеркивании из исходных реализаций некоторых исходов испытаний с последующей перенумерацией оставшихся исходов в естественном порядке, мы будем называть разрежением цепи. Замену некоторых состояний одной случайной последовательности синхронными (с тем же номером) состояниями другой мы называем замещениями.

Наконец, мы будем рассматривать случайное чередование детерминированно построенных сочетаний исходов испытаний. Такие сочетания мы будем называть пакетами. Последовательность пакетов трансформируется следующим образом: границы между пакетами стираются, а образующие эти пакеты состояния нумеруются в естественном порядке. Такое преобразование мы называем стиранием границ в последовательности пакетов.

Указанные преобразования, а также их суперпозиции могут быть широко использованы во многих задачах математической геологии. Во второй части этой книги мы будем использовать теорию этих преобразований, изучая процессы кристаллизации магматических пород, метасоматическую переработку этих пород и явление стратификации осадочных образований.

Пример V.1. Рассмотрим кристаллизацию системы из трех компонентов с одной тройной эвтектикой, о которой мы упоминали только что (рис. V.4). Процесс кристаллизации начинается в точке a . На участке $a-b$ выделяются зерна состава Ab_1 , где 1 отмечает стадию кристаллизации. На участке $b-c$ образуются зерна двух компонентов — продолжается выделение зерен, сложенных Ab , и появляются зерна Q . Соответствующие состояния будем обозначать как Ab_2 и Q_2 . Наконец, в точке c (тройной эвтектики) выделяются зерна, которым отвечают состояния Ab_3 , Q_3 и Or_3 . Если мы зафиксируем зерна, попадающие на прямую, проходящую

через продукты кристаллизации указанной системы, то получим случайную последовательность. При принятой кодировке типов зерен с помощью двух индексов (состав и стадия кристаллизации) возникает последовательность на множестве состояний $\Sigma = \{I_1, I_2, I_3, J_2, J_3, K_3\}$. Как будет показано во второй части книги, рассмотренный процесс кристаллизации приводит к возникновению на прямых, проходящих через породу, стационарных эргодических простых цепей Маркова на множестве состояний Σ . При лабораторном исследовании мы, как правило, не можем выде-

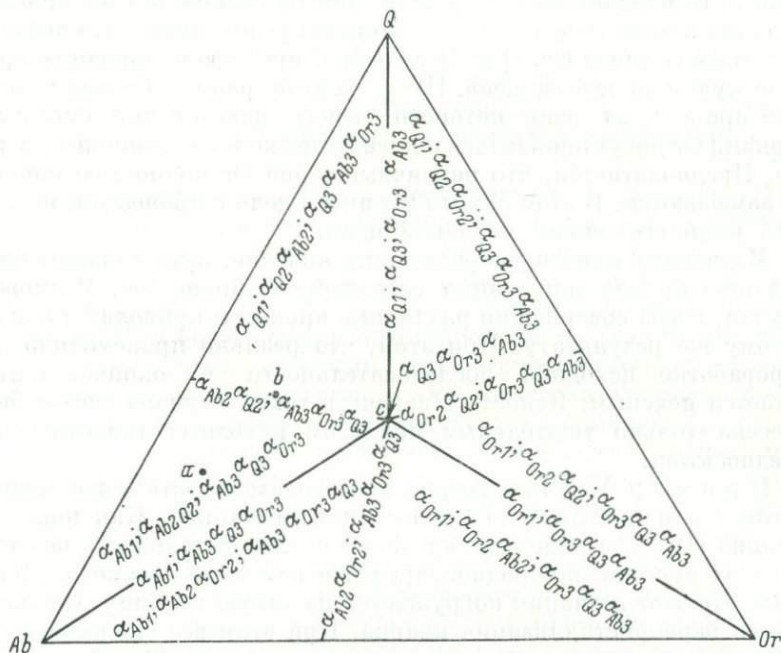


Рис. V.1. Схема кристаллизации эвтектической системы с состояниями Ab, Oг и Q.

α — реализация состояния, первый нижний индекс — состояние, второй — последовательность выделения (меньше цифра — более раннее выделение). a — точка, в которой началась кристаллизация, b — начало котектической и c — эвтектической кристаллизации.

лять зерна по их принадлежности к той или иной стадии кристаллизации (второму индексу) и вынуждены фиксировать состояния

$$I = I_1 \cup I_2 \cup I_3; \quad J = J_2 \cup J_3; \quad K = K_3.$$

Таким образом, наблюдается последовательность на множестве состояний:

$$\mathcal{S} = \{I, J, K\}.$$

Очевидно, эта последовательность получена с помощью укрупнения на множестве состояний Σ . Новая последовательность будет

стационарной и обратимой, если таковой была исходная последовательность. Однако последовательность на множестве \mathcal{S} может быть как простой или сложной марковской цепью, так и немарковской. Для того чтобы укрупнение в этом случае не разрушало простую марковость, а также марковость вообще, необходимо соблюдение ряда ограничений, связанных со структурой исходной цепи.

Предположим теперь, что по изложенной схеме образовался гранит, состоящий почти целиком из ортоклаза (Or), кварца (Q) и кислого плагиоклаза (Ab). При этом он возник за счет кристаллизации в эвтектической точке. Можно предполагать, что последовательность зерен Or, Q и Ab образует простую однородную эргодическую марковскую цепь. Пусть на этот гранит наложен вторичный процесс, за счет которого между некоторыми смежными зернами Or могут появляться одно или несколько вторичных зерен Ab. Предполагается, что первичные зерна Or полностью никогда не замещаются. В этом случае мы имеем дело с процессом, названным нами сгущением исходной цепи.

Изучаемые нами преобразования, конечно, представляют идеализацию крайне запутанных естественных процессов. Возможны случаи, когда совершенно различные процессы приводят к одному и тому же результату. При этом, что реально происходило при переработке исходной последовательности в окончательную, остается неясным. Ясность в задачу в таких случаях может быть внесена только тщательным анализом исходных геологических предпосылок.

Пример V.2. Рассмотрим последовательность зерен кварца, ортоклаза и плагиоклаза в массивном песчанике. Как показала Демина (1978), в ряде случаев такая последовательность не отличается от простой однородной эргодической цепи Маркова. Допустим, что этот песчаник погружается на такую глубину, где начинается перекристаллизация кварца. При этом все смежные кварцевые зерна сливаются в одно крупное зерно. Такой процесс перекристаллизации без привноса материала приводит к окончательной последовательности, полученной путем укрупнения по времени. Действительно, если каждую серию кварцевых зерен в исходной последовательности воспринимать как одно зерно, то это эквивалентно тому, что мы перестали различать номера зерен внутри этой серии, т. е. всем зернам внутри серии приписали один и тот же номер. С другой стороны, такой же окончательный результат получился бы, если бы процесс шел так, что происходило бы, скажем, выщелачивание кварцевых зерен, при этом в каждой серии из нескольких кварцевых зерен в исходной последовательности удалялись бы все зерна за исключением одного. Если же серия состояла из одного зерна, то с ним ничего не происходило. Очевидно, что здесь имело бы место преобразование путем разрежения. Это преобразование при описании процесса лишь в терминах состава зерен формально дало бы тот же эф-

фект, что и укрупнение по времени. Прямые наблюдения в шлифах, очевидно, позволили бы определить, с каким типом реального процесса мы имеем дело.

Рассмотрим теперь процесс формирования слоистой толщи, который может быть представлен как чередование взмучивания и осадки частиц, отвечающих по размеру глинистому и песчаному осадку. При этом возникает толща, состоящая из слоев глинистого (γ) и песчаного (π) состава. Одному акту седиментации отвечает одно взмучивание с последующим осадением и погребением осадка. Затем поступает новый материал, и процесс повторяется сначала.

В ходе одного седиментационного акта образуется либо один слой (слой γ или слой π , когда взмучивался гранулометрически однородный осадок), либо два слоя, образующих пакет, который мы будем обозначать ($\pi\gamma$) (когда взмучивался неоднородный осадок).

Таким образом, здесь мы имеем пакеты $A_1=(\pi)$, $A_2=(\gamma)$ и $A_3=(\pi\gamma)$. Скобки указывают, что соответствующие слои образуют пакеты, хотя бы в такой пакет входил всего один слой. Разрез можно интерпретировать как случайную последовательность на множестве пакетов $\Sigma=\{A_1, A_2, A_3\}$. Может возникнуть вопрос, не является ли более естественным описание разреза не как итога актов седиментации, где каждый акт рассматривается как отдельная единица в составе разреза, а как последовательность слоев π и γ , т. е. последовательность на множестве состояний $\Sigma_1=\{\pi, \gamma\}$.

Какую из приведенных интерпретаций разреза следует предпочесть? По-видимому, простота вероятностной структуры последовательности может рассматриваться как аргумент в пользу выбираемого варианта интерпретации, если, конечно, нет прямых геологических наблюдений в пользу того или иного варианта.

Изучение некоторых разрезов показало, что последовательность на состояниях π и γ обладает достаточно сложной вероятностной структурой. В то же время последовательность, образованная пакетами, представляет собой последовательность независимых испытаний или простую цепь Маркова. Такая простая вероятностная структура кажется в данном случае весьма естественной, что заставляет предпочесть пакетную интерпретацию разрезов.

Из сказанного ясно, что мы рассматриваем простую марковскую цепь вида

$$\dots, (\pi)^{(h)}, (\pi\gamma)^{(h+1)}, (\pi\gamma)^{(h+2)}, (\gamma)^{(h+3)}, \dots \quad (V. 1. 1)$$

Во время полевых работ установить границы между пакетами обычно не удается и вместо последовательности (V.1.1) фиксируется последовательность

$$\dots, \pi^{(l)}, \pi^{(l+1)}, \gamma^{(l+2)}, \pi^{(l+3)}, \gamma^{(l+4)}, \gamma^{(l+5)}, \dots \quad (V. 1. 2)$$

Иными словами, мы применили здесь преобразование, названное ранее стиранием границ. Ту же окончательную последова-

тельность (V.1.2) можно получить с помощью другого преобразования. Состояние π из пакета (π) и состояние π из пакета ($\pi\gamma$) имеют разные вероятностные свойства. То же относится к состоянию γ . Поэтому будем обозначать через π_1 песчаный слой из пакета A_1 , через π_2 — песчаный слой из пакета A_3 , соответственно через γ_1 — глинистый слой из пакета A_2 и γ_2 — глинистый слой из пакета A_3 . Очевидно, что случайная последовательность на множестве состояний, $\Sigma_2 = \{\pi_1, \pi_2, \gamma_1, \gamma_2\}$, полностью эквивалентна последовательности пакетов на множестве Σ . Так, например, часть последовательности, приведенная в (V. 1.1), перекодируется в терминах Σ_2 как

$$\dots, \pi_1^{(l)}, \pi_2^{(l+1)}, \gamma_1^{(l+2)}, \pi_2^{(l+3)}, \gamma_2^{(l+4)}, \gamma_1^{(l+5)}, \dots \quad (V. 1.3)$$

Если теперь в (V.1.3) произведем укрупнение $\gamma = \gamma_1 \cup \gamma_2$ и $\pi = \pi_1 \cup \pi_2$, то мы получим последовательность (V.1.2), возникшую в результате стирания границ.

Таким образом, два преобразования, т. е. стирание границ, примененное к (V.1.1), и укрупнение по множеству состояний, примененное к (V.1.3), дают одинаковый результат — (V.1.2).

V.2. УКРУПНЕНИЕ ПО МНОЖЕСТВУ СОСТОЯНИЙ

Введем следующие обозначения. Пусть множество состояний исходной цепи Σ разбито на непересекающиеся подмножества $\Sigma_1, \dots, \Sigma_L$, как это было описано в предыдущем параграфе.

Для исходов испытаний в первоначальной цепи используем обозначения вида $a_i^{(h)}$, где a означает испытание в исходной последовательности, h — его номер; индекс I указывает, что исход этого испытания принадлежит подмножеству Σ_I , а i указывает на состояние i в множестве Σ_I , которое реализовалось в h -м испытании.

Таким образом, i — некоторое состояние из Σ_I , т. е. $i \in \Sigma_I$. После укрупнения по множеству состояний получим случайную последовательность, испытания в которой будем обозначать как $a_I^{(h)}$. Очевидно,

$$a_I^{(h)} = \bigcup_{i \in \Sigma_I} a_i^{(h)}.$$

Укрупнение системы приводит к соответствующему разбиению матрицы переходных вероятностей P на клетки так, что IJ -той клетке соответствуют строки из P , отвечающие состояниям из подмножества Σ_I , и столбцы, отвечающие состояниям из подмножества Σ_J .

Пусть, например, в плите первично магматического гранита имелась последовательность зерен Q, Or и Ab. В результате калиевого метасоматоза, наложенного на этот гранит, зерна плагиоклаза могли обрастать каймой вторичного ортоклаза. Это обрастание могло происходить вокруг всего зерна, вокруг правой

или вокруг левой его частей. Предположим также, что первичные зерна ортоклаза содержат внутри вростки выделений вторичного кварца. Эти вторичные выделения кварца при подсчетах фиксируются как отдельные индивиды. На прямой, пересекающей зерна, могут фиксироваться агрегаты вида $Or_2, Ab_1, Ab_1, Or_3, Or_2, Ab_1, Or_3$, где Ab_1 — зерно первичного плагиоклаза, Or_2 — левая часть вторичной оторочки, Or_3 — правая часть каймы. Могут также встречаться агрегаты Or_1, Q_2, Or_1 , где Or_1 — первичное зерно ортоклаза, Q_2 — вторичное зерно кварца. Кроме того, могут, естественно, наблюдаться сочетания первичных зерен Q_1, Or_1, Ab_1 . В итоге мы будем иметь последовательность зерен до укрупнения на множестве состояний

$$\Sigma = \{Q_1, Q_2, Ab_1, Or_1, Or_2, Or_3\}.$$

Под микроскопом определить первичность или вторичность каждого зерна невозможно. Поэтому безопаснее вообще не выделять генетических типов, а отмечать только состав зерна, т. е. отмечать просто Q, Or, Ab . Эта операция определена нами как укрупнение состояний $Or = Or_1 \cup Or_2 \cup Or_3, Q = Q_1 \cup Q_2, Ab = Ab_1$.

Таким образом, мы имеем разбиение

$$\Sigma = \Sigma_Q \cup \Sigma_{Or} \cup \Sigma_{Ab}.$$

Соответственно этому матрица для последовательности зерен до укрупнения

$$P = \begin{matrix} & Or_1 & Or_2 & Or_3 & Q_1 Q_2 & \cdot & Ab_1 \\ \begin{matrix} Or_1 \\ Or_2 \\ Or_3 \\ Q_1 \\ Q_2 \\ Ab_1 \end{matrix} & \begin{pmatrix} & & & & & & \\ & & & & & & \\ & & P_{Or; Or} & P_{Or; Q} & & & P_{Or; Ab} \\ & & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & P_{Q; Or} & P_{Q; Q} & & P_{Q; Ab} \\ & & & \cdot & \cdot & \cdot & \cdot \\ & & & & & & P_{Ab; Or} & P_{Ab; Q} & P_{Ab; Ab} \end{pmatrix} & \end{matrix} \quad (V. 2. 1)$$

разбивается на клетки, показанные пунктиром. Матрица (V. 2. 1) содержит следующие клетки: $Or, Or; Or, Q; Or, Ab; Q, Or; Q, Q; Q, Ab; Ab, Or; Ab, Q$ и Ab, Ab . Горизонтальный ряд клеток, у которых первым индексом является I , мы называем полосой I . В матрице (V. 2. 1) I принимает значения Or, Q и Ab . Таким образом, в ней мы имеем три полосы: Or, Q и Ab .

V.2.1. Сильное и слабое укрупнение

В качестве исходной цепи рассмотрим простую однородную марковскую цепь, не предполагая сначала ее стационарности. Если последовательность, возникшая в результате заданной системы укрупнения G , снова будет простой марковской цепью, то говорят, что исходная цепь допускает данное укрупнение G (она

укрупняется с помощью G). Будет ли допускать исходная цепь некоторое укрупнение G , зависит от вида G , матрицы переходных вероятностей исходной цепи P и вектора начального распределения $p^{(0)}$. Нетрудно убедиться на простейших числовых примерах, что при фиксированной матрице P и заданном укрупнении G потеря или сохранение марковости могут зависеть от выбора начального распределения $p^{(0)}$. При одних $p^{(0)}$ простая марковость сохраняется, при других теряется. Поэтому свойство укрупняемости классифицируется по отношению к начальному распределению. Если при заданном разбиении (укрупнении) G и заданной матрице P цепь допускает укрупнение при всяком начальном распределении, то говорят, что цепь допускает сильное укрупнение. Если цепь допускает укрупнение хотя бы при одном каком-либо начальном распределении, то говорят, что возможно слабое укрупнение.

Простая однородная марковская цепь с матрицей переходных вероятностей P и разбиением (укрупнением) G допускает сильное укрупнение в том и только в том случае, когда каждая клетка IJ , порожденная разбиением G , обладает следующим свойством: сумма вероятностей по строке внутри клетки является постоянной c_{IJ} , зависящей только от выбранной клетки. Иными словами, для всех $I, J \in \mathcal{S}$ должно выполняться равенство

$$\sum_{j \in \Sigma_J} p(\alpha_{Jj} | \alpha_{Ii}) = c_{IJ}, \quad (V. 2. 2)$$

где константа c_{IJ} не зависит от i . Подчеркнем еще раз, что константа c_{IJ} должна быть одной и той же для каждой строки Ii в клетке IJ , но может меняться от клетки к клетке (Кемени, Снел, 1970).

Пример V.3. Пусть имеется случайная последовательность зерен из первично магматического гранита, зафиксированная на прямой, проходящей через шлиф. Соответствующая последовательность представляет собой однородную марковскую цепь на множестве состояний $\{Or_1, Q_1, Ab_1\}$ с произвольной матрицей переходных вероятностей

$$P = \begin{matrix} & \begin{matrix} Or_1 & Q_1 & Ab_1 \end{matrix} \\ \begin{matrix} Or_1 \\ Q_1 \\ Ab_1 \end{matrix} & \begin{pmatrix} p(Or_1 | Or_1) & p(Q_1 | Or_1) & p(Ab_1 | Or_1) \\ p(Or_1 | Q_1) & p(Q_1 | Q_1) & p(Ab_1 | Q_1) \\ p(Or_1 | Ab_1) & p(Q_1 | Ab_1) & p(Ab_1 | Ab_1) \end{pmatrix} \end{matrix}. \quad (V. 2. 3)$$

О начальном распределении изучаемой последовательности ничего неизвестно: таким образом, стационарность не предполагается. Допустим далее, что в этой исходной последовательности некоторая часть зерен Ab была замещена полностью, зерно на зерно, ортоклазом. Предполагается, что замещение данного зерна Ab происходит с постоянной вероятностью π , не зависящей от возможного замещения остальных зерен. Исследуемый шлиф

изготовлен из породы, претерпевшей изменение по изложенной схеме. При этом мы не можем отличить первичного ортоклаза от вторичного и ведем описание породы в терминах Or , Q и Ab , не выделяя генерации. Спрашивается, какой должна быть матрица переходных вероятностей (V.2.3) для того, чтобы при любом начальном распределении вероятностей $p^{(0)}$ в магматическом граните наблюдаемая последовательность была простой цепью Маркова?

Если ввести состояние Or_2 , где 2 означает вторичное происхождение, то очевидно, что речь пойдет о сильном укрупнении. Рассмотрим сначала последовательность на множестве состояний $\Sigma = \{Or_1, Or_2, Q_1, Ab_1\}$. Легко убедиться, что последовательность на Σ является также простой однородной цепью Маркова. Действительно, что касается каждого зерна, то известно происхождение каждого из них. Это позволяет последовательность на множестве Σ свести к последовательности на исходном множестве. Матрица переходных вероятностей для цепи на множестве Σ будет следующей:

$$\begin{matrix} & Or_1 & Or_2 & Q_1 & Ab_1 \\ Or_1 & p(Or_1|Or_1) \cdot \pi p(Ab_1|Or_1) \cdot p(Q_1|Or_1) \cdot (1-\pi) p(Ab_1|Or_1) \\ Or_2 & p(Or_1|Ab_1) \cdot \pi p(Ab_1|Ab_1) \cdot p(Q_1|Ab_1) \cdot (1-\pi) p(Ab_1|Ab_1) \\ Q_1 & p(Or_1|Q_1) \cdot \pi p(Ab_1|Q_1) \cdot p(Q_1|Q_1) \cdot (1-\pi) p(Ab_1|Q_1) \\ Ab_1 & p(Or_1|Ab_1) \cdot \pi p(Ab_1|Ab_1) \cdot p(Q_1|Ab_1) \cdot (1-\pi) p(Ab_1|Ab_1) \end{matrix} \quad (V.2.4)$$

где переходные вероятности взяты из матрицы (V.2.3).

Для получения наблюдаемой последовательности на множестве $\mathcal{S} = \{Or, Q, Ab\}$ в последовательности на Σ производится укрупнение $Or = Or_1 \cup Or_2$, $Q = Q_1$, $Ab = Ab_1$. Этому укрупнению отвечает разбиение матрицы (V.2.4) на клетки, показанное пунктиром. Применив к (V.2.4) критерий сильной укрупняемости (V.2.2), приходим к следующему выводу: наблюдаемая последовательность после реализации метасоматоза указанного типа гарантированно будет простой марковской цепью в том и только в том случае, когда у матрицы переходных вероятностей магматического гранита будут полностью идентичные строки переходных вероятностей с Or_1 и Ab_1 .

Таким образом, для гарантированного сохранения марковости после метасоматоза должны выполняться следующие равенства:

$$\begin{aligned} p(Q_1|Ab_1) &= p(Q_1|Or_1); & p(Or_1|Ab_1) &= p(Or_1|Or_1); \\ p(Ab_1|Ab_1) &= p(Ab_1|Or_1). \end{aligned} \quad (V.2.5)$$

Если хотя бы одно из равенств (V.2.5) не выполняется, то всегда найдется какое-либо специфическое распределение $p^{(0)}$, являющееся начальным для магматического гранита, такое, что наблюдаемая последовательность не будет простой цепью.

Понятие сильного укрупнения кажется мало полезным в геологических исследованиях. Обычно существуют такие стохастические векторы, которые заведомо невозможны в качестве начального распределения в рассматриваемой задаче.

Рассмотрим множество $\{p^{(0)}\}_{G, P}$ таких начальных распределений, при которых укрупнение g сохраняет простую марковость, когда P — матрица переходных вероятностей исходной цепи. Если множество $\{p^{(0)}\}_{G, P}$ не пусто, то, как известно (Burke, Rosenblatt, 1958), в него входит стационарное начальное распределение $p_{st} \in \{p^{(0)}\}_{G, P}$.

Таким образом, если укрупнение вообще возможно, то оно возможно при стационарном начальном распределении. В связи с этим можно дать новое определение слабой укрупняемости, эквивалентное старому определению, приведенному на с. 304. *Если простая стационарная цепь Маркова при заданном разбиении G и заданной матрице переходных вероятностей P допускает укрупнение, то говорят, что цепь допускает слабое укрупнение.*

В геологических задачах обычно встречается одна из двух ситуаций. Мы можем располагать информацией, к какому классу относится начальное распределение $p^{(0)}$. В этом случае нет необходимости в сильном укрупнении, так как укрупнение проверяется лишь по отношению к указанному классу. Однако во многих случаях информация о начальном распределении отсутствует. Тем не менее, как мы видели в гл. IV, марковская цепь, как правило, через несколько переходов устанавливается.

Таким образом, если мы имеем дело с последовательностью, где первое зафиксированное испытание фактически достаточно удалено от начала процесса, то можно при известной осторожности считать, что цепь уже установилась. В таком случае укрупнение всегда будет слабым. Например, если мы изучаем однообразно построенный разрез слоистой толщи, то имеет смысл начать его исследование на некотором расстоянии от подошвы; если изучается последовательность зерен в шлифе магматической породы, то контакт этой породы с вмещающей не должен находиться в пределах шлифа.

Далее будет рассматриваться только слабое укрупнение. В связи с этим исходная последовательность будет предполагаться всегда стационарной простой марковской цепью.

Необходимые и достаточные условия слабого укрупнения, насколько нам известно, не публиковались. Если бы алгоритм проверки слабой укрупняемости существовал, то ввиду стационарности цепи он сводился бы к конечному числу операций с элементами матрицы переходных вероятностей.

В следующем параграфе будут даны некоторые достаточные условия слабого укрупнения. При этом будет исследовано, когда сохраняется ограниченно марковское свойство и марковский признак у частных переходов при укрупнении (о понятиях ограни-

ченно марковского свойства и марковского признака см. гл. IV, с. 289, 291).

Очевидно, что информация о сохранении ограниченно марковского свойства у частных переходов при укрупнении дает больше, чем информация о сохранении марковости в целом. В гл. IV (п. IV.4.2) отмечалось, что если ограниченная марковость сохраняется для перехода через каждое состояние, то сохраняется и обычная марковость для цепи в целом. Справедливо и обратное. Однако ограниченно марковское свойство может сохраниться при укрупнении для переходов через одни состояния и потеряться при переходе через другие состояния.

V.2.2. Некоторые достаточные условия слабого укрупнения цепи и частных переходов

Предполагается, как отмечалось, что исходная последовательность является стационарной марковской цепью первого порядка на множестве состояний Σ , разбитом на классы $\Sigma_I (I \in \mathcal{S})$ с помощью укрупнения G .

Пусть матрица переходных вероятностей этой последовательности $P_{I_i; J_j}$. Наряду с исходной цепью рассматривается обращенная цепь с матрицей переходных вероятностей $P_{I_i; J_j}^{(-)}$ (об обращенной цепи см. гл. IV, п. IV.3.4).

В книге (Кемени и Снелл, 1970) приводятся следующие достаточные условия слабой укрупняемости: *если хотя бы для одной из матриц $P_{I_i; J_j}$ и $P_{I_i; J_j}^{(-)}$ выполняются условия сильного укрупнения (V.2.2), то допустимо слабое укрупнение.*

Еще раз подчеркнем, что приведенные условия не являются необходимыми для слабого укрупнения.

Пример V.4. Пусть исходная цепь на трех состояниях $\{1, 2, 3\}$ имеет следующую матрицу переходных вероятностей:

$$\begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \end{matrix},$$

причем состояния укрупняются следующим образом: $1 \cup 2 = I$; $3 = J$. Необходимо выяснить, допускает ли цепь слабое укрупнение. Очевидно, что условия (V.2.2) для матрицы P не выполняются. Найдем теперь матрицу переходных вероятностей обращенной цепи. Поскольку P является дважды стохастической

матрицей, стационарное распределение для нее будет (см. гл. IV, с. 252) следующим:

$$p^{(0)} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}, \quad p_1 = p_2 = p_3 = \frac{1}{3}.$$

Формула (IV.3.16) обращения цепи показывает, что в данном случае

$$P^{(-)} = P^T = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \cdot & \cdot & \cdot \\ \frac{1}{4} & \frac{3}{4} & 0 \end{pmatrix}.$$

Очевидно, что условия (V.2.2) для $P^{(-)}$ выполняются.

Таким образом, приведенные выше достаточные условия Кемени и Снелла имеют место, и новая укрупненная последовательность на множестве состояний $\{I, J\}$ является простой марковской цепью.

Достаточные условия Кемени и Снелла редко выполняются на практике. Кроме того, они не применимы к частным переходам. Поэтому мы рассмотрим более гибкие условия, налагая, однако, дополнительные ограничения.

Предположим, что в результате укрупнения получается обратимая последовательность. Если исходная последовательность обратима, то укрупненная также обратима, но обратимость укрупненной последовательности не означает, что исходная была обратима. Таким образом, предположение об обратимости наблюдаемой последовательности охватывает более широкий класс последовательностей, чем предположение об обратимости исходной последовательности. Кроме того, обратимость окончательной последовательности может быть проверена с помощью статистических тестов. Так, большой опыт исследования последовательностей зерен в шлифах гранитов верхнего структурного этажа из различных районов мира показал, что все эти последовательности были обратимыми, независимо от того, представляли ли они простые или сложные цепи Маркова.

Исследованные нами разрезы осадочных толщ также показали статистическую неотличимость чередования слоев от обратимых последовательностей. Подчеркнем, что в соответствии с п. IV.2.2 гл. IV обратимость определена независимо от наличия у последовательности марковских свойств. Имея возможность получить эмпирическое, а иногда и теоретическое подтверждение обратимости в окончательных последовательностях, мы в то же время не можем, как правило, проверить опытным путем обратимость исходной последовательности.

Теорема V.1. Пусть исходная последовательность представляет эргодическую и стационарную цепь Маркова первого порядка, а последовательность после укрупнения обратима. Для того чтобы переход через состояние I был ограниченно марковским, достаточно, чтобы в полосе I выполнялось хотя бы одно из следующих равенств:

$$\begin{aligned} \alpha) \sum_{j \in \Sigma_J} p(\alpha_{Jj}^{(h)} | \alpha_{Ii}^{(h-1)}) &= c_{IJ}, \\ \beta) \sum_{j \in \Sigma_J} p(\alpha_{Jj}^{(h-1)} | \alpha_{Ii}^{(h)}) &= k_{IJ}, \end{aligned} \quad (V. 2.6)$$

для любых $J \in \mathcal{S}$ и произвольных целых $h \geq 1$.

Доказательство. Требуется установить равенство

$$p(\alpha_J^{(h)} | \alpha_I^{(h-1)}, A) = p(\alpha_J^{(h)} | \alpha_I^{(h-1)}) \quad (A \in \mathfrak{A}_{h-2}), \quad (V. 2.7)$$

где $h = 1, 2, \dots, I$ — фиксированное состояние; A — любое событие, зависящее только от исходов испытаний с номерами $0, 1, \dots, h-2$; J — любое состояние из множества \mathcal{S} . Вместо того чтобы доказывать (V. 2.7), можно ограничиться доказательством равенства

$$p(\alpha_J^{(h)} | \alpha_I^{(h-1)}, A) = p(\alpha_J^{(h)} | \alpha_I^{(h-1)}), \quad (V. 2.8)$$

где A — любое событие вида

$$\{a_I^{(0)}, a_K^{(1)}, \dots, a_L^{(h-2)}\},$$

поскольку, как отмечалось в гл. IV, из (V. 2.8) следует (V. 2.7).

Представим условную вероятность в левой части (V. 2.8) в виде отношения безусловных вероятностей (см. гл. II)

$$\frac{p(A, \alpha_J^{(h-1)}, \alpha_J^{(h)})}{p(A, \alpha_I^{(h-1)})}. \quad (V. 2.9)$$

Событиям $\alpha_I^{(h-1)}$ и $\alpha_J^{(h)}$ укрупненной последовательности взаимно однозначно соответствуют события

$$\bigcup_{i \in \Sigma_I} \alpha_{Ii}^{(h-1)} \text{ и } \bigcup_{j \in \Sigma_J} \alpha_{Jj}^{(h)}$$

в исходной цепи. При этом событию A соответствует некоторое событие B в терминах исходной цепи. Используя простую марковость исходной цепи, (V. 2.9) можно представить как

$$\begin{aligned} & \frac{p\left(B, \bigcup_{i \in \Sigma_I} \alpha_{Ii}^{(h-1)}, \bigcup_{j \in \Sigma_J} \alpha_{Jj}^{(h)}\right)}{p\left(B, \bigcup_{i \in \Sigma_I} \alpha_{Ii}^{(h-1)}\right)} = \\ &= \frac{\sum_{i \in \Sigma_I, j \in \Sigma_J} p(B) p(\alpha_{Ii}^{(h-1)} | B) p(\alpha_{Jj}^{(h)} | \alpha_{Ii}^{(h-1)})}{\sum_{i \in \Sigma_I} p(B) p(\alpha_{Ii}^{(h-1)} | B)}. \end{aligned} \quad (V. 2.10)$$

Предположим теперь, что в (V.2.6) выполняется равенство α). Если это так, то произведем сначала суммирование в числителе (V.2.10) по j на h -м месте. Вынося общий множитель c_{IJ} за знак суммирования и сокращая, получим

$$p(a_j^{(h)} | a_I^{(h-1)}, A) = c_{IJ}.$$

Аналогично для правой части (V.2.8)

$$p(a_j^{(h)} | a_I^{(h-1)}) = c_{IJ}.$$

Итак, теорема доказана при условии α).

Допустим теперь, что выполняется равенство β) из (V.2.6). Тогда используем предположенную обратимость укрупненной последовательности и ее стационарность. Это дает возможность переписать (V.2.9) как

$$\frac{p(a_J^{(0)}, a_I^{(1)}, A'_{h-2})}{p(a_I^{(1)}, A'_{h-2})}, \quad (V.2.11)$$

где A' — событие вида

$$\{a_L^{(2)}, \dots, a_K^{(h)}, a_T^{(h+1)}\},$$

полученное из A изменением порядка и сдвигом временных параметров. Подстановка A' вместо A допустима, так как последовательность в терминах a_M обратима (это дает возможность переписать ее в обратном направлении) и стационарна (это дает возможность сдвинуть ее к началу 0).

Преобразуем далее (V.2.11). При этом используем условие, что обращенная исходная последовательность также является простой марковской:

$$\frac{p\left(B', \bigcup_{i \in \Sigma_I} a_{Ii}^{(1)}, \bigcup_{j \in \Sigma_J} a_{Jj}^{(0)}\right)}{p\left(B'_{h-2}, \bigcup_{i \in \Sigma_I} a_{Ii}^{(1)}\right)} = \frac{\sum_{i \in \Sigma_I, j \in \Sigma_J} p(B') p(a_{Ii}^{(1)} | B') p(a_{Jj}^{(0)} | a_{Ii}^{(1)})}{\sum_{i \in \Sigma_I} p(B') p(a_{Ii}^{(1)} | B')}, \quad (V.2.12)$$

где B' соответствует событию A' в терминах исходной цепи.

Суммируем в (V.2.12) сначала по j на 0-м месте. Условие

$$\sum_{j \in \Sigma_J} p(a_{Jj}^{(h-1)} | a_{Ii}^{(h)}) = k_{IJ}$$

позволяет вынести постоянный множитель k_{IJ} за знак суммирования. Последующее сокращение в (V.2.12) приводит к равенству

$$p(a_j^{(h)} | a_I^{(h-1)}, A') = k_{IJ}$$

и аналогично к

$$p(a_j^{(h)} | a_I^{(h-1)}) = k_{IJ}.$$

Этим заканчивается доказательство теоремы.

Поскольку наличие ограниченно марковского перехода по каждому состоянию, как отмечалось в гл. IV, влечет простую марковость последовательности в целом, то в качестве следствия из теоремы V.1 мы получаем следующее достаточное условие.

Т е о р е м а V.2. *Если исходная простая цепь Маркова является эргодической и стационарной, а укрупненная последовательность обратима, то для сохранения простой марковости после укрупнения достаточно, чтобы хотя бы одно из условий α) или β) в (V.2.6.) выполнялось для каждой полосы в матрице P .*

Допустим теперь, что равенства α) или β) в (V.2.6) выполняются для полосы матрицы P , однако мы не знаем, обратима ли окончатальная (наблюдаемая, т. е. укрупненная) последовательность. В этом случае для укрупненной последовательности простая марковость либо марковость вообще может теряться, но гарантировано сохранение марковского признака.

Т е о р е м а V.3. *Если исходная последовательность является стационарной и эргодической цепью Маркова первого порядка и одно из равенств α) или β) выполняется в полосе I матрицы P , то после укрупнения переход через I обладает марковским признаком (см. гл. IV).*

Д о к а з а т е л ь с т в о. Требуется показать, что выполнение одного из равенств α) или β) в (V.2.6) влечет равенство

$$p(a_J^{(h)} | a_I^{(h-1)}, a_H^{(h-2)}) = p(a_J^{(h)} | a_I^{(h-1)}). \quad (V. 2. 13)$$

Если выполняется α) в (V.2.6), то (V.2.13) доказывается с помощью простых вычислений, аналогичных приведенным при доказательстве теоремы (V.1). Если выполняется β) в (V.2.6), то вычисления несколько усложняются. Рассмотрим последний случай.

Левую часть (V.2.13) можно представить в виде

$$p(a_J^{(h)} | a_I^{(h-1)}, a_H^{(h-2)}) = \frac{p(a_J^{(h)}) p(a_I^{(h-1)} | a_J^{(h)}) p(a_H^{(h-2)} | a_I^{(h-1)}, a_J^{(h)})}{p(a_I^{(h-1)}) p(a_H^{(h-2)} | a_I^{(h-1)})}. \quad (V. 2. 14)$$

Как мы только что установили при доказательстве теоремы V.1, условие β) из (V.2.6) влечет равенства

$$p(a_H^{(h-2)} | a_I^{(h-1)}, a_J^{(h)}) = k_{IH}, \quad p(a_H^{(h-2)} | a_I^{(h-1)}) = k_{IH}.$$

Учитывая эти равенства, приводим (V.2.14) к виду

$$p(a_J^{(h)} | a_I^{(h-1)}, a_H^{(h-2)}) = \frac{p(a_J^{(h)}) p(a_I^{(h-1)} | a_J^{(h)})}{p(a_I^{(h-1)})} = p(a_J^{(h)} | a_I^{(h-1)}),$$

что и требовалось доказать.

Распространяя выполнение условий теоремы V.3 на все полосы, т. е. на всю матрицу P в целом, из теоремы V.2 получаем следующее.

С л е д с т в и е. Если условия α) или β) из (V.2.6) выполняются для каждой полосы матрицы P , то после укрупнения гарантировано сохранение марковского признака для перехода через любое состояние. Настоящая марковость может при этом теряться.

Мы подробно остановились на условиях сохранения марковского признака при слабом укрупнении, так как в этих условиях могут возникать случайные последовательности, весьма опасные с точки зрения правильности интерпретации значений статистических тестов, проверяющих порядок марковости. Действительно, пусть при применении тестов, проверяющих марковские альтернативы, которые далее подробно рассматриваются в гл. VI, производится браковка нулевого порядка в пользу первого, затем второй порядок бракуется в пользу первого. Если на этом остановиться, то можно принять неправильное решение о первом порядке марковости, хотя фактически мог иметь место только марковский признак. Возможно, что дальнейшая статистическая проверка альтернатив дала бы решение в пользу третьего порядка в альтернативе второго порядка против третьего.

Подчеркнем еще раз, что для получения последовательности с марковским признаком, но без марковского свойства требуется, чтобы для полос матрицы P выполнялись условия α) или β) и в то же время укрупненная последовательность не обладала обратимостью. Это следует из теоремы V.3. Отметим также, что для возникновения подобного рода случайной последовательности необходимо укрупнение не менее чем трех состояний в одном классе. Это утверждение оказывается простым следствием теоремы V.4, которая будет дана в следующем параграфе.

П р и м е р V.5. Рассмотрим еще раз схему слоеобразования, приводящую к последовательности пакетов (π) , (γ) и $(\pi\gamma)$, которую мы уже изучали во второй части примера V.2. При этом мы предположим, что последовательность пакетов является бернуллиевской. Наблюдаемая последовательность слоев в разрезе получается, как отмечалось в примере V.2, за счет стирания границ между пакетами. Опираясь на материалы этого параграфа о достаточных условиях сохранения марковости, рассмотрим, какой марковской структурой будет обладать наблюдаемая последовательность слоев.

Последовательности пакетов отвечает следующая матрица переходных вероятностей:

$$P_1 = \begin{matrix} & \begin{matrix} (\pi) & (\gamma) & (\pi\gamma) \end{matrix} \\ \begin{matrix} (\pi) \\ (\gamma) \\ (\pi\gamma) \end{matrix} & \begin{pmatrix} p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \end{pmatrix} \end{matrix}.$$

Равные вероятности в каждом из столбцов соответствуют предположению, что пакеты образуют последовательность Бернулли. Обозначим, как и ранее в примере V.2, π_1 — песчаный слой из

пакета (π); π_2 — песчаный слой из пакета ($\pi\gamma$); γ_1 — глинистый слой из пакета (γ) и γ_2 — глинистый слой из пакета ($\pi\gamma$). Тогда на множестве состояний $\Sigma = \{\pi_1, \pi_2, \gamma_1, \gamma_2\}$, как уже было показано в примере V.2, возникнет простая марковская цепь, стационарная и эргодическая, с матрицей переходных вероятностей

$$P_2 = \begin{matrix} & \pi_1 & \pi_2 & \cdot & \gamma_1 & \gamma_2 \\ \begin{matrix} \pi_1 \\ \pi_2 \\ \cdot \\ \gamma_1 \\ \gamma_2 \end{matrix} & \begin{pmatrix} p_1 & p_3 & \cdot & p_2 & 0 \\ 0 & 0 & \cdot & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p_1 & p_3 & \cdot & p_2 & 0 \\ p_1 & p_3 & \cdot & p_2 & 0 \end{pmatrix} \end{matrix},$$

где параметры для переходных вероятностей взяты из P_1 . Пунктиром в P_2 показано то укрупнение, которое приводит к наблюдаемой последовательности слоев. Стационарное распределение, как нетрудно показать (гл. IV, с. 255),

$$P_{st} = P_0 = \begin{pmatrix} p(\pi_1) = \frac{p_1}{1+p_3} \\ p(\pi_2) = \frac{p_3}{1+p_3} \\ p(\gamma_1) = \frac{p_2}{1+p_3} \\ p(\gamma_2) = \frac{p_3}{1+p_3} \end{pmatrix}.$$

Соответственно обращенной цепи будет отвечать матрица переходных вероятностей

$$P_2^{(-)} = \begin{matrix} & \pi_1 & \pi_2 & \cdot & \gamma_1 & \gamma_2 \\ \begin{matrix} \pi_1 \\ \pi_2 \\ \cdot \\ \gamma_1 \\ \gamma_2 \end{matrix} & \begin{pmatrix} p_1 & 0 & \cdot & p_2 & p_3 \\ p_1 & 0 & \cdot & p_2 & p_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p_1 & 0 & \cdot & p_2 & p_3 \\ 0 & 1 & \cdot & 0 & 0 \end{pmatrix} \end{matrix}.$$

Рассмотрим теперь достаточные условия Кемени и Снелла (1970, с. 176). Поскольку, как в P_2 , так и в $P_2^{(-)}$ равенства сумм по строчкам внутри клеток не выдерживаются, эти условия не выполняются. В силу того что условия эти достаточные, вывод о наличии или отсутствии марковской структуры у последовательности на состояниях π и γ сделать нельзя.

Перейдем теперь к теореме V.2.

В матрице P_2 имеется равенство сумм по строкам внутри клетки в полосе γ . В матрице $P_2^{(-)}$ наблюдается равенство сумм по строкам внутри клетки в полосе π . Таким образом, условия теоремы V.2 выполняются. В результате мы можем утверждать, что наблюдаемая последовательность слоев должна обладать марковским признаком для перехода как через π , так и через γ .

Наличие марковского признака в отличие от марковости легко проверяется статистически по наблюдениям. Отметим, что фактическая проверка переходов через состояния (слои) π и γ в реально изучавшихся разрезах (Челекен; о-в Жилой, Азербайджан) показала, что переход через одно состояние обладал марковским признаком, а через другое состояние он этим признаком не обладал. Из этого следует, что для изучавшихся разрезов либо схема седиментации пакетами не верна, либо предположение о том, что пакеты чередовались бернуллиевским способом, ошибочно. Этот вопрос подробно будет рассмотрен во второй части этой книги.

Используем теперь теорему V.1.

Нетрудно показать, что окончательная последовательность слоев (т. е. на множестве $\{\pi, \gamma\}$) будет обратимой, если последовательность пакетов была бернуллиевской. Так, например, установим, что

$$p(\pi^{(h-1)}, \gamma^{(h)}) = p(\gamma^{(h-1)}, \pi^{(h)}),$$

что должно иметь место для обратимой последовательности.

Действительно,

$$\begin{aligned} p(\pi^{(h-1)}, \gamma^{(h)}) &= p(\pi_1^{(h-1)}, \gamma_1^{(h)}) + p(\pi^{(h-1)}, \gamma_2^{(h)}) + p(\pi_2^{(h-1)}, \gamma_1^{(h)}) + \\ &+ p(\pi_2^{(h-1)}, \gamma_2^{(h)}) = p(\pi_1) p(\gamma_1 | \pi_1) + p(\pi_1) p(\gamma_2 | \pi_1) + \\ &+ p(\pi_2) p(\gamma_1 | \pi_2) + p(\pi_2) p(\gamma_2 | \pi_2) \end{aligned} \quad (V. 2. 15)$$

и

$$\begin{aligned} p(\gamma^{(h-1)}, \pi^{(h)}) &= p(\gamma_1^{(h-1)}, \pi_1^{(h)}) + p(\gamma_1^{(h-1)}, \pi_2^{(h)}) + p(\gamma_2^{(h-1)}, \pi_1^{(h)}) + \\ &+ p(\gamma_2^{(h-1)}, \pi_2^{(h)}) = p(\gamma_1) p(\pi_1 | \gamma_1) + p(\gamma_1) p(\pi_2 | \gamma_1) + \\ &+ p(\gamma_2) p(\pi_1 | \gamma_2) + p(\gamma_2) p(\pi_2 | \gamma_2). \end{aligned} \quad (V. 2. 16)$$

Подставляя в (V.2.15) и в (V.2.16) безусловные и переходные вероятности соответственно из P , p_{st} и P_2 , убеждаемся, что

$$p(\pi^{(h-1)}, \gamma^{(h)}) = p(\gamma^{(h-1)}, \pi^{(h)}).$$

Точно так же легко доказать обратимость относительно других сочетаний слоев.

Таким образом, условия теоремы V.1 выполняются, и мы можем сделать заключение, что в случае бернуллиевского чередования пакетов чередование слоев, наблюдаемое в разрезе, должно отвечать простой марковской цепи.

V.2.3. Критерий слабого укрупнения при дополнительных ограничениях

Для получения критериев слабого укрупнения рассмотрим следующие дополнительные ограничения:

- а) исходная цепь является обратимой,
- б) укрупненная последовательность является обратимой,
- в) состояния укрупняются не более, чем по два.

Иными словами, в случае в) разбиение таково, что либо $\Sigma_I = \{I_1, I_2\}$, либо $\Sigma_I = \{I\}$.

В геологических исследованиях вопрос о том, обратима ли исходная цепь, определяется природными соотношениями и, как правило, не может быть проверен статистическими методами. В силу каких причин окончательная последовательность (после укрупнения) может считаться обратимой, рассмотрено в п. V.2.2. Наконец, случай, когда проводится укрупнение состояний по два, имеет решение и представляется естественным во многих задачах геологии. Например, при изучении горных пород естественно различать первичные и вторичные минералы (одного состава), в сериях осадочных пород — аналогичные породы трансгрессивных и регрессивных членов серии и т. д.

Рассмотрим сначала случай, когда исходная цепь предполагается обратимой. В книге Кемени и Снелла (1970, теоремы 6.4.6—6.4.8) для него доказан следующий критерий: *если исходная эргодическая и стационарная простая марковская цепь обратима, то критерий слабого укрупнения совпадает с критерием сильного укрупнения, приведенным нами на с. 304 под номером V.2.2.*

Пример V.6. Рассмотрим снова кристаллизацию тройной системы с эвтектикой, разбиравшуюся нами в начале примера V.1, с зернами I_1 , выделившимися на первой стадии кристаллизации, с зернами I_2, J_2 , выделившимися на второй стадии, и с зернами I_3, J_3, K_3 в эвтектической точке. Если в шлифе через агрегат, образованный указанными индивидами, провести прямую, то естественно будет предположить, что последовательность зерен на этой прямой окажется обратимой.

В примере V.1 было высказано предположение, что эта последовательность является простой марковской цепью, это предположение мы примем и здесь. В силу изотропности расплава, из которого кристаллизуются индивиды, для любой прямой, проходящей через агрегат, кажется, что обратимость этой цепи совершенно банальна. Однако предположение об обратимости последовательности приводит к некоторым далеко идущим следствиям относительно процесса кристаллизации.

Рассмотрим матрицу переходных вероятностей для исходной последовательности

$$\begin{array}{c}
 I_1 \quad I_2 \quad I_3 \quad J_2 \quad J_3 \quad K_3 \\
 \begin{array}{c}
 I_1 \\
 I_2 \\
 I_3 \\
 J_2 \\
 J_3 \\
 K_3
 \end{array}
 \left(
 \begin{array}{cccccc}
 & & & & & p(K_3 | I_1) \\
 & p_{I; I} & & p_{I; J} & & p(K_3 | I_2) \\
 & & & & & p(K_3 | I_3) \\
 & \dots & \dots & \dots & \dots & \dots \\
 & & & & & p(K_3 | J_2) \\
 & p_{J; I} & & p_{J; J} & & p(K_3 | J_3) \\
 & \dots & \dots & \dots & \dots & \dots \\
 p_{K; I} & & p_{K; J} & & p(K_3 | K_3) &
 \end{array}
 \right)
 \end{array}$$

Предположим, что исходная цепь обратима. Кроме того, на основании статистических расчетов примем, что окончательная последовательность — простая марковская цепь. Тогда из только что приведенного критерия будет следовать, что сумма по строчкам внутри клеток постоянна. Взяв клетки IK, JK , получим

$$p(K_3 | I_1) = p(K_3 | I_2) = p(K_3 | I_3); \quad p(K_3 | J_2) = p(K_3 | J_3).$$

Таким образом, вероятность перехода на состояние K_3 не зависит от генерации, с которой осуществляется переход на K_3 . Эти равенства переходных вероятностей не выводятся непосредственно из модели кристаллизации (см. гл. IV и особенно работу автора (Vistelius, 1972)). Поэтому должен существовать некоторый специфический механизм размещения зерен по генерациям в пространстве, занятом кристаллизующимся расплавом.

Перейдем теперь к случаю, когда состояния укрупняются по два. Приведем сначала необходимые условия сохранения марковости в этом случае.

Теорема V.4. Пусть в исходной стационарной эргодической цепи Маркова первого порядка подмножество Σ_I содержит два состояния I_1 и I_2 . Для того чтобы при укрупнении частный переход через I сохранил марковский признак или ограниченно марковское свойство, или марковское свойство, необходимо, чтобы выполнялось хотя бы одно из условий α) или β) в V.2.6.

Д о к а з а т е л ь с т в о. Если переход через I после укрупнения частный марковский или ограниченно марковский, то тем более этот переход должен обладать марковским признаком, т. е. должно выполняться равенство

$$p(a_J^{(h)} | a_I^{(h-1)}, a_R^{(h-2)}) = p(a_J^{(h)} | a_I^{(h-1)}) \quad (V. 2. 17)$$

для всех $J, R \in \mathcal{S}$ и данного I .

Используя стационарность и простую марковость исходной цепи, представим равенство (V. 2. 17) в виде

$$\left[\sum_{\substack{i \in \Sigma_I, \\ j \in \Sigma_J, \\ r \in \Sigma_R}} p(\alpha_{Rr}^{(h-2)}) p(\alpha_{Ii}^{(h-1)} | \alpha_{Rr}^{(h-2)}) p(\alpha_{Jj}^{(h)} | \alpha_{Ii}^{(h-1)}) \right] \sum_{i \in \Sigma_I} p(\alpha_{Ii}^{(h-1)}) - \\ - \sum_{\substack{i \in \Sigma_I, \\ r \in \Sigma_R}} p(\alpha_{Rr}^{(h-2)}) p(\alpha_{Ii}^{(h-1)} | \alpha_{Rr}^{(h-2)}) \sum_{\substack{i \in \Sigma_I, \\ j \in \Sigma_J}} p(\alpha_{Ii}^{(h-1)}) p(\alpha_{Jj}^{(h)} | \alpha_{Ii}^{(h-1)}) = 0 \quad (V. 2. 18)$$

или в эквивалентной форме:

$$\sum_{i \in \Sigma_I} p(\alpha_{Ii}^{(h)}, \alpha_R^{(h-1)}) [p(\alpha_{Ii}^{(h-1)}) p(\alpha_{Ji}^{(h)} | \alpha_{Ii}^{(h-1)}) - p(\alpha_{Ji}^{(h)}, \alpha_{Ii}^{(h-1)})] = 0. \quad (V. 2. 19)$$

Если в (V. 2. 19) заменить первый множитель

$$p(\alpha_{Ii}^{(h)}, \alpha_R^{(h-1)}) \text{ на } p(\alpha_{Ii}^{(h-1)})$$

или, что то же самое, просуммировать по всем $R \in \mathcal{S}$, то получится тождество

$$\sum_{i \in (\Sigma_I)} p(\alpha_{I_i}^{(h-1)}) [p(\alpha_{I_i}^{(h-1)}) p(\alpha_{J_i}^{(h)} | \alpha_{I_i}^{(h-1)}) - p(\alpha_{J_i}^{(h)}, \alpha_{I_i}^{(h-1)})] = 0. \quad (\text{V. 2. 20})$$

Обозначим одинаковые выражения в квадратных скобках в (V.2.19) и (V.2.20) как $x(i, J)$. Учитывая, что i может принимать только значения 1 и 2, мы видим, что (V.2.19) вместе с (V.2.20) представляет линейную однородную систему из двух уравнений с двумя неизвестными $x(1, J)$ и $x(2, J)$. Известно, что однородная линейная система либо имеет только нулевое решение, либо ее определитель равен нулю. В первом случае мы имеем

$$x(1, J) = x(2, J) = 0, \quad (\text{V. 2. 21})$$

во втором (заменяя $(h-1, h-2)$ на $(h, h-1)$):

$$\Delta = \begin{vmatrix} p(\alpha_{I_1}^{(h)}, \alpha_{I_1}^{(h-1)}) & p(\alpha_{I_2}^{(h)}, \alpha_{I_2}^{(h-1)}) \\ p(\alpha_{I_1}^{(h)}) & p(\alpha_{I_2}^{(h)}) \end{vmatrix} = 0. \quad (\text{V. 2. 22})$$

Если (V.2.21) выполняется при всех $J \in \mathcal{S}$, то это эквивалентно условию α) из (V.2.6).

Если при каком-нибудь J (V.2.21) не выполняется, то тогда (V.2.22) должно выполняться при всех $R \in \mathcal{S}$ и данном I . Поскольку R можно выбрать независимо от I , равенство (V.2.22) эквивалентно условию β) из (V.2.6). Этим завершается доказательство теоремы.

Если мы теперь объединим необходимые условия доказанной теоремы V.4 с достаточными условиями теоремы V.1, то получим следующее следствие.

С л е д с т в и е. Пусть исходная простая цепь эргодична и стационарна, а укрупненная последовательность обратима. Пусть кроме того подмножество Σ_I содержит не более двух состояний. Тогда для сохранения после укрупнения ограниченной марковости при переходе через I необходимо и достаточно, чтобы в полосе I выполнялось хотя бы одно из условий α) или β) из (V.2.6).

Если ограниченная марковость сохраняется у перехода через каждое состояние, то сохраняется и простая марковость в целом, как неоднократно отмечалось ранее. Поэтому в качестве другого следствия получаем следующее утверждение.

Т е о р е м а V.5. Пусть исходная простая цепь эргодична и стационарна, а укрупненная последовательность обратима. Пусть кроме того укрупнение производится так, что одновременно укрупняется не более двух состояний в одном классе. Тогда необходимые и достаточные условия сохранения простой марковости после укрупнения заключаются в выполнении хотя бы одного из равенств α) или β) в (V.2.6) для каждой полосы матрицы переходных вероятностей P ,

Пример V.7. Пусть исходная последовательность есть последовательность зерен кварца (Q), ортоклаза (Or) и плагиоклаза (Ab), являющихся главными породообразующими минералами гранита магматического происхождения. Такая последовательность, как отмечалось в литературе (Vistelius, 1972), скорее всего является эргодической простой цепью Маркова, что мы и примем в этом примере. Допустим теперь, что порода подверглась вторичным изменениям, заключавшимся в том, что некоторые индивиды плагиоклаза оказались окруженными непрерывной каймой ортоклаза. При этом предполагается, что ни одно из зерен, соседних с окаймляемым, полностью этой каймой не замещается. На пересечении с прямой каждое зерно плагиоклаза, окруженное каймой, дает то, что мы называем пакетом, т. е. последовательность из трех зерен Or, Ab, Or. Предположим, что окаймление каждого такого зерна плагиоклаза происходит с постоянной вероятностью $0 < \pi < 1$.

Таким образом, окаймление такого зерна плагиоклаза не зависит от того, что произойдет с остальными зернами плагиоклаза. Возникает вопрос — может ли последовательность, образованная зернами Or, Q и Ab, после возникновения каем из Or остаться простой марковской цепью. Покажем, что ответ отрицателен, т. е. марковость первого порядка всегда будет теряться. Иными словами, простая марковость исчезает при любом выборе параметра $0 < \pi < 1$ и любых параметрах исходной стационарной эргодической цепи Маркова первого порядка.

Для того чтобы доказать это, введем случайную последовательность на множестве состояний $\Sigma = \{Ab_1, Ab_2, Or_1, Or_2, Or_3, Q_1\}$, где Ab_1 — зерно плагиоклаза, не подвергшееся окаймлению; Ab_2 — зерно плагиоклаза, окруженное каймой; Or_1 — зерно первичного ортоклаза; Or_2 — участок каймы, находящейся слева от окаймленного зерна плагиоклаза; Or_3 — участок той же каймы, находящейся справа от окаймленного зерна плагиоклаза; Q_1 — зерно кварца.

Вопрос о том, на каком множестве состояний рационально строить исходную (укрупняемую) последовательность для того, чтобы решить задачу о влиянии укрупнения на марковские свойства последовательности, может решаться разными путями. Так, например, можно взять исходную последовательность на множестве $\Sigma_1 = \{Ab_1, Or_1, Q_1, Or_2\}$, где 1 означает первичное, а 2 — вторичное происхождение зерна. Однако эта последовательность оказывается существенно немарковской. Действительно,

$$P(Ab_1^{(h)} | Or_2^{(h-1)}, Ab_1^{(h-2)}, Or_2^{(h-3)}, Ab_1^{(h-4)}, Or_2^{(h-5)}, Ab_1^{(h-6)}) < 1,$$

но

$$P(Ab_1^{(h)} | Or_2^{(h-1)}, Ab_1^{(h-2)}, Or_2^{(h-3)}, Ab_1^{(h-4)}, Or_2^{(h-5)}, Q_1^{(h-6)}) = 1.$$

Поскольку конструирование подобных переходных вероятностей можно продолжать как угодно далеко, то, очевидно, здесь мы

имеем дело с существенной немарковостью, а не с марковостью шестого порядка.

В то же время для последовательности на Σ легко устанавливается ее простая марковость. Перейдем к доказательству того, что последовательность на Σ — простая цепь Маркова. Рассмотрим вероятности переходов через состояния Ab_1 :

$$p(Ab_1^{(h)} | Ab_1^{(h-1)}, A) = p(Ab | Ab) \quad (A \in \mathfrak{A}_{h-2}),$$

где переходная вероятность $p(Ab | Ab)$ взята для последовательности зерен до того, как порода подверглась преобразованию.

Далее

$$p(Ab_2^{(h)} | Ab_1^{(h-1)}, A) = 0, \quad p(Or_1^{(h)} | Ab_1^{(h-1)}, A) = p(Or | Ab),$$

$$p(Or_2^{(h)} | Ab_1^{(h-1)}, A) = \pi p(Ab | Ab), \quad p(Or_3^{(h)} | Ab_1^{(h-1)}, A) = 0,$$

$$p(Q_1^{(h)} | Ab_1^{(h-1)}, A) = p(Q | Ab) \quad (A \in \mathfrak{A}_{h-2}),$$

где переходные вероятности для состояний без индексов отнесены к последовательности до ее преобразования. Таким образом, переход через Ab_1 обладает ограниченно марковским свойством.

Аналогично можно показать, что переходы через все другие состояния из Σ являются ограниченно марковскими. Из этого следует, что последовательность на Σ является простой марковской цепью. Иными словами, мы укрупняем простую марковскую цепь. Легко также показать, что эта цепь является стационарной и эргодической. В этой цепи состояния Ab_1 и Ab_2 укрупняются по два. Это позволяет нам использовать необходимые условия из теоремы V.4.

Рассмотрим в матрицах переходных вероятностей P и $P^{(-)}$ (обращенной цепи) для последовательности на множестве Σ полюсу Ab . Без вычислений ясно, что

$$p(Q_1 | Ab_1) > 0, \quad \text{но} \quad p(Q_1 | Ab_2) = 0$$

и что

$$p^{(-)}(Q_1 | Ab_1) > 0, \quad \text{но} \quad p^{(-)}(Q_1 | Ab_2) = 0. \quad (V.2.23)$$

Из (V.2.23) следует, что необходимые условия теоремы V.4 не выполняются. Поэтому простая марковость при укрупнении будет потеряна по крайней мере для перехода через Ab . Доказать этот факт можно было бы значительно короче; более длинный путь был выбран для того, чтобы показать типичные рассуждения, применяемые при изучении вопроса об укрупнении состояний.

V.3. СГУЩЕНИЕ И РАЗРЕЖЕНИЕ

Как отмечалось в § V.4, сгущением мы называем такое преобразование, при котором в реализацию добавляются новые исходы испытаний, размещаемые между старыми исходами. После этого производится перенумерация всех испытаний в естественном порядке. Предполагается, что взаимное положение исходов старых испытаний не меняется. Исходы старых и новых испытаний, от-

вещающие одному и тому же состоянию, в окончательной последовательности не различаются, т. е. производится укрупнение. Очевидно, что если исходы новых испытаний не принадлежат множеству \mathcal{S} состояний цепи, то укрупнения состояний в окончательной последовательности не будет. Сложнее, но важнее для геологических задач случай, когда исходами дополнительных испытаний могут быть состояния из множества \mathcal{S} . Тогда появляются исходы испытаний, отвечающие одному и тому же состоянию из множества \mathcal{S} , часть которых принадлежит старой цепи, а часть введена в нее новыми исходами испытаний.

Рассмотрим случай, когда исходы новых испытаний могут давать лишь одно состояние, скажем, $I \in \mathcal{S}$. Остальные состояния, принадлежащие множеству $\mathcal{S}' = \mathcal{S} \setminus \{I\}$, во вторичном преобразовании не участвуют, не появляются как исходы новых испытаний.

При рассмотрении разрежения мы ограничиваемся лишь простейшим случаем, когда из реализации исходной цепи удаляются (вычеркиваются) все те испытания, в которых появились состояния из некоторого подмножества состояний \mathcal{S}_1 основного множества состояний \mathcal{S} . После этого остающиеся испытания перенумеровываются в естественном порядке. Для сокращения выкладок мы часто будем использовать следующее положение.

Лемма V.1. Пусть в некоторой случайной последовательности $\mathcal{A}^{(+)}$ есть событие, отнесенное к будущему, $\mathcal{A}^{(0)}$ — к настоящему, а $\{\mathcal{A}_S^{(-)}\}$ — множество, каждое событие $\mathcal{A}_S^{(-)}$ которого принадлежит прошлому. Пусть события из $\{\mathcal{A}_S^{(-)}\}$ несовместимы, т. е.

$$\mathcal{A}_K^{(-)} \cap \mathcal{A}_L^{(-)} = \emptyset \quad (K \neq L), \quad (V.3.1)$$

если $\mathcal{A}_K^{(-)}, \mathcal{A}_L^{(-)} \in \{\mathcal{A}_S^{(-)}\}$, и таковы, что настоящее $\mathcal{A}^{(0)}$ осуществляется обязательно по крайней мере с одним из этих событий, т. е.

$$p \left[\bigcup_S (\mathcal{A}_S^{(-)} | \mathcal{A}^{(0)}) \right] = 1. \quad (V.3.2)$$

Если вероятность $p(\mathcal{A}^{(+)} | \mathcal{A}^{(0)} \mathcal{A}_S^{(-)})$ не зависит от выбора $\mathcal{A}_S^{(-)}$ во множестве $\{\mathcal{A}_S^{(-)}\}$, то

$$p(\mathcal{A}^{(+)} | \mathcal{A}^{(0)} \mathcal{A}_S^{(-)}) = p(\mathcal{A}^{(+)} | \mathcal{A}^{(0)}). \quad (V.3.3)$$

Если (V.3.1) и (V.3.2) выполняются, то (V.3.3) справедливо независимо от наличия или отсутствия марковской структуры у последовательности.

Условия (V.3.1) и (V.3.2) существенны, и без них равенство (V.3.3) в общем случае неверно.

Доказательство. Согласно условию (V.3.2),

$$p(\mathcal{A}^{(+)} | \mathcal{A}^{(0)}) = p \left[\mathcal{A}^{(+)} \left| \left(\mathcal{A}^{(0)}, \bigcup_S \mathcal{A}_S^{(-)} \right) \right. \right] = \frac{p \left(\bigcup_S \mathcal{A}_S^{(-)}, \mathcal{A}^{(0)}, \mathcal{A}^{(+)} \right)}{p \left(\bigcup_S \mathcal{A}_S^{(-)}, \mathcal{A}^{(0)} \right)}. \quad (V.3.4)$$

В силу условия (V.3.1) правую часть (V.3.4) можно переписать в виде

$$\frac{\sum_S p(\mathcal{A}_S^{(-)}, \mathcal{A}^{(0)}, \mathcal{A}^{(+)})}{\sum_S p(\mathcal{A}_S^{(-)}, \mathcal{A}^{(0)})} = \frac{\sum_S p(\mathcal{A}_S^{(-)}, \mathcal{A}^{(0)}) p(\mathcal{A}^{(+)} | \mathcal{A}^{(0)}, \mathcal{A}_S^{(-)})}{\sum_S p(\mathcal{A}_S^{(-)}, \mathcal{A}^{(0)})}. \quad (\text{V.3.5})$$

Поскольку $p(\mathcal{A}^{(+)} | \mathcal{A}^{(0)}, \mathcal{A}_S^{(-)})$ по предположению не зависит от S , то этот множитель в числителе можно вынести за знак суммирования. В итоге получим

$$p(\mathcal{A}^{(+)} | \mathcal{A}^{(0)}) = p(\mathcal{A}^{(+)} | \mathcal{A}^{(0)}, \mathcal{A}_S^{(-)}).$$

Итак, лемма доказана.

В качестве простейшего примера использования леммы V.1 рассмотрим однородную немарковскую последовательность или однородную непростую марковскую цепь на двух состояниях: I и J .

Пусть нам известно, что

$$p(a_I^{(h)} | a_J^{(h-1)}, a_I^{(h-10)}) = \frac{1}{8}$$

и

$$p(a_I^{(h)} | a_J^{(h-1)}, a_J^{(h-10)}) = \frac{1}{8}.$$

Тогда, согласно лемме V.1, без всяких вычислений заключаем, что $p(a_I | a_J) = \frac{1}{8}$.

V.3.1. Внедрение между исходами смежных испытаний

Предположим, что изложенные предпосылки относительно процесса сгущения выполнены. Рассмотрим теперь конкретную схему, которую назовем «внедрением между исходами смежных испытаний».

Пусть между исходами двух смежных испытаний¹ $u^{(h-1)}, v^{(h)}$, ($u, v \in \mathcal{S}$) начальной последовательности может размещаться l вторичных исходов I с вероятностью $p_I(u^{(h-1)}, v^{(h)})$.

Относительно этих вероятностей внедрения мы принимаем две аксиомы.

а). Аксиома независимости внедрения новых зерен между различными парами. Предполагается, что вероятности $p_I(u^{(h-1)}, v^{(h)})$ постоянны и не зависят от того, как были заполнены промежутки между любыми другими парами исходов смежных испытаний. Таким образом, условная вероятность того, скажем, что между первым и вторым зернами, которые оказались сложенными кварцем, появилось вторичное зерно альбита и что между девятым и десятым зернами, которые оказались зернами ортоклаза, возникло

¹ Здесь и в дальнейшем используется сокращение: вместо $a_u^{(h)}$ пишется $u^{(h)}$ и соответственно вместо $a_v^{(h)}$ пишем $v^{(h)}$.

два вторичных зерна альбита, равна произведению условных вероятностей этих событий, т. е. для данного примера она равна произведению $P_1(a_Q^{(1)}, a_Q^{(2)}) \cdot P_2(a_{Og}^{(g)}, a_{Og}^{(10)})$.

б). Аксиома однородности вторичного процесса. Это означает, что вероятности внедрения не зависят от h и могут быть обозначены как $p_i(u, v)$.

Такая аксиоматика введена на основе тех представлений, которые возникли при изучении начальных стадий метасоматического преобразования гранитов (Романова, 1976). При этом создается впечатление, что вторичные зерна возникли за счет инфильтрации флюида равномерно, по фронту, проходящему через породу. Это отражено аксиомой б). В итоге деятельности этого флюида возникли зерна альбита. Миграция флюида проходила по межзерновым контактам, которые открывались под действием тектонических напряжений, растягивавших соответствующую часть массива. Изучение деталей минералообразования во вскрытых каналах показывает, что для образования вторичного зерна необходимо, но недостаточно наличия канала. Это определяется чисто локальными причинами.

Таким образом, если зерно альбита возникло в одном канале, то в первом приближении это не влияло на вероятность появления зерен альбита во всех других каналах, что обосновывает аксиому а). Схема подобного внедрения показана на рис. V.2.

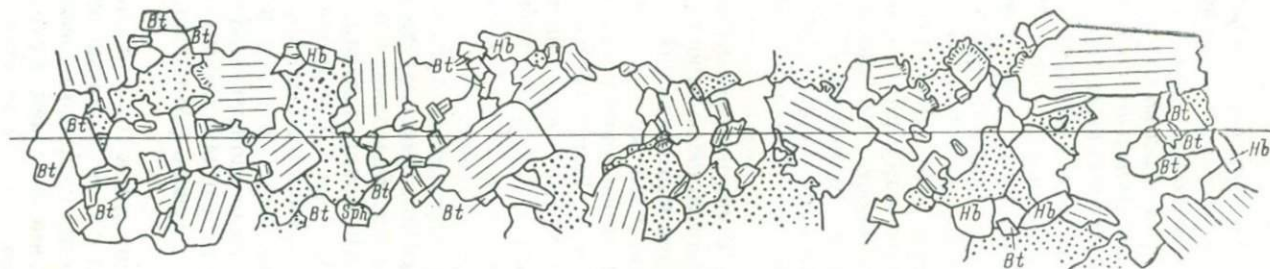
Детально подобные процессы будут изучены во второй части этой книги.

Пример V.8. В породах массива Арга-Ыннх-Хая, согласно исследованиям М. А. Романовой (1978), под микроскопом можно видеть мелкие вторичные зерна альбита на контакте между крупными индивидами плагиоклаза (по составу почти чистого альбита). Такие соотношения могут вытекать из изложенной аксиоматики; они подтверждаются статистическими тестами, приведенными в цитированной работе.

С другой стороны, в докембрийских гранитах Бразилии (лекция Д. Гуимареса) нами многократно наблюдались каймы калиевого полевого шпата вокруг индивидов плагиоклаза. Этот случай не отвечает изложенной аксиоматике (нарушается аксиома а)). Нарушение вызвано тем, что каждое пересечение каймы с прямой, проходящей через окаймленное зерно плагиоклаза, дает при подсчете два зерна Og .

Таким образом, если внутри между парой зерен $u^{(h-1)}$ и $Ab^{(h)}$ возникло зерно Og , то оно возникает и внутри пары $Ab^{(h)}$ $v^{(h+1)}$. Здесь предположение о независимости явно нарушается. Такой процесс описывается схемой сгущения, а не рассматриваемой схемой внедрения между исходами смежных испытаний.²

² Вопрос происхождения соответствующих гранитов Бразилии здесь не рассматривается, пример приведен только для иллюстрации специфических соотношений.



123 233 123 323 232 213 233 221 333 212 212 23

Рис. V.2. Последовательность образования зерен в граните из долины Йосемити (западный склон Сьерра-Невады, Калифорния; зарисовка Ю. А. Высоцкого; материалы А. Б. Вистелиуса и Дж. Хэрбо, 1975).

Гранит претерпел переработку со вторичным образованием плагиоклаза. 1 — калиевый полевой шпат; 2 — кварц; 3 — плагиоклаз; 4 — мirmekиты; 5 — биотит; 6 — роговая обманка; 7 — сфен. Цифры *внизу* отвечают последовательности минералов вдоль линии.

Марковская структура последовательности, построенной по схеме внедрения между исходами смежных испытаний, характеризуется следующей леммой.

Лемма V.2. Исходная стационарная простая марковская цепь, преобразованная по схеме внедрения состояния I между исходами смежных испытаний, дает однородную случайную последовательность, в которой переход через каждое состояние из \mathcal{S}' является частным марковским. Последовательность будет обратимой, если $p_I(u, v) = p_I(v, u)$ и исходная цепь была обратимой.

Доказательство можно провести на основе интерпретации аксиоматики, положенной в основу рассматриваемой статистической схемы. Окончательная последовательность представляет наложение на исходную марковскую цепь процесса сгущения. Поскольку оба эти процесса однородны, то этим же свойством обладает и окончательная последовательность. Если $p_I(u, v) = p_I(v, u)$, то оба процесса, положенные в основание схемы, обратимы, обратной будет и окончательная последовательность. Из аксиоматики также следует, что переход через любое состояние из \mathcal{S}' в окончательной последовательности будет частным марковским, т. е. появление в любом испытании состояния из \mathcal{S}' является восстанавливающим событием.

Действительно, пусть $a^{(h)} = J$, $J \neq I$. Тогда все дальнейшее вероятностное течение исходной последовательности (простой марковской цепи) определяется этим событием полностью. Прохождение же вторичного процесса (заполнение промежутков между соседними исходами в марковской цепи) определяется однозначно тем, как реализовался первичный процесс. Таким образом, реализация $a^{(h)}$ будет восстанавливающим событием для окончательной последовательности. Тем самым переход через J в окончательной последовательности, как указывалось в гл. IV на с. 287, будет обладать марковским свойством.

Пример V.9. Изучение белых гранитов Арга-Бинных Хая (Якутия) показало, что есть основания предполагать происхождение структур этих пород за счет преобразования простых марковских цепей по схеме внедрения Ab между исходами смежных испытаний (Романова, 1976, 1978). Лемма V.2 утверждает, что если единственным внедрившимся минералом был Ab , то мы должны наблюдать однородную последовательность, в которой переходы через Q и Og обладают марковским свойством. Конкретное исследование показало, что именно такую картину мы наблюдаем в абсолютно доминирующем числе случаев (Романова, 1976, 1978).

Рассмотренная схема внедрения приводит к простым марковским переходам через все состояния из \mathcal{S}' не зависимо от того, как выбраны вероятности $p_I(u, v)$. Что же касается перехода через I , то его марковская структура, а вместе с ней и структура наблюдаемой последовательности в целом зависит от конкретного задания этих вероятностей. Чтобы задать конкретную схему внед-

рения, надо указать все вероятности $p_i(u, v)$, т. е. задать распределения по длине l -серий со всевозможными фиксированными концами $u, v \in \mathcal{S}'$ (о распределении серий по длине с фиксированными концами см. гл. IV, теорема IV.1).

Простейший случай схемы внедрения — когда между смежными исходами испытаний в первичной цепи может внедриться лишь один новый исход I , либо ни одного нового исхода. В этом случае $p_i(u, v) = 0$ для всех $u, v \in \mathcal{S}$, если $l > 1$. Даже в этом простейшем случае могут возникать разнообразные марковские структуры в зависимости от того, как выбирать вероятности $p_i(u, v)$.

Рассмотрим две крайние возможности.

В первом варианте предполагается, что дополнительный исход испытания I может возникать лишь между старыми смежными исходами испытаний, в которых оба эти исхода дали состояние I . Такая схема преобразования полностью определяется заданием одной вероятности $p_1(I, I)$. Обращаясь к петрографическим аналогиям, эту схему мы условно назовем «схемой очагового зерна».

Во втором варианте схемы внедрения одного исхода предполагается, что этот исход, дающий состояние I , может внедряться лишь между старыми исходами испытаний, давшими состояния из \mathcal{S}' . Такая схема определяется заданием вероятностей $p_1(u, v)$ для всех $u, v \in \mathcal{S}'$. Пользуясь петрографическими аналогиями, назовем ее схемой «чужеродного зерна».

Пример V.10. Алгоритм для получения реализаций существенно немарковских последовательностей

Марковскую цепь конечного порядка r нетрудно моделировать с помощью таблицы случайных чисел, фиксируя исход r -го последовательного испытания по правилу, учитывающему реализацию предшествующих $r-1$ испытаний. Техника подобных операций хорошо известна и будет показана в гл. VI. Однако таким путем нельзя моделировать существенно немарковскую последовательность, так как невозможно применять бесконечно усложняющиеся правила для идентификации исходов. Покажем, что схема внедрения в ее очаговом варианте приводит к такой реализации.

Теорема V.6. Пусть исходная последовательность является эргодической и стационарной простой цепью Маркова, в которой $0 < p(I|I) < 1$. Пусть реализации этой цепи преобразованы по схеме внедрения так, что между любыми двумя смежными исходами типа $a_i^{(h-1)}$, $a_i^{(h)}$ может внедриться новый исход I с вероятностью $0 < p < 1$. Тогда окончательная последовательность является существенно немарковской.

Доказательство. Допустим, что окончательная последовательность есть марковская цепь некоторого конечного

порядка. Рассмотрим в этой цепи распределение по длине I -серий с фиксированными концами $u, v \in \{\mathcal{S}'\}$. Согласно теореме IV.1 о распределении серий с фиксированными концами в марковской цепи, известно, что, начиная с некоторой длины l , вероятности серий в этом распределении образуют геометрическую прогрессию. Пусть $\pi_l(u, v)$ — вероятность для I -серии длины l с концами u, v из окончательной последовательности. Соответствующую вероятность для исходной цепи будем обозначать $\tau_l(u, v)$. Тогда при достаточно большом n (большем чем неизвестный, но конечный, по предположению, порядок марковости) справедлива пропорция

$$\frac{\pi_{2n+1}(u, v)}{\pi_{2n}(u, v)} = \frac{\pi_{2n}(u, v)}{\pi_{2n-1}(u, v)}. \quad (\text{V. 3. 6})$$

Вычислим теперь вероятности в (V.3.6) и докажем, что равенство между правой и левой частями не может иметь место. Очевидно, что получить I -серию длины $2n+1$ в окончательной последовательности можно лишь на базе уже имевшейся I -серии в исходной последовательности. Эта исходная серия должна иметь длину от $n+1$ до $2n+1$. Более длинные или более короткие I -серии не могут преобразоваться в рамках рассматриваемой схемы в серию длины $2n+1$. Если исходная серия имела длину l , причем $n+1 \leq l \leq 2n+1$, то внутри нее должно образоваться по очаговой схеме $2n+1-l$ вторичных исходов I . Условная вероятность образования в окончательной последовательности I -серии с концами u и v длины $2n+1$ совпадает с вероятностью образования такой серии, взятой при следующем условии.

Допустим, что I -серия началась с некоторого фиксированного места, непосредственно вслед за появлением состояния $u \in \mathcal{S}'$, имевшим в исходной цепи номер h . В силу стационарности исходной цепи и однородности процесса внедрения вероятность соответствующего события не будет зависеть ни от h , ни от номера, отвечающего началу серии в окончательной последовательности.

В силу леммы V.1 вероятность образования серии, начинающейся в данном месте, совпадает с вероятностью ее появления в случайной последовательности вообще. Поэтому мы можем вычислить вероятность $\pi_{2n+1}(u, v)$ как вероятность следующего события. В начальной цепи, вслед за появлением исхода $a_u^{(h)}$, образуется I -серия длины l ($n+1 \leq l \leq 2n+1$), а затем в этой серии по очаговой схеме формируется еще $2n+1-l$ вторичных исходов I .

Вероятность появления первичной серии длины l равна $\tau_l(u, v)$. Вероятность добавления в эту серию новых i исходов по очаговой схеме будет $C_{l-1}^i p^i q^{l-i-1}$ (здесь $q = 1 - p$), так как в серию можно добавить i дополнительных исходов различными способами в $l-1$ -й промежуток между исходами первичных испытаний. Поскольку

$$l + i = 2n + 1,$$

получаем

$$\begin{aligned} \tau_{2n+1}(u, v) &= C_{2n}^0 p^0 q^{2n} \tau_{2n+1}(u, v) + C_{2n-1}^1 p^1 q^{2n-2} \tau_{2n}(u, v) + \dots \\ &\dots + C_{n+1}^{n-1} p^{n-1} q^2 \tau_{n+2}(u, v) + C_n^n p^n q^0 \tau_{n+1}(u, v) = \\ &= \sum_{i=0}^n C_{2n-i}^i p^i q^{2n-2i} \tau_{2n+1-i}(u, v). \end{aligned} \quad (\text{V. 3. 7})$$

I -серии длины $2n$ в окончательной последовательности могут образовываться лишь из I -серий исходной последовательности с длиной от $n+1$ до $2n$; I -серии длины $2n-1$ в окончательной последовательности могут образовываться лишь из серий исходной последовательности с длиной от n до $2n-1$.

Учитывая это, после вычислений, аналогичных приведем к (V. 3. 7), получаем

$$\tau_{2n}(u, v) = \sum_{i=0}^{n-1} C_{2n-1-i}^i p^i q^{2n-2i-1} \tau_{2n-i}(u, v) \quad (\text{V. 3. 8})$$

и

$$\tau_{2n-1}(u, v) = \sum_{i=0}^{n-1} C_{2n-2-i}^i p^i q^{2n-2i-2} \tau_{2n-1-i}(u, v). \quad (\text{V. 3. 9})$$

Подставляя (V. 3. 7), (V. 3. 8) и (V. 3. 9) в (V. 3. 6), имеем

$$\begin{aligned} &\sum_{\substack{i=0, 1, \dots, n \\ j=0, 1, \dots, n-1}} C_{2n-i}^i C_{2n-2-j}^j p^{i+j} q^{4n-2i-2j-2} \tau_{2n+1-i}(u, v) \tau_{2n-1-j}(u, v) = \\ &= \sum_{\substack{i=0, 1, \dots, n-1 \\ j=0, 1, \dots, n-1}} C_{2n-i}^i C_{2n-1-j}^j p^{i+j} q^{4n-2i-2j-2} \tau_{2n-i}(u, v) \tau_{2n-j}(u, v). \end{aligned} \quad (\text{V. 3. 10})$$

Учитывая, согласно теореме IV. I, что в силу простой марковости исходной цепи

$$\tau_m(u, v) = p(I|I) [p(I|I)]^{m-1} \quad (m = 1, 2, \dots),$$

где $p(\bar{I}|I)$, $p(I|I)$ — вероятности, отнесенные к исходной цепи, после некоторых преобразований из (V. 3. 10) получаем следующее равенство, равносильное (V. 3. 6):

$$p^{2n-1} [p(I|I)]^2 [p(I|I)]^{2n-1} = 0. \quad (\text{V. 3. 11})$$

Из (V. 3. 11) следует справедливость по крайней мере одного из равенств

$$p = 0, \quad p(I|I) = 0, \quad p(I|I) = 0.$$

Первое из этих равенств означает, что внедрения по очаговой схеме не происходило, а второе и третье равенства невозможны по условиям теоремы. Таким образом, доказана существенная немарковость, возникающая при внедрении одного дополнительного испытания по схеме очагового зерна.

Согласно теореме V.6, алгоритм для получения реализации существенно немарковской последовательности может быть осуществлен, например, следующим путем. По таблице случайных чисел строим реализацию последовательности независимых испытаний с двумя состояниями I и J и вероятностями $p(I) = p(J) = 1/2$. Для каждой пары смежных исходов $a_I^{(h-1)} a_I^{(h)}$ из этой реализации по таблице случайных чисел разыгрываем, внедряется ли между этими исходами еще один исход с состоянием I . Вероятность такого внедрения снова примем равной $1/2$. Перенумеруем теперь все исходы заново. Тогда получим реализацию существенно немарковской последовательности. Действительно, теорема V.6 не ограничивает выбора значений параметров и числа состояний, а последовательность Бернулли с постоянными вероятностями исходов является частным случаем эргодической и стационарной простой цепи Маркова.

Отметим, что рассмотренная в теореме V.6 схема очагового зерна, по-видимому, реализуется в белых гранитах Арга-Бннх-Хая с $I=Ab$ в тех образцах, где статистические тесты позволяют сделать заключение о том, что порядок марковости был выше второго (Романова, 1978). Возможно, что эти последовательности являются существенно немарковскими.

Как видим, существенно немарковские последовательности могут возникать при некоторых геологических явлениях в самых обычных условиях. В связи с этим возникают большие трудности для уверенной идентификации марковского характера изучаемых последовательностей. Обычные статистические методы описания и классификации марковских цепей, излагаемые нами в гл. VI, в принципе для изучения существенно немарковских последовательностей не пригодны. Опознание таких последовательностей связано с большими трудностями. Действительно, с одной стороны, существенная немарковость представляет наиболее сильное нарушение марковского свойства. С другой стороны, для того чтобы отличить такую последовательность от простой цепи, при малом значении параметра p в теореме V.6, требуется, по-видимому, чрезвычайно большое число наблюдений.

Статистические методы, пригодные для изучения существенно немарковских последовательностей с точки зрения отделения таких последовательностей на основании наблюдений от марковских, в настоящее время, видимо, не разработаны. Точно так же, по-видимому, нет статистических методов для оценивания всех переходных вероятностей в существенно немарковской последовательности. Однако вероятностная основа для постановки этих задач имеется (Harris, 1955).

В настоящее время при геологических исследованиях в случае подозрения, что изучаемая последовательность является существенно немарковской, приходится строить выводы на основании статистики, не опирающейся на марковскую специфику последо-

вательности. Этот сложный вопрос в практической плоскости будет рассмотрен во второй части этой книги.

Рассмотрим теперь схему чужеродного зерна, в частности тот случай, когда $p_1(u, v) > 0$ лишь при $u, v \in \mathcal{S}'$.

Теорема V.7. Пусть исходная последовательность является эргодической стационарной марковской цепью первого порядка. При этом схема внедрения, примененная к реализации исходной цепи, предполагает, что между любыми двумя смежными исходами $a_u^{(h)}$ и $a_v^{(h+1)}$ ($u, v \in \mathcal{S}'$) исходной цепи может внедриться с вероятностью $0 < p < 1$ один новый исход I . Тогда окончательная (наблюдаемая) последовательность оказывается эргодической однородной цепью Маркова второго порядка. Эта цепь имеет следующую марковскую структуру частных переходов:

а) переходы первого порядка через каждое состояние из \mathcal{S}' марковские;

б) переходы второго порядка через $u^{(h-1)}I^{(h)}$ для всех $u \in \mathcal{S}$ также марковские;

в) переход через $I^{(h-1)}I^{(h)}$, имея второй порядок, обладает специальным свойством

$$p(a_I^{(h+1)} | a_I^{(h)} a_I^{(h-1)} A) = p_1(a_I^{(h+1)} | a_I^{(h)}) \quad (A \in \mathcal{A}_{h-2}),$$

где p_1 — вероятность, отнесенная к исходной цепи.

Доказательство. Утверждения теоремы, касающиеся однородности и частных марковских переходов из \mathcal{S}' , являются следствиями из леммы V.2. Для установления марковости второго порядка у окончательной последовательности остается доказать, что

$$\begin{aligned} & p(B | a_I^{(h)}, a_J^{(h-1)}, A) = \\ & = p(B | a_I^{(h)}, a_J^{(h-1)}) (B \in \mathcal{B}_{h+1}, A \in \mathcal{A}_{h-2} \text{ при всех } J \in \mathcal{S}') \end{aligned} \quad (\text{V.3.12})$$

и

$$p(B | a_I^{(h)}, a_J^{(h-1)}, A) = p(B | a_I^{(h)}, a_J^{(h-1)}). \quad (\text{V.3.13})$$

Перепишем левую часть равенства (V.3.12) в виде

$$\frac{p(A, a_J^{(h-1)} a_I^{(h)} B)}{p(A, a_J^{(h-1)}, a_I^{(h)})} = \frac{p(B, a_I^{(h)} | a_J^{(h-1)}, A)}{p(a_I^{(h)} | a_J^{(h-1)}, A)}. \quad (\text{V.3.14})$$

Согласно лемме V.2, событие $a_J^{(h-1)}$ является восстанавливающим для окончательной последовательности. Поэтому в числителе и знаменателе (V.3.14) можно отбросить событие из прошлого A .

Тогда

$$\frac{p(a_I^{(h)}, B | a_J^{(h-1)})}{p(a_I^{(h)} | a_J^{(h-1)})} = p(B | a_I^{(h)}, a_J^{(h-1)}).$$

Таким образом, (V.3.12) доказано.

Переходим к рассмотрению (V.3.13). Исход испытания $a_I^{(h)}$ в (V.3.13) принадлежит начальной цепи. Действительно, ему пред-

шествовал исход $a_I^{(h-1)}$. Так как мы имеем дело со схемой чужеродного зерна, то состояние I на h -м месте должно быть первичным. Поэтому в соответствии с рассуждениями леммы V.2 прошлое до этого испытания (т. е. в данном случае $a_I^{(h-1)}$, A) не влияет на вероятность B . Этим доказано (V.3.13).

Применение леммы V.1 показывает, что в этом случае не только выполняется равенство (V.3.13), но и специальное равенство

$$p(a_I | a_I^{(h)}, a_I^{(h-1)}, A) = p_1(a_I | a_I^{(h)}).$$

Этим завершается доказательство теоремы V.7. Можно показать, что те же результаты теоремы V.7 имеют место и в случае внедрения нового состояния I , не принадлежавшего множеству состояний \mathcal{S} исходной цепи.

Пример V.11. Допустим, что исходная последовательность из теоремы V.7 — это последовательность зерен, пересеченных прямой в шлифе первично магматического гранита, фигурировавшего во многих наших предыдущих примерах, на множестве состояний $\mathcal{S} = \{Og, Q, Ab\}$. Допустим далее, что реализация этой последовательности преобразуется по общей схеме внедрения (в ее варианте) чужеродного зерна с возможностью образования лишь одного вторичного зерна Ab . Тогда наблюдаемая последовательность, согласно теореме V.7, должна иметь чрезвычайно специфические свойства а), б) и в) из теоремы V.7. Применительно к исследуемой породе эти свойства будут такими.

а). Переходы как через Q , так и через Og будут марковскими. Иными словами,

$$p(B | a_Q^{(h)}, A) = p(B | a_Q^{(h)}) \quad (B \in \mathfrak{Z}_{h+1}, A \in \mathfrak{A}_{h-1})$$

и

$$p(B | a_{Og}^{(h)}, A) = p(B | a_{Og}^{(h)}).$$

б). Переходы второго порядка через $Q^{h-1}Ab^{(h)}$ и $Og^{(h-1)}Ab^{(h)}$ также являются марковскими. Иными словами,

$$p(B | a_{Ab}^{(h)}, a_Q^{(h-1)}, A') = p(B | a_{Ab}^{(h)} a_Q^{(h-1)}) \quad (A' \in \mathfrak{A}_{h-2})$$

и

$$p(B | a_{Ab}^{(h)}, a_{Og}^{(h-1)}, A') = p(B | a_{Ab}^{(h)}, a_{Og}^{(h-1)}).$$

в). Для перехода через $Ab^{(h-1)}$, $Ab^{(h)}$ должно выполняться равенство

$$p(B | a_{Ab}^{(h)}, a_{Ab}^{(h-1)}, \mathfrak{A}_{h-2}) = p(B | a_{Ab}^{(h)}).$$

Если из изученных образцов белых гранитов Арга-Ынних-Хая (Романова, 1976) выделить образцы, в которых последовательности зерен Og , Q и Ab были статистически не отличимы от цепей Маркова второго порядка, то окажется, что среди этих образцов подавляющее большинство будет удовлетворять соотношениям

б) и в) настоящего примера. Это показывает, что метасоматическое преобразование отмеченных пород, по-видимому, протекало по схеме чужеродного зерна.

Варианты чужеродного и очагового зерна в схеме внедрения приводят после укрупнения к последовательностям с резко различающимися свойствами. До укрупнения матрицы переходных вероятностей, получающиеся в результате приложения этих схем, могут быть очень близкими.

Пример V.12. Рассмотрим две матрицы переходных вероятностей

$$P_1 = \begin{matrix} & \begin{matrix} x_1 & x_2 & y \end{matrix} \\ \begin{matrix} x_1 \\ x_2 \\ y \end{matrix} & \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \\ \dots & \dots & \dots \\ 1/2 & 1/4 & 1/4 \end{pmatrix} \end{matrix}, \quad P_2 = \begin{matrix} & \begin{matrix} x_1 & x_2 & y \end{matrix} \\ \begin{matrix} x_1 \\ x_2 \\ y \end{matrix} & \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1 & 0 & 0 \\ \dots & \dots & \dots \\ 1/2 & 0 & 1/2 \end{pmatrix} \end{matrix},$$

укрупняемые так, как показано пунктиром. Если допустить, что до укрупнения этим матрицам отвечали стационарные простые цепи Маркова, то мы обнаружим большое сходство. А именно: обеим цепям отвечает одно и то же начальное и стационарное распределение $p(x_1)=4/9$, $p(x_2)=1/9$, $p(y)=4/9$. Укрупнение производится одинаково, переходные вероятности близки. Однако в результате укрупнения в первом случае (с матрицей P_1) возникает марковская цепь второго порядка, в то время как укрупнение во второй последовательности приводит к бесконечному порядку марковости. Доказательство этого, основанное на вычислении переходных вероятностей, было бы очень громоздким. На основе теорем V.5. и V.6 может быть получено следующее краткое доказательство.

Рассмотрим исходную последовательность Бернулли на двух состояниях $\{x, y\}$ с вероятностями $p(x)=1/2$, $p(y)=1/2$. Преобразуем реализации этой последовательности по схеме внедрения одного исхода, дающего состояние x , полагая, что это внедрение происходит по схеме чужеродного зерна с вероятностью $p=1/2$. Тогда получим случайную последовательность с матрицей переходных вероятностей P_1 , где первичное состояние x обозначено как x_1 , а вторичное — как x_2 . Применим к той же начальной последовательности преобразование внедрения x по схеме очагового зерна с вероятностью $p=1/2$. Это даст матрицу P_2 , если снова обозначить первичное состояние x как x_1 , а вторичное — как x_2 .

Применим теперь к обеим последовательностям одно и то же укрупнение G :

$$x_1 \cup x_2 = X, \quad y = Y.$$

Тогда возникнут две последовательности на $\{X, Y\}$. Применяя теоремы V.6 и V.7 к результатам укрупнения P_1 и P_2 , получаем для первой последовательности марковость второго порядка, а для второй — существенную немарковость.

Теорема V.8. Пусть исходная последовательность является эргодической и стационарной простой цепью Маркова на множестве состояний \mathcal{S} , не содержащем состояние I . В соответствии с аксиоматикой схемы сгущения (см. с. 321, 322) в каждую реализацию этой цепи может добавляться между двумя смежными старыми исходами серия, образованная новыми элементами I . Распределение серий I по длине произвольное с единственным ограничением, что оно не является усеченным (т. е. $p_l(I) > 0$ при всех $l=1, 2, \dots$). Тогда окончательная последовательность на множестве состояний $\mathcal{S}' = \mathcal{S} \cup I$ будет существенно немарковской.

Доказательство. Рассмотрим вероятность

$$p(\mathcal{A}_h | I^{(h-1)}, \dots, I^{(h-r)}, y^{(h-r-1)}), \quad (\text{V. 3. 15})$$

где $y \in \mathcal{S}$, а \mathcal{A}_h событие: « I серия на местах $h-r, h-r+1, \dots, h-1, \dots$ завершится состоянием $z \in \mathcal{S}$ ». Тогда вероятность (V.3.15) совпадает с $p(z|y)$, которая относится к исходной цепи. В то же время

$$p(\mathcal{A}_h | I^{(h-r)}, \dots, I^{(h-r)}) = p(z), \quad (\text{V. 3. 16})$$

где $p(z)$ — безусловная вероятность из исходной цепи. Поскольку исходная последовательность, по предположению, не является последовательностью Бернулли, найдутся такие $y, z \in \mathcal{S}$, что $p(z|y) \neq p(z)$, и тем самым (V.3.15) не совпадает с вероятностью (V.3.16). Поскольку r в (V.3.15) — произвольное целое число, несовпадение вероятностей (V.3.15) и (V.3.16) означает существенную немарковость окончательной последовательности.

Тот же результат имел бы место, если бы исходная последовательность была бы цепью Маркова k -порядка (внедрение серий нового состояния приводит к существенной немарковости). В качестве примера для применения теоремы V.8 возьмем последовательность зерен в оливиновом габбро, представляющую простую стационарную марковскую цепь на состояниях оливин, пироксен и плагиоклаз. Пусть при этом некоторые зерна оливина частично замещаются с одной стороны агрегатом зерен серпентина, так что старое зерно оливина сохраняется в последовательности. Тогда последовательность на состояниях оливин, пироксен, плагиоклаз и серпентин будет существенно немарковской.

Следует подчеркнуть, что внедрение серий любого из состояний $x \in \mathcal{S}$ не обязательно нарушает марковость последовательности.

V.3.2 Схема увеличения серий

Теорема V.9. Пусть [исходная последовательность есть эргодическая и стационарная простая цепь Маркова на множестве состояний \mathcal{S} , содержащем состояние I . Пусть, далее, длина каждой I -серии исходной цепи может либо увеличиться за счет присоединения нового исхода I к серии на один элемент с постоянной ве-

роятностью π , либо сохранить прежнюю длину с вероятностью $1-\pi$. При этом предполагается, что вероятность изменения данной I -серии не зависит от того, как будут видоизменены остальные I -серии. Тогда окончательная последовательность будет однородной марковской цепью второго порядка с марковскими переходами первого порядка через любое состояние из $\mathcal{S}' = \mathcal{S} \setminus \{I\}$.

Подчеркнем, что схема увеличения серий не является частным случаем схемы очагового зерна. Обратное соотношение также не имеет места. Так, в схеме очагового зерна из I -серии длины l исходной цепи может образоваться серия длины $2l-1$, тогда как в настоящей схеме может возникнуть серия длины только $l+1$. При серии длины 2 в исходной реализации очаговая схема допускает только один вариант размещения вторичного зерна, в то время как схема удлинения серий допускает три таких варианта.

Д о к а з а т е л ь с т в о. Пусть I_1 — состояние в исходной цепи, а I_2 — состояние, удлиняющее исходную I -серию. Рассмотрим случайную последовательность на множестве состояний $\{\Sigma\} = \{I_1, I_2, J, \dots, L\}$. Условимся размещать в этой случайной последовательности I_2 всегда перед началом I -серии. Мы можем это сделать, так как вид окончательной последовательности, возникшей после укрупнения $I = I_1 \cup I_2$, не зависит от того, как мы размещаем I_2 по отношению к I_1 -серии — до серии, после нее или внутри серии. При таком предположении относительно I_2 последовательность на множестве состояний Σ будет простой марковской цепью. Действительно, легко показать, что

$$\begin{aligned} p(a_{I_1}^{(h)} | a_{I_1}^{(h-1)}, A) &= p_1(a_{I_1}^{(h)} | a_{I_1}^{(h-1)}) \text{ для всех } J \neq I_2 (A \in \mathfrak{A}_{h-2}), \\ p(a_{I_2}^{(h)} | a_{I_1}^{(h-1)}, A) &= 0, \quad p(a_L^{(h)} | a_{I_1}^{(h-1)}, A) = p_1(a_L^{(h)} | a_{I_1}^{(h-1)}) \\ &\text{для всех } L, J \in \mathcal{S}', \quad (V. 3.17) \\ p(a_{I_2}^{(h)} | a_{I_2}^{(h-1)}, A) &= \pi p_1(a_{I_2}^{(h)} | a_{I_2}^{(h-1)}), \quad p(a_{I_1}^{(h)} | a_{I_2}^{(h-1)}, A) = \\ &= p_1(a_{I_1}^{(h)} | a_{I_2}^{(h-1)}) \text{ для всех } J \in \mathcal{S}', \\ p(a_{I_1}^{(h)} | a_{I_2}^{(h-1)}, A) &= 1, \end{aligned}$$

где переходные вероятности p_1 в правых частях (V. 3.17), за исключением второго равенства, относятся к исходной цепи. Таким образом, вероятности в (V. 3.17) слева не зависят от A . Применяя лемму V.1 к равенствам (V. 3.17), получаем

$$\begin{aligned} p(a_{I_1}^{(h)} | a_{I_1}^{(h-1)}, A) &= p(a_{I_1}^{(h)} | a_{I_1}^{(h-1)}) \text{ для всех } J \neq I_2, \quad p(a_{I_2}^{(h)} | a_{I_1}^{(h-1)}, A) = \\ &= p(a_{I_2}^{(h)} | a_{I_1}^{(h-1)}) = 0, \\ p(a_L^{(h)} | a_{I_1}^{(h-1)}, A) &= p(a_L^{(h)} | a_{I_1}^{(h-1)}) \text{ для всех } L, J \in \mathcal{S}', \quad (V. 3.18) \\ p(a_{I_2}^{(h)} | a_{I_2}^{(h-1)}, A) &= \pi p(a_{I_2}^{(h)} | a_{I_2}^{(h-1)}), \\ p(a_{I_1}^{(h)} | a_{I_2}^{(h-1)}, A) &= (1 - \pi) p(a_{I_1}^{(h)} | a_{I_2}^{(h-1)}). \\ p(a_{I_1}^{(h)} | a_{I_2}^{(h-1)}, A) &= 1. \end{aligned}$$

Равенства (V.3.18) показывают, что переход через любое состояние из Σ является ограниченно марковским, а это (гл. IV, с. 289) влечет простую марковость последовательности на множестве Σ .

Покажем теперь, что для укрупненной случайной последовательности ($I = I_1 \cup I_2$) справедливы следующие равенства:

$$\begin{aligned} p(a_I^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}, A') &= p(a_I^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}), p(a_J^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}, A') = \\ &= p(a_J^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}), \\ p(a_I^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}, A') &= p(a_I^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}), \\ p(a_J^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}, A') &= p(a_J^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}), \\ p(a_I^{(h)} | a_I^{(h-1)}, A) &= p(a_I^{(h)} | a_I^{(h-1)}), p(a_I^{(h)} | a_I^{(h-1)}, A) = \\ &= p(a_I^{(h)} | a_I^{(h-1)}) \text{ для всех } J, L \in \{\mathcal{S}'\}, A \in \mathfrak{A}_{h-2}, A' \in \mathfrak{A}_{h-3}. \end{aligned} \quad (\text{V. 3.19})$$

Равенства (V.3.19) означают, что все переходы второго порядка в укрупненной последовательности обладают ограниченно марковским свойством. Из этого, как отмечалось, следует, что укрупненная последовательность есть марковская цепь второго порядка. Два последних равенства в (V.3.19) кроме того показывают, что переходы первого порядка через состояния из \mathcal{S}' являются марковскими.

Все равенства из (V.3.19) доказываются непосредственным вычислением. Наиболее трудоемок расчет, отвечающий первому равенству (V.3.19). Этот расчет дается ниже.

Как известно, событие A' можно представить в виде конечной последовательности результатов испытаний с номерами $0, 1, \dots, h-3$ (см. гл. IV). Переходя от условных вероятностей к безусловным и учитывая, что $I = I_1 \cup I_2$, получаем

$$\begin{aligned} p(a_I^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}, A') &= \\ p(A', a_{I_1}^{(h-2)}, a_{I_1}^{(h-1)}, a_{I_1}^{(h)}) &+ p(A', a_{I_1}^{(h-2)}, a_{I_1}^{(h-1)}, a_{I_2}^{(h)}) + p(A', a_{I_1}^{(h-2)}, a_{I_2}^{(h-1)}, a_{I_1}^{(h)}) + \\ + p(A', a_{I_1}^{(h-2)}, a_{I_2}^{(h-1)}, a_{I_2}^{(h)}) &+ p(A', a_{I_2}^{(h-2)}, a_{I_1}^{(h-1)}, a_{I_1}^{(h)}) + p(A', a_{I_2}^{(h-2)}, a_{I_1}^{(h-1)}, a_{I_2}^{(h)}) + \\ + p(A', a_{I_2}^{(h-2)}, a_{I_2}^{(h-1)}, a_{I_1}^{(h)}) &+ p(A', a_{I_2}^{(h-2)}, a_{I_2}^{(h-1)}, a_{I_2}^{(h)}) \\ = \frac{p(A', a_{I_1}^{(h-2)}, a_{I_1}^{(h-1)}, a_{I_1}^{(h)}) &+ p(A', a_{I_1}^{(h-2)}, a_{I_2}^{(h-1)}, a_{I_1}^{(h)}) + \\ + p(A', a_{I_2}^{(h-2)}, a_{I_1}^{(h-1)}, a_{I_1}^{(h)}) &+ p(A', a_{I_2}^{(h-2)}, a_{I_2}^{(h-1)}, a_{I_1}^{(h)})}{p(A', a_{I_1}^{(h-2)}, a_{I_1}^{(h-1)}, a_{I_1}^{(h)}) + p(A', a_{I_1}^{(h-2)}, a_{I_2}^{(h-1)}, a_{I_1}^{(h)}) + \\ + p(A', a_{I_2}^{(h-2)}, a_{I_1}^{(h-1)}, a_{I_1}^{(h)}) &+ p(A', a_{I_2}^{(h-2)}, a_{I_2}^{(h-1)}, a_{I_1}^{(h)})} \end{aligned} \quad (\text{V. 3.20})$$

Учтем теперь, что последовательность на множестве состояний Σ простая марковская, что эта цепь однородна и, наконец, что для этой цепи

$$p(a_{I_1} | a_{I_2}) = 1, \quad p(a_{I_2} | a_{I_1}) = 0 \quad \text{и} \quad p(a_{I_2} | a_{I_2}) = 0.$$

Тогда вместо правой части (V.3.20) получим

$$\begin{aligned} \frac{p(A') p(a_{I_1} | A') p(a_{I_1} | a_{I_1}) p(a_{I_1} | a_{I_1}) + p(A') p(a_{I_2} | A') \cdot 1 \cdot p(a_{I_1} | a_{I_1})}{p(A') p(a_{I_1} | A') p(a_{I_1} | a_{I_1}) + p(A') p(a_{I_2} | A') \cdot 1} = \\ = p(a_{I_1} | a_{I_1}). \end{aligned}$$

В силу леммы V.1 заключаем, что

$$p(a_I^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}, A') = p(a_I^{(h)} | a_I^{(h-1)}, a_I^{(h-2)}). \quad (\text{V.3.21})$$

Очевидно, что вероятность в (V.3.21) справа не совпадает с $p(a_I | a_I)$. Таким образом, цепь не может быть простой марковской. Итак, теорема V.9 доказана. Аналогично можно доказать и несколько более общий результат. В том случае, когда с вероятностью π осуществляется приращение I серии на l элементов (при сохранении всех остальных предположений теоремы V.9), появляется марковская цепь порядка $1+l$ с марковскими переходами первого порядка через все состояния из \mathcal{S}' и с марковским переходом $l+1$ -го порядка через $\underbrace{a_I^{(h-1)}, \dots, a_I^{(1)}}_h$.

Из рассмотренной стохастической схемы следует, что простая марковская цепь легко преобразуется в цепь Маркова порядка l , если к I -сериям в ней может прибавляться по $l-1$ элементов.

Пример V.13. До сих пор мы интерпретировали рассматриваемые стохастические схемы главным образом на материале вторичного преобразования первично магматических пород. Посмотрим, насколько реально для такого типа задач может выполняться аксиоматика только что изученной схемы увеличения серий. Имеется предположение, что I серия увеличивается с постоянной вероятностью π независимо от длины этой серии. Таким образом, если мы рассмотрим серию из n зерен кварца и наряду с ней серию из одного зерна кварца, то, согласно указанной схеме, можно будет предположить, что вероятность превращения серии в n зерен в серию из $n+1$ -го зерна и вероятность превращения серии из n -го зерна в серию из $n+1$ зерен одна и та же. Реализация такой схемы неправдоподобна. Однако имеются случаи, когда аксиоматика такого типа оказывается убедительной. Это случай, когда в исходной последовательности все серии имеют одинаковую длину. Тогда независимость вероятности π изменения длины серии от этой длины будет очевидной. Указанный случай может реализоваться, например, если все I -серии имеют длину 1. Это произойдет в том и только в том случае, когда в исходной цепи $p(a_I) > 0$ и $p(a_I | a_I) = 0$.

Практически близкая картина наблюдается при кристаллизации в эвтектической системе, если I минерал кристаллизуется только в эвтектической точке. Некоторые иллюстрации к этому дает эксперимент, опубликованный Д. Н. Ивановым (1975). В таких условиях аксиоматика схемы увеличения серий скорее всего соответствует реальности. Другой вариант, если фактическое увеличение серий осуществляется только с конца (или с двух концов).

V.3.3. Простейшие схемы разрежения

Предположим, что некоторое подмножество состояний $\mathcal{S}_1 \subset \mathcal{S}$ удаляется полностью из реализаций исходной цепи. Таким образом, любое испытание $a_I^{(h)}$ исключается из реализаций, если $I \in \mathcal{S}_1$.

Остающиеся испытания перенумеровываются в естественном порядке. В результате возникает «разряженная» случайная последовательность на множестве состояний $\mathcal{S}' = \mathcal{S} \setminus \mathcal{S}_1$. Для удобства предположим, что состояния исходной цепи обозначены так, что им отвечают матрицы переходных вероятностей P со следующим клеточным строением:

$$P = \begin{matrix} & \mathcal{S}' & \cdot & \mathcal{S}_1 \\ \mathcal{S}' & \left(\begin{array}{c} P_{\mathcal{S}', \mathcal{S}'} \cdot P_{\mathcal{S}', \mathcal{S}_1} \\ \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \end{array} \right) \\ \mathcal{S}_1 & \left(\begin{array}{c} P_{\mathcal{S}_1, \mathcal{S}'} \cdot P_{\mathcal{S}_1, \mathcal{S}_1} \end{array} \right) \end{matrix}, \quad (\text{V. 3. 22})$$

где $P_{\mathcal{S}', \mathcal{S}'}$ — квадратная матрица, отвечающая остающимся состояниям; $P_{\mathcal{S}', \mathcal{S}_1}$ — прямоугольная матрица, отвечающая переходным вероятностям из множества \mathcal{S}' в множество \mathcal{S}_1 ; $P_{\mathcal{S}_1, \mathcal{S}'}$ — прямоугольная матрица, отвечающая переходным вероятностям из множества \mathcal{S}_1 в множество \mathcal{S}' ; $P_{\mathcal{S}_1, \mathcal{S}_1}$ — квадратная матрица, отвечающая переходам между исключаемыми состояниями.

Теорема V.10. Удаление из реализаций исходной эргодической и стационарной простой цепи Маркова некоторого подмножества состояний $\mathcal{S}_1 \subset \mathcal{S}$ с последующей естественной перенумерацией остающихся исходов приводит к однородной простой цепи Маркова на множестве состояний $\mathcal{S}' = \mathcal{S} \setminus \mathcal{S}_1$. Матрица переходных вероятностей P' для новой цепи определяется матричным уравнением

$$P' = P_{\mathcal{S}', \mathcal{S}'} + P_{\mathcal{S}', \mathcal{S}_1} (I - P_{\mathcal{S}_1, \mathcal{S}_1})^{-1} P_{\mathcal{S}_1, \mathcal{S}'}, \quad (\text{V. 3. 23})$$

где I — единичная матрица того же порядка, что и $P_{\mathcal{S}_1, \mathcal{S}_1}$. Вектор стационарного распределения для новой цепи p'_{st} получается в результате приведения к единице остающихся компонент стационарного вектора исходной цепи, принадлежащих множеству остающихся состояний \mathcal{S}' .

Доказательство. При доказательстве части теоремы, касающейся сохранения марковости, можно ограничиться случаем, когда вычеркивается лишь одно состояние I . Действительно, после такого вычеркивания теорему можно применить снова, рассматривая в качестве исходной цепь после удаления состояния, и т. д.

Итак, пусть состояние I удаляется из реализации. События, относящиеся к новой случайной последовательности, будем отмечать *. Достаточно показать, что

$$p(a_J^* \mid a_R^* \text{ (} h-1 \text{)}, A^*) = p(a_J^* \mid a_R^* \text{ (} h-1 \text{)}) \quad (\text{V. 3. 24})$$

для всех $J, R \in \mathcal{S}'$, где событие $A^* \in \mathcal{Q}_{h-2}^*$ — реализация некоторой конечной последовательности испытаний с номерами $0, 1, \dots, h-2$. Как неоднократно отмечалось в этой и предыдущей главе, равенство (V.3.24) влечет справедливость равенства

$$p(B^* \mid a_R^* \text{ (} h-1 \text{)}, A^*) = p(B^* \mid a_R^* \text{ (} h-1 \text{)}) \quad (\text{V. 3. 25})$$

для всех $R \in \mathcal{S}'$, где $B^* \in \mathfrak{B}_h^*$ — событие из будущего, а $A^* \in \mathfrak{A}_{h-2}^*$ — событие из прошлого, отнесенные к окончательной последовательности.

Равенство (V.3.25) для всех $R \in \mathcal{S}'$ означает простоту марковской последовательности, возникшей после удаления состояния I . Чтобы установить (V.3.24), представим левую часть этого равенства в виде

$$\frac{p(A^*, a_R^{*(h-1)}, a_J^{*(h)})}{p(A^*, a_R^{*(h-1)})}. \quad (\text{V. 3. 26})$$

Пусть событию A^* в окончательной последовательности отвечает некоторое событие $A' \in \mathfrak{A}_{l-1}$ исходной последовательности, где l — номер $a_R^{*(h-1)}$ до преобразования, т. е. A^* в окончательной последовательности происходит в том и только в том случае, когда A' происходит в исходной.

Тогда числитель в (V.3.26) будет равен

$$p\left\{A', a_R^{(l)}, \bigcup_{s=0}^{\infty} [(a_I^{(l+1)}, a_I^{(l+2)}, \dots, a_I^{(l+s)}, a_J^{(l+s+1)})]\right\},$$

если условиться, что случаю $S=0$ отвечает событие $a_J^{(l+1)}$ под знаком объединения.

Таким образом, (V.3.26) можно представить как

$$p\left[a_J^{(l+s+1)}, \bigcup_{s=0}^{\infty} (a_I^{(l+s)}, \dots, a_J^{(l+1)} \mid a_R^{(l)}, A')\right].$$

Поскольку в исходной последовательности $a_R^{(l)}$ является восстанавливающим событием, это выражение совпадает с

$$p\left[a_J^{(l+s+1)}, \bigcup_{s=0}^{\infty} (a_I^{(l+s)}, \dots, a_J^{(l+1)}) \mid a_R^{(l)}\right].$$

Из этого следует, что $p(a_J^{*(h)} \mid a_R^{*(h-1)}, A^*)$ не зависит от A^* . Конечные последовательности A^* являются событиями такого типа, к которым относится лемма V.1, — объединение всех этих конечных реализаций дает достоверное событие, отдельные реализации которого не пересекаются, и, наконец, $a_R^{*(h-1)}$ обязательно происходит совместно с одним из событий типа A^* . Согласно лемме V.1, заключаем, что

$$p(a_J^{*(h)} \mid a_R^{*(h-1)}, A^*) = p(a_J^{*(h)} \mid a_{R_I}^{*(h-1)}).$$

Итак, утверждение теоремы V.9, касающееся простой марковости окончательной последовательности, доказано.

Представление (V.3.23) для матрицы переходных вероятностей окончательной последовательности и утверждение о векторе стационарного распределения этой последовательности доказаны в книге Кемени и Снелла (1970, с. 74—76, 147—150).

Аналогичными методами можно показать, что рассмотренное в теореме V.9 преобразование, примененное к марковской цепи k -го порядка, сохраняет этот порядок.

Проиллюстрируем теорему V.10, рассмотрев последовательность зерен Or, Q, Ab в аплитовидном граните из района Омсукчана (обр. 0-51). Статистические данные делают правдоподобным предположение о том, что эта последовательность — простая цепь Маркова. Эргодичность и стационарность ее не вызывают сомнения. Матрица переходных вероятностей

$$P = \begin{matrix} & \begin{matrix} \text{Or} & \text{Q} & \text{Ab} \end{matrix} \\ \begin{matrix} \text{Or} \\ \text{Q} \\ \text{Ab} \end{matrix} & \begin{pmatrix} 0.142 & 0.294 & 0.564 \\ 0.394 & 0.083 & 0.523 \\ 0.434 & 0.329 & 0.236 \end{pmatrix} \end{matrix}, \quad P_{st} = \begin{pmatrix} P(\text{Or}) = 0.326 \\ P(\text{Q}) = 0.256 \\ P(\text{Ab}) = 0.416 \end{pmatrix}.$$

Допустим, что нас интересуют соотношения только между полевыми шпатами, т. е. случайная последовательность, из которой удалены зерна кварца. Тогда, не производя непосредственного удаления зерен из реализации, можно воспользоваться теоремой V.10 для получения последовательности на множестве {Or, Ab}.

Согласно теореме V.10, эта последовательность должна быть стационарной и простой цепью Маркова. Представим матрицу P в клеточной форме, как это показано в (V. 3. 22). Тогда получим

$$\begin{matrix} & \begin{matrix} \text{Or} & \text{Ab} & \text{Q} \end{matrix} \\ \begin{matrix} \text{Or} \\ \text{Ab} \\ \dots \\ \text{Q} \end{matrix} & \begin{pmatrix} 0.142 & 0.564 & 0.294 \\ 0.434 & 0.236 & 0.329 \\ \dots & \dots & \dots \\ 0.394 & 0.523 & 0.083 \end{pmatrix} \end{matrix}.$$

Таким образом,

$$P_{\mathcal{S}', \mathcal{S}'} = \begin{pmatrix} 0.142 & 0.564 \\ 0.434 & 0.236 \end{pmatrix}, \quad P_{\mathcal{S}', \mathcal{S}_1} = \begin{pmatrix} 0.294 \\ 0.329 \end{pmatrix}, \quad P_{\mathcal{S}_1, \mathcal{S}'} = (0.394 \ 0.523), \\ P_{\mathcal{S}_1, \mathcal{S}_1} = (0.083).$$

Подстановка этих матриц в формулу (V. 3. 23) приводит к следующему выражению:

$$P' = \begin{pmatrix} 0.142 & 0.564 \\ 0.434 & 0.236 \end{pmatrix} + \begin{pmatrix} 0.294 \\ 0.329 \end{pmatrix} (1 - 0.083)^{-1} (0.394 \ 0.523),$$

откуда

$$P' = \begin{matrix} & \begin{matrix} \text{Or}' & \text{Ab}' \end{matrix} \\ \begin{matrix} \text{Or}' \\ \text{Ab}' \end{matrix} & \begin{pmatrix} 0.268 & 0.732 \\ 0.575 & 0.425 \end{pmatrix} \end{matrix} \quad \text{и} \quad P'_{st} = \begin{pmatrix} P(\text{Or}') = 0.439 \\ P(\text{Ab}') = 0.561 \end{pmatrix}. \quad (\text{V. 3. 27})$$

Если наши представления об исходной последовательности были верны, то непосредственное удаление из реализации состоя-

ния Q должно привести к реализации, у которой матрица переходных вероятностей и вектор стационарного распределения близки к (V. 3. 27). Прямые наблюдения дали

$$\tilde{P}' = \begin{matrix} & \text{Oг}' & \text{Аб}' \\ \text{Oг}' & (0.275 & 0.725) \\ \text{Аб}' & (0.568 & 0.432) \end{matrix}, \quad \bar{P}'_{st} = \begin{pmatrix} p(\text{Oг}') = 0.437 \\ p(\text{Аб}') = 0.563 \end{pmatrix}.$$

Очевидно, что \tilde{P}' и P' и \bar{P}'_{st} и P'_{st} статистически неотличимы.

Рассмотрим теперь один из вариантов укрупнения по времени—слияние в сериях. Реализация исходной последовательности преобразуется так, что любая I -серия сливается в одно испытание, дающее исход I . Поскольку внутри такой серии ($a_i^{(h)}$, $a_i^{(h+1)}$, ..., $a_i^{(h+l)}$), мы не различаем номеров испытаний, слившихся в одно, здесь имеет место укрупнение по времени. Это же преобразование, как отмечалось в § V.1, формально эквивалентно процессу разрежения.

Теорема V.11. Пусть исходная последовательность является эргодической стационарной цепью Маркова первого порядка на множестве состояний $\mathcal{S} = \{I, J, \dots, K, L\}$. Допустим, что реализации этой последовательности преобразуются путем слияния всех элементов каждой I -серии в I -серию из одного элемента с последующей перенумерацией исходов испытаний в естественном порядке. Тогда возникает эргодическая и однородная цепь Маркова первого порядка.

Доказательство. Рассмотрим последовательность на расширенном множестве состояний $\Sigma = \{I_1, I_2, J, \dots, K, L\}$, где I_1 — первый элемент I -серии; I_2 — все остальные элементы I -серии (если I -серия состоит из одного элемента, то, согласно определению, этот элемент будет I_1).

Переходы через все состояния в Σ простые марковские. Действительно,

$$\begin{aligned} p(I_1 | I_1, A) &= 0, \quad p(I_1 | I_2, A) = 0, \quad p(I_1 | J, A) = \pi(I | J), \\ p(I_2 | I_1, A) &= \pi(I | I), \quad p(I_2 | I_2, A) = \pi(I | I), \quad p(I_2 | J, A) = 0, \\ p(J | I_1, A) &= \pi(J | I), \quad p(J | I_2, A) = \pi(J | I), \quad p(K | J, A) = \pi(K | J), \end{aligned}$$

где $A \in \mathcal{A}_{h-2}$; p — вероятность для последовательности на Σ ; π — вероятность для последовательности на \mathcal{S} .

Поскольку переходы через все состояния на Σ простые марковские, то как отмечалось в гл. IV, последовательность на Σ простая марковская, т. е.

$$p(B | L, A) = p(B | L) \text{ при всех } L \in \Sigma, B \in \mathcal{B}_{h+1}, A \in \mathcal{A}_{h-2}.$$

Последовательность, возникающая после слияния элементов I -серии, формально эквивалентна последовательности, возникшей в результате вычеркивания всех состояний I_2 из последова-

тельности на Σ . Согласно теореме V.9, полное удаление какого-либо состояния из простой цепи Маркова не изменяет порядка марковости. Итак, слияние элементов не нарушает простой марковости. Наличие однородности у новой последовательности и ее эргодичность очевидны.

Случай, рассмотренный в теореме V.11, очень типичен во многих геологических ситуациях. Он возникает, например, при слиянии зерен, образующих серию в первичной породе, в одно зерно после преобразования этой породы. Такие примеры встречаются при прямой перекристаллизации в процессе метаморфизма как горных пород, так и руд (скажем, крупнозернистые галенит-сфалеритовые руды многих полиметаллических месторождений). Не менее часто условия теоремы возникают вследствие методических затруднений. Например, при изучении первично магматических пород в шлифах, последовательности зерен в которых являются простыми стационарными цепями Маркова. В этом случае зерна какого-либо минерала могут быть целиком преобразованы. Границы между зернами таких минералов провести невозможно — их приходится выделять как одно состояние (серпентинизированные оливины в габбро, гаюины в окаитах, иногда калиевые полевые шпаты и т. п.). В этом случае для проверки моделей кристаллизации приходится прибегать к теореме V.11.

В.4. О НЕКОТОРЫХ СХЕМАХ ЗАМЕЩЕНИЯ

Рассмотрим случай, когда замещающим является лишь одно состояние I , не принадлежащее исходной марковской цепи на множестве состояний \mathcal{S} . Схема такого замещения предполагается следующей.

Пусть $a^{(h)}$ — случайная последовательность на множестве состояний \mathcal{S} ; $b^{(h)}$ — случайная последовательность на множестве из двух состояний $\{I, 0\}$, причем $I, 0 \in \mathcal{S}$. Введем третью случайную последовательность

$$c^{(h)} = \begin{cases} a^{(h)}, & \text{если } b^{(h)} = 0, \\ I, & \text{если } b^{(h)} = I, \end{cases}$$

означающую, что происходит независимое синхронное наложение $b^{(h)}$ на $a^{(h)}$, причем состояние I из $b^{(h)}$ замещает синхронное состояние (с тем же номером) из $a^{(h)}$, тогда как наложение 0 из $b^{(h)}$ не меняет синхронного состояния из $a^{(h)}$.

Т е о р е м а. V.12. Пусть в рассмотренной схеме замещения $a^{(h)}$ представляет простую стационарную марковскую цепь, $b^{(h)}$ — последовательность Бернулли. Тогда последовательность, возникающая после замещения, т. е. $c^{(h)}$, является существенно немарковской.

Докажем теорему методом «от обратного».

Доказательство. Пусть условия теоремы выполнены и последовательности $b^{(h)}$ отвечает матрица переходных вероятностей

$$B = \begin{matrix} & 0 & I \\ \begin{matrix} 0 \\ I \end{matrix} & \begin{pmatrix} \pi & 1 - \pi \\ \pi & 1 - \pi \end{pmatrix} \end{matrix}.$$

Предположим также, что $c^{(h)}$ является марковской цепью порядка k . Вычеркнем теперь из всех реализаций $c^{(h)}$ состояние I по схеме разрежения (теорема V.10). Тогда получим последовательность $c^{*(h)}$ на множестве состояний \mathcal{S} , которое будет, согласно теореме V.10, марковской цепью порядка k . Применим теперь к $c^{*(h)}$ новое преобразование, заключающееся во внедрении I -серий согласно теореме V.9.

Пусть распределение серий по длине является геометрическим вида

$$p_l(I) = (1 - \pi)^{l-1}\pi,$$

где $p_l(I)$ — вероятность серии I длины l ; π — элемент матрицы B .

Тогда условия теоремы V.8 выполнены, поскольку реализации $c^{*(h)}$ не содержат состояния I ; кроме того геометрическое распределение не является усеченным. Возникающая в результате внедрения I -серий последовательность, которую обозначим $c^{r(h)}$, согласно теореме V.8, будет существенно немарковской. Но $c^{r(h)}$ по построению совпадает с $c^{(h)}$.

Таким образом, получаем противоречие — $c^{(h)}$ является марковской конечного порядка k (по предположению) и в то же время $c^{(h)}$ оказывается существенно немарковской. Следовательно, неверно предположение, сделанное в доказательстве теоремы V.11, что $c^{(h)}$ может быть марковской цепью конечного порядка k .

Мы видели, что чисто случайное замещение (по схеме Бернулли) ведет к возникновению существенной немарковости. Однако замещение по марковской схеме, когда результат данного замещения учитывает результаты прошлых, может сохранять марковское свойство, как правило, изменяя порядок марковости.

К подобной ситуации относится следующая теорема.

Теорема V.13. Пусть $a^{(h)}$ — стационарная простая марковская цепь с матрицей переходных вероятностей A , $b^{(h)}$ — стационарная простая марковская цепь с запрещенным переходом с I на I . Тогда $c^{(h)}$, представляющая последовательность, возникшую по изложенной выше схеме замещения, в общем случае является цепью Маркова второго порядка. В частном случае марковская цепь $c^{(h)}$ является простой цепью Маркова. Это осуществляется

тогда и только тогда, когда A^2 является матрицей бернуллиевской последовательности (с равными столбцами).

Доказательство. Чтобы доказать, что $c^{(h)}$ является марковской цепью второго порядка, достаточно установить ограниченно марковское свойство второго порядка при переходе через любое состояние $\mathcal{S}' = \mathcal{S} \cup \{I\}$ (относительно эквивалентности ограниченно марковского свойства для переходов через все состояния и марковости того же порядка см. с. 289).

Установим, что переходы через любое состояние из \mathcal{S} обладают ограниченно марковским свойством первого порядка и только переход через состояние I имеет ограниченно марковское свойство второго порядка.

Действительно,

$$p(R^{(h)} | K^{(h-1)} A_{h-2}) = p(R^{(h)} | K^{(h-1)}),$$

где $R, K \in \mathcal{S}$, $A_{h-2} \in \mathcal{A}_{h-2}$; $p(I^{(h)} | I^{(h-1)} A_{h-2}) = 0$, $p(I^{(h)} | K^{(h-1)} A_{h-2}) = \pi$; здесь π — вероятность появления I , т. е. элемент матрицы

$$B = \begin{matrix} & I & 0 \\ I & 0 & 1 \\ 0 & \pi & 1 - \pi \end{matrix}.$$

Остается еще рассмотреть вероятность перехода

$$p(K^{(h)} | I^{(h-1)} R^{(h-2)} A_{h-3}),$$

где $K, R \in \mathcal{S}$, а $A_{h-3} \in \mathcal{A}_{h-3}$. Указанная вероятность вычисляется как

$$\frac{p(A_{h-3}, R^{(h-2)}, \cdot^{(h-1)}, K^{(h)})}{p(A_{h-3}, R^{(h-2)}) (1 - \pi) \pi}, \quad (\text{V. 4. 1})$$

где \cdot означает суммирование по всем состояниям из \mathcal{S} .

Преобразуя это выражение, выносим общий множитель

$$\sum_{J \in \mathcal{S}} p(J | R) p(K | J). \quad (\text{V. 4. 2})$$

Безусловные вероятности события A_{h-3} , очевидно, сокращаются, а переходные вероятности с A_{h-3} на следующее состояние сокращаются в том и только в том случае, когда (V. 4. 2) не зависит от R .

Очевидно, что указанная сумма представляет вероятность перехода из состояния R в состояние K , осуществленного в простой цепи Маркова с матрицей A за два шага, т. е. совпадает с соответствующим элементом квадрата матрицы A . Для сохранения простой марковости в последовательности $c^{(h)}$ необходимо и достаточно, чтобы выражение для этой вероятности было одним и тем же при всех $R \in \mathcal{S}$. Это означает, что A^2 является матрицей со столбцами, составленными из одинаковых элементов.

Таким образом, схемы марковского замещения, при которых не может происходить два замещения подряд (в соседние моменты времени), приводят к цепям второго и первого порядка.

Разработки настоящего параграфа возникли при опытном прогнозировании землетрясений в долине гейзеров (Камчатка) по дискретной последовательности длины интервалов между извержениями гейзера «Великан», наблюдавшимися Н. Г. Сугрובовой.

Изучение замещений различного типа также важно при исследовании процессов преобразования горных пород, особенно с псевдоморфизацией первичных выделений.

У.5. ПОСЛЕДОВАТЕЛЬНОСТИ ПАКЕТОВ

Ранее мы неоднократно встречались с последовательностями, образованными пакетами, т. е. с детерминированно построенными сочетаниями, образующими случайное чередование. Как уже отмечалось, при изучении таких последовательностей часто приходится иметь дело с чередованием не пакетов, а состояний. Последовательность пакетов оказывается более простой структуры и легче поддается аксиоматизации. Основное затруднение при описании разрезов — невозможность установления границы между пакетами. Таким образом, изучая объект (скажем, разрез осадочной толщи), мы исследуем последовательность, в которой стерты границы между пакетами. Такая операция была определена нами в У.1, как стирание границ в последовательности пакетов. Это преобразование может очень сильно изменить марковскую структуру последовательности, существовавшую на множестве пакетов. Так, из бернуллиевской последовательности пакетов после стирания границ может возникнуть существенно немарковская последовательность состояний, образовавших пакеты.

Пример У.14. Рассмотрим, как влияет стирание границ между пакетами на свойства последовательности. Пусть имеются пакеты $A_1 = (\pi, \pi)$ и $A_2 = (\gamma)$, где π — песчаный, а γ — глинистый слой. Эти пакеты слагают разрез. При этом, естественно, один за другим могут следовать несколько пакетов A_1 . Предполагается, что последовательность на множестве пакетов является бернуллиевской. После стирания границ между пакетами образуется последовательность слоев вида

$$\dots, a_{\pi}^{(h)}, a_{\pi}^{(h+1)}, a_{\gamma}^{(h+2)}, a_{\pi}^{(h+3)}, a_{\pi}^{(h+4)}, a_{\pi}^{(h+5)}, a_{\pi}^{(h+6)}, a_{\gamma}^{(h+7)}, \dots$$

Покажем, что эта последовательность существенно немарковская. Рассмотрим для этого две вероятности

$$p(a_{\pi}^{(2h+1)} / a_{\pi}^{(2h)}, \dots, a_{\pi}^{(2)}, \gamma^{(1)}) \text{ и } p(a_{\pi}^{(2h+1)} / a_{\pi}^{(2h)}, \dots, a_{\pi}^{(2)}, a_{\pi}^{(1)}).$$

Первая из этих вероятностей равна единице, так как π серия может иметь лишь четную длину, вторая вероятность — меньше единицы, поскольку здесь возможно появление в $2h+1$ -м испыта-

нии состояния γ . Поскольку h можно выбрать произвольно, это доказывает существенную немарковость последовательности.

Если положение состояния в пакете известно, то мы называем такое состояние «пакетным» и обозначаем двумя нижними индексами. Первый индекс указывает тип пакета, второй — номер, занимаемый данным состоянием в этом пакете. Последовательность таких состояний обозначается как $\dots, \alpha_{Ii}^{(h)}$, где α — символ испытания, h — номер испытания в последовательности пакетных состояний, I — вид пакета, i — место состояния в пакете. Так, например, в структуре рапакиви можно выделить индивиды (овоиды) калиевого полевого шпата (Or), окруженные каймой олигоклаза (Ab). В сечении таких образований прямой возникает пакет (Ab, Or, Ab). В терминах пакетных состояний это сочетание можно представить как $\alpha_{Ab,1}^{(h)}$, $\alpha_{Or,2}^{(h+1)}$, $\alpha_{Ab,3}^{(h+2)}$, где индексы 1 и 3 у Ab соответственно обозначают левое и правое пересечения каймы, а индекс 2 у Or показывает, что индивид Or подвергся окаймлению. Существенно, что при переходе от последовательности пакетов к последовательности пакетных состояний марковость не теряется. При этом справедливо следующее утверждение.

Л е м м а V.3. *Если последовательность пакетов — однородная простая марковская цепь, то последовательность пакетных состояний также является однородной простой цепью Маркова.*

Отметим также, что в том случае, когда последовательность пакетов является последовательностью Бернулли, последовательность пакетных состояний будет простой цепью Маркова, а не последовательностью Бернулли.

Д о к а з а т е л ь с т в о. Достаточно показать, что

$$p(\alpha_{Jj}^{(h)} | \alpha_{Ii}^{(h-1)}, A) = p(\alpha_{Jj}^{(h)} | \alpha_{Ii}^{(h-1)}) (A \in \mathfrak{A}_{h-2}),$$

(где A — событие из прошлого, отнесенное к последовательности пакетных состояний) выполняется для всех пар пакетных состояний $I, i; J, j$.

Поскольку состояние I, i в условии фиксировано, то это означает, что даны как I , так и i . Если состояние i не занимает последнего места в пакете I , то далее с вероятностью 1 следует переход на $\alpha_{I(i+1)}^{(h)}$. В таком случае предыстория не оказывает влияния на эту вероятность. Если состояние i является последним в пакете I , то переход с $\alpha_{Ii}^{(h-1)}$ на $\alpha_{Jj}^{(h)}$ осуществляется с вероятностью $p(A_J/A_I)$, когда j — первое состояние в пакете J , и с вероятностью нуль, когда j не занимает первого места в пакете J . В обоих этих случаях предыстория также не оказывает влияния на переходную вероятность. Лемма доказана.

Существуют ситуации, когда пакетная последовательность не рассматривается как исходная. Она может возникать как преобразование первичной последовательности в результате замещения пакетами отдельных состояний. Допустим, что при подобном замещении соблюдаются следующие условия. Различающиеся

состояния не могут замещаться пакетами одного вида, замещение $a_j^{(h)}$ на пакет $A_p^{(h)}$ происходит с постоянной вероятностью $p_{j,p}$, не зависящей от возможного замещения всех остальных состояний в реализации. Если при замещении такого рода рассматривать незамещенные состояния также как некоторые пакеты, то простая марковская цепь на множестве состояний преобразуется в простую марковскую цепь на множестве пакетов. Это очевидно, так как всегда возможно обратное перекодирование от последовательности пакетов к исходной последовательности (до того, как она прервется замещением).

Пример V.15. Преобразование, заключающееся в окаймлении зерна Ab каймой Og , уже рассматривалось в примере V.7 на с. 318. В том же примере разбирались исходные последовательности. При этом мы стремились выбрать исходную последовательность так, чтобы это была однородная простая цепь Маркова и чтобы укрупнение состояний в этой цепи приводило к изучаемой нами последовательности. После довольно громоздких вычислений в примере V.7 было показано, что такой последовательностью в данном случае является цепь на множестве состояний

$$\Sigma = \{Og_1, Og_2, Og_3, Ab_1, Ab_2, Q_1\},$$

где Og_1 — первичное зерно ортоклаза; Og_2 — левая, а Og_3 — правая части каймы; Ab_1 — неокаймленное зерно; Ab_2 — окаймленное зерно; Q_1 — первичное зерно кварца.

Если использовать символику для пакетных состояний, то этому множеству соответствовало бы множество, отличающееся только обозначениями, т. е.

$$\{\Sigma_1\} = \{Og_I, Og_{I,1}, Og_{I,3}, Ab_{I,2}, Ab_{I,1}, Q_1\},$$

где Og_1 , Ab_1 и Q_1 — первичные зерна; I — пакет (Og, Ab, Og) , вторые индексы — места состояний в пакете.

Как уже отмечалось в примере V.7, для установления простой марковости последовательности на Σ имеется более короткий способ, чем прямые вычисления переходных вероятностей, использованных в этом примере. Что эта последовательность является простой цепью Маркова, можно показать без каких-либо вычислений, например, следующим путем. Окаймление зерен Ab , как отмечалось, приводит к образованию пакетов $Og Ab Og$. Таким образом, с формальной точки зрения образование каймы можно рассматривать как преобразование, заменяющее состояния пакетами. Состояние Ab замещается на пакет $A_I = (Og, Ab, Og)$ с постоянной вероятностью π , если зерно окаймлялось, либо на пакет $A_2 = (Ab)$, если окаймления не происходило, что имеет вероятность $1 - \pi$. Состояния Q и Og замещались пакетами $A_3 = (Q)$ и $A_4 = (Og)$ с вероятностью единица. Такое замещение состояний пакетами соответствует рассмотренной на предыдущей странице схеме замещения. Поэтому последовательность на множестве пакете-

тов $\{(Q), (Og), (Ab), (Og, Ab, Q)\}$ является простой марковской цепью.

Перейдем теперь к последовательности на множестве пакетных состояний. Таким множеством — с точностью до обозначений — будет Σ . Согласно лемме V.3, последовательность на $\{\Sigma\}$ есть простая цепь Маркова.

Приведенное рассуждение типично и позволяет во многих случаях выбрать подходящую для дальнейшего укрупнения исходную последовательность.

Опыт показывает, что результаты многих геологических явлений порождают последовательности, не отличимые от простых цепей Маркова. В этой главе был рассмотрен ряд типичных случаев, когда эти первичные последовательности преобразовывались. Преобразования осуществлялись различными путями. Типичные случаи наблюдались при изучении процессов метасоматического преобразования первично магматических гранитов и при исследовании явления слоеобразования. Естественно, что все случаи преобразований не исчерпаны. Однако рассмотренный материал показывает, как поступать, сталкиваясь с аналогичными процессами. Следует подчеркнуть, что мы исследовали преобразование простых цепей Маркова, а не вообще случайных последовательностей.

Относительно подробный разбор преобразований марковских цепей был вызван опытом геологических исследований. При этом оказалось, что в тех задачах, которые интересуют геологию, математическая сторона вопроса была разработана очень мало. В этой главе дан ряд новых математических разработок, но они отнюдь не исчерпывают тематики. В настоящее время математикам следовало бы разработать следующие, важные для геологии задачи.

1. Получить критерии (необходимые и достаточные условия) слабой укрупняемости. Нужен алгоритм, позволяющий по элементам матрицы переходных вероятностей судить, сохранится ли простая марковость при (слабом) укрупнении. Нами были рассмотрены частные случаи, необходимо дать полное решение.

2. Установить конструктивный критерий возникновения существенно немарковских последовательностей при слабом укрупнении.

3. Разработать математический аппарат для описания существенно немарковских последовательностей.

Решение отмеченных вероятностных задач, возникающих при изучении укрупнения в марковских цепях, позволило бы в дальнейшем разработать соответствующий статистический аппарат. Следует подчеркнуть, что отмеченные задачи имеют большое значение прежде всего для разработки методов исследования различ-

ных метасоматических преобразований. В ряде случаев они необходимы в чисто методических целях, когда то или иное преобразование вызывается спецификой наблюдений.

Л и т е р а т у р а

- Д е м и н а М. Е. О механизме фиксации зерен при формировании песчаных отложений. — В кн.: Исследования по математической геологии. Л., «Наука», 1978, с. 62—69.
- И в а н о в Д. Н. Результаты кристаллизации эвтектической системы $\text{Na}_2\text{CO}_3\text{—NaCl—NaF}$ как аналог идеальных гранитов. — Докл. АН СССР, 1975, т. 222, № 2, с. 448—451.
- К е м е н и Дж., С н е л л Дж. Конечные цепи Маркова. М., «Наука», 1970. 272 с.
- Р о м а н о в а М. А. Марковские свойства последовательностей зерен в редкометалльных гранитах, их использование при поисковых работах и петрологических исследованиях. — В кн.: Геологическая информация и математическая геология. М., «Недра», 1976, с. 53—65.
- Р о м а н о в а М. А. Белые граниты Арга-Ынных-Хая (Якутия) и механизм их образования. — В кн.: Исследования по математической геологии. Л., «Наука», 1978, с. 25—39.
- Ф е л л е р В. Введение в теорию вероятностей и ее приложения. М., «Мир», 1964. 498 с.
- B u r k e С. К., R o s e n b l a t t А. Markovian function of a Markov chain. — Ann. Math. Stat., 1958, v. 29, p. 1112—1122.
- H a r r i s Т. Е. On chains of infinite order. — Pac. J. Math., 1955, suppl. 1, v. 5, p. 707—724.
- V i s t e l i u s А. В. Ideal granite and its properties I. The stochastic model. — J. Int. Assoc. Math. Geol., 1972, v. 4, No. 2, p. 89—102.

Глава VI

СТАТИСТИЧЕСКИЕ ЗАКЛЮЧЕНИЯ О СВОЙСТВАХ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И МАРКОВСКИХ ГИПОТЕЗАХ

Общая постановка вопроса о проверке однородности, обратимости и марковости случайных последовательностей. Статистическая техника, привлекаемая для решения этих задач. Проверка статистических гипотез о частных марковских переходах. Расчетные примеры из геологической практики. Результаты численных экспериментов на ЭВМ, показывающие возможные источники ошибок при заключении о марковском свойстве последовательности.

Ключевые слова:

проверка обратимости, однородности и порядка марковости в марковской цепи, статистика частных марковских переходов, статистические эффекты при работе с немарковскими последовательностями.

VI.1. ВВЕДЕНИЕ

В целом, задача, интересующая геолога, прибегающего к статистике марковских цепей, сводится к следующему. Имеется марковская модель явления, которой отвечает статистическая гипотеза. При исследовании геологического объекта получены наблюдения над случайной последовательностью. Эта последовательность рассматривается как реализация предполагаемого случайного явления. Считая, что модель вытекает из геологической аксиоматики, а статистическая гипотеза отражает черты модели решающей важности, для суждения о справедливости геологической аксиоматики требуется проверить соответствие гипотезы наблюдениям. Иными словами, ситуация в точности отвечает той, что была рассмотрена в гл. III. Единственное отличие состоит в том, что наши наблюдения не являются независимыми, а представляют, вообще говоря, случайную последовательность марковского типа (напомним, что в предельном случае марковская последовательность может вырождаться в детерминированную последовательность, с другой стороны, последовательность независимых

испытаний (бернуллиевскую последовательность) мы считаем марковской цепью с нулевым порядком марковости).

Сказанное повторяет мысли, изложенные в предыдущих главах. Однако это необходимо для того, чтобы еще раз оттенить, что заключение должно строиться на основе априорного введения модели, а не путем индукции из наблюдений, обработанных тем или иным методом. Нужно проверять марковские гипотезы по наблюдениям, а не принимать их на основе расчетов, вводимых без предварительной геологической аксиоматики.

Итак, в этой главе с основным будут рассматриваться различные статистические критерии, используемые для проверки гипотез о структуре случайных, главным образом марковских, последовательностей с конечным порядком марковости.

Как это было принято во всех остальных главах, критерии, заимствованные из литературы, будут излагаться менее подробно. Первые введенные критерии будут приводиться с соответствующими доказательствами.

Рассматриваемые в этой главе критерии получены с помощью различных методов. Критерии для проверки однородности, обратимости и статистики частных марковских переходов построены методом χ^2 , точнее модифицированным методом χ_x^2 , для которого использовано обозначение χ_x^2 (Ван-дер Варден, 1960, с. 233). Критерии для оценки пропорциональности переходных вероятностей и порядка марковской цепи в целом получены методом отношения максимума правдоподобия (статистика Λ). Все три статистики — Λ , χ^2 и χ_x^2 — имеют одно и то же асимптотическое распределение. Однако точечные оценки, используемые для получения Λ , требуют больше вычислений. Поэтому иногда мы пользовались более простыми расчетами, необходимыми для нахождения χ^2 и χ_x^2 . При этом испытание гипотез осуществлялось последовательной проверкой каждого свойства в отдельности в предположении, что остальные свойства имеют место (т. е. включаются в аксиоматику H_0). Таким образом, когда мы проверяли однородность, мы допускали, что порядок марковской цепи известен; проверяя марковский порядок цепи, мы считали, что она однородна. При этом все время предполагалось, что изучаемая последовательность заведомо является марковской цепью с конечным порядком марковости, хотя порядок марковости цепи неизвестен. Оценки, которые строились бы с учетом как однородности, так и марковских свойств одновременно, неизвестны. Случаи, когда порядок марковости бесконечен, имеющимися критериями не охватываются.

Как подчеркивалось в гл. IV, при конечном числе наблюдений невозможно доказать наличие или отсутствие того или иного марковского порядка. Для такого доказательства требуется бесконечная последовательность. Оперировав конечным числом наблюдений, мы, как отмечалось, оценивали не марковость, как часто пишут, а то, что в гл. IV называлось «марковским признаком». Эта терминология будет использована в дальнейшем.

Поскольку аксиоматика, на которой основываются проверочные процедуры, содержит целый ряд условий, на практике может оказаться, что то или иное условие не выполняется и аксиоматика нарушается. В связи с этим устойчивость (robustness) рассматриваемых критериев вряд ли высока. Для суждения об устойчивости критериев в конце главы даны результаты численных экспериментов. Они показали, что устойчивость одного из критериев при умеренном числе наблюдений действительно невысока.

VI.2. ПРОВЕРКА ОДНОРОДНОСТИ

На практике наблюдаемая последовательность на множестве изучаемых состояний всегда оказывается образованной отдельными кусками, разобценными либо другими состояниями, либо пропусками в наблюдениях. Эти куски, совокупность которых образует последовательность, мы будем называть подпоследовательностями. В этом случае проверка однородности у последовательности наблюдений, относительно которой предполагается, что она является цепью Маркова m -го порядка, сводится к проверке того, что все подпоследовательности, из которых состоит эта последовательность, имеют одну и ту же матрицу переходных вероятностей $\underbrace{P_{I, J, \dots, R; L}^{(h)}}_m$, не зависящую от h , где h — номер

подпоследовательности. В этом состоит гипотеза H_0 . H_1 заключается в предположении, что $P_{I, J, \dots, R; L}^{(h)}$ существенно зависит от h . Процедура проверки заключается в том, что для всей последовательности в целом строится одна общая матрица. Затем строятся матрицы для каждой подпоследовательности отдельно. Принимая общую матрицу за теоретическую, с ней по очереди сравнивают каждую из матриц для подпоследовательностей. При этом для каждого сравнения находят значение статистики χ_h^2 , где h — номер подпоследовательности. Для оценки однородности используется

$$\chi^2 = \sum_{h=1}^k \chi_h^2, \quad (\text{VI. 2. 1})$$

где k — число подпоследовательностей в изучаемой последовательности.

Если имеет место однородность, то статистика (VI.2.1) распределена асимптотически как центральное χ^2 . Для случайной последовательности с марковостью порядка m число степеней свободы для статистики (VI. 2. 1) рассчитывается на основании следующих соображений.

Для матрицы, отвечающей любой подпоследовательности, сравниваемой с общей матрицей, имеет место ограничение, заключающееся в том, что сумма вероятностей для любой строки равна единице. При числе состояний в цепи, равном s , общее число

строк в матрице переходных вероятностей равно s^m . Таким образом, марковской цепи порядка m отвечает матрица переходных вероятностей с $s^m \times s$ клеток, откуда число независимых параметров стационарной цепи равно

$$s^m \times s - s^m = s^m (s - 1).$$

Как известно (Ван-дер Варден, 1960), стохастически независимые величины χ^2 обладают аддитивностью (т. е. их можно складывать), причем аддитивны для независимых χ^2 отвечающие им степени свободы. Всего в исследуемой последовательности имеется k подпоследовательностей (кусков). У каждой из $k-1$ этих подпоследовательностей имеется $s^m (s-1)$ независимых параметров. Последняя k -тая подпоследовательность целиком определяется предыдущими $k-1$ -й подпоследовательностями и общей матрицей для всей последовательности.

Таким образом, на k -тую подпоследовательность не расходуются степени свободы. В итоге имеем

$$v = s^m \times (s - 1) (k - 1), \quad (\text{VI.2.2})$$

где v — число степеней свободы у статистики (VI. 2. 1) с центральным χ^2 -распределением.

Если мы принимаем марковость порядка m , то полезно проверять однородность последовательности для всех порядков марковости от 0 до m . Такая проверка не требуется прямым смыслом задачи. Однако при неоднородности последовательности для какого-либо порядка $l < m$ может оказаться, что эта неоднородность исказит истинные соотношения и заключение о том, что последовательность имеет порядок m , окажется ложным. Так как вычислительная сторона при проверке однородности не вызывает затруднений, то кажется, что такую проверку разумно делать для всех $l < m$.

Пример VI. 1. Численная проверка однородности. Последовательность зерен калиевого полевого шпата, кварца и плагиоклаза из аплитовидного гранита Омсукчана (бассейн р. Колымы) состоит из 18 отдельных подпоследовательностей. В целом последовательность не отличима от простой цепи Маркова, что отвечает модели. Покажем оценку однородности, приведя детальные расчеты для первой подпоследовательности.

Для всей последовательности в целом матрица переходов в числе зерен, т. е. матрица частот

$$\tilde{N}_{I; L} = \begin{matrix} & \text{Or} & \text{Q} & \text{Ab} \\ \text{Or} & & & \\ \text{Q} & \begin{pmatrix} 44 & 91 & 175 \\ 95 & 20 & 126 \end{pmatrix} & & \\ \text{Ab} & & \begin{pmatrix} 173 & 131 & 94 \end{pmatrix} & \end{matrix}$$

в оценках переходных вероятностей матрица частот

$$\hat{P}_{I;L} = Q \begin{array}{c} \text{Or} \quad \text{Q} \quad \text{Ab} \\ \left(\begin{array}{ccc} 0.142 & 0.294 & 0.565 \\ 0.394 & 0.083 & 0.523 \\ 0.435 & 0.329 & 0.236 \end{array} \right).$$

Эту матрицу принимаем за теоретическую. Первая подпоследовательность (в кодовых номерах: 1 — Or, 2 — Q, 3 — Ab) такова:

333121232313233133232131311323213112131.

Общее число зерен в подпоследовательности равно сорока.

Матрица $\tilde{N}_{I;L}^{(1)}$ (верхний индекс указывает номер подпоследовательности) для этой подпоследовательности такова:

$$\tilde{N}_{I;L}^{(1)} = Q \begin{array}{c} \text{Or} \quad \text{Q} \quad \text{Ab} \\ \left(\begin{array}{ccc} 2 & 2 & 7 \\ 4 & 0 & 5 \\ 6 & 7 & 6 \end{array} \right).$$

Рассчитываем $\tilde{P}_{I;L}^{(1)}$ для первой подпоследовательности. Для этого берем суммы по строкам матрицы $\tilde{N}_{I;L}^{(1)}$ и умножаем их соответственно на частоты матрицы $P_{I;L}$, которую мы приняли за теоретическую. Получаем

$$N_{I;L}^{(1)} = Q \begin{array}{c} \text{Or} \quad \text{Q} \quad \text{Ab} \\ \left(\begin{array}{ccc} 1.526 & 3.234 & 6.204 \\ 3.546 & 0.747 & 4.707 \\ 8.265 & 6.251 & 4.484 \end{array} \right).$$

Сравнивая $\tilde{N}_{I;L}$ с «теоретической матрицей» $N_{I;L}^{(1)}$, находим

$$\begin{aligned} \chi_1^2 = & \frac{(2.000 - 1.562)^2}{1.562} + \frac{(2.000 - 3.234)^2}{3.234} + \frac{(7.000 - 6.204)^2}{6.204} + \\ & + \frac{(4.000 - 3.546)^2}{3.540} + \frac{(0.000 - 0.747)^2}{0.747} + \frac{(5.000 - 4.707)^2}{4.707} + \\ & + \frac{(6.000 - 8.265)^2}{8.265} + \frac{(7.000 - 6.251)^2}{6.251} + \frac{(6.000 - 4.484)^2}{4.484} = 2.742. \end{aligned}$$

Проводя аналогичные расчеты для всех подпоследовательностей и суммируя χ_i^2 по i , получаем

$$\sum_{i=1}^{18} \chi_i^2 = 2.742 + \dots + 1.77 = 95.4 \quad (\chi_{18}^2 = 1.177).$$

Подсчитывая число степеней свободы (ν), по (VI. 2. 2) находим

$$\nu = 3^1 \times (3 - 1) (18 - 1) = 102.$$

Из таблиц для распределения χ^2 получаем:

$$0.25 < P(\chi_{(102)}^2 \geq 95.4) < 0.5.$$

Значение критерия показывает, что наблюдаемая последовательность не противоречит предположению о том, что она однородна, если порядок марковости $m=1$.

VI.3. ПРОВЕРКА ОБРАТИМОСТИ

Обратимость последовательности $\{a^{(h)}\}$, как мы указывали в гл. IV, означает, что все вероятностные характеристики последовательности должны совпадать независимо от того, как мы опишем ее — слева направо или справа налево.

Любой марковской цепи порядка l соответствует прямоугольная матрица, образованная элементами $\pi_{I, J, \dots, R; L}$.

Как указано в гл. III, с. 162, эти элементы являются параметрами полиномиального распределения. Параметры полиномиального распределения, естественно, не совпадают с переходными вероятностями, так как они нормированы не по строкам, а по общей сумме. Очевидно, что переходные вероятности полностью определены этими параметрами. В такой ситуации обратимость влечет «симметрию» матрицы по отношению к элементам $\pi_{I, J, \dots, R; L}$. Термин «симметрия» у прямоугольной матрицы означает, что

$$\pi_{I, J, \dots, R; L} = \pi_{L, R, \dots, J; I}. \quad (\text{VI.3.1})$$

В дальнейшем мы будем проверять обратимость последовательности, являющейся однородной марковской цепью порядка l , путем статистической проверки равенства (VI.3.1).

Пусть $\tilde{n}_{I, J, \dots, R; L}$ — число наблюдений в клетке $I, J, \dots, R; L$. Если выполняется (VI.3.1), то должно иметь место приближенное равенство

$$\tilde{n}_{I, J, \dots, R; L} \approx \tilde{n}_{L, R, \dots, J; I}. \quad (\text{VI.3.2})$$

Однако на практике почти всегда эти числа несколько отличаются друг от друга. Таким образом, возникает задача оценки значимости разности между ними. Для решения поставленной задачи воспользуемся методом χ^2 , т. е. будем проверять (VI.3.2), рассчитывая

$$\chi^2 = \sum_{I, J, \dots, R; L \in \mathcal{S}} \frac{(\tilde{n}_{I, J, \dots, R; L} - \check{n}_{I, J, \dots, R; L})^2}{\check{n}_{I, J, \dots, R; L}}, \quad (\text{VI.3.3})$$

где величины $\check{n}_{I, J, \dots, R; L}$ образованы так, что

$$\text{а) } \check{n}_{I, J, \dots, R; L} = \check{n}_{L, R, \dots, J; I} > 0,$$

$$\text{б) } \sum_{I, J, \dots, R; L} \check{n}_{I, J, \dots, R; L} = N,$$

в) χ^2 в (VI. 3. 3) достигает условного минимума, если а) и б)

выполнены.

Таким образом, если \check{n} — наблюдаемая частота и если гипотеза обратимости верна, то \check{n} есть оценка минимума χ^2 для n . Для упрощения вместо (VI. 3. 3) мы будем вычислять

$$\chi_x^2 = \sum_{I, J, \dots, R; L \in \mathcal{S}} \frac{(\check{n}_{I, J, \dots, R; L} - \tilde{n}_{I, J, \dots, R; L})^2}{\tilde{n}_{I, J, \dots, R; L}}, \quad (\text{VI. 3. 4})$$

что возможно, так как статистика (VI. 3. 4) асимптотически распределена так же как (VI. 3. 3) и отвечает тем же ограничениям а) и б) для $\check{n}_{I, J, \dots, R; L}$. Таким образом, мы получаем статистику модифицированного χ^2 (VI. 3. 4) (Ван-дер Варден, 1960).

χ_x^2 вычисляется следующим образом.

Пусть

$$\chi_x^2 = \sum_{I, J, \dots, R; L \in \mathcal{S}} \left(\frac{\check{n}_{I, J, \dots, R; L} - \tilde{n}_{I, J, \dots, R; L}}{\sqrt{\tilde{n}_{I, J, \dots, R; L}}} \right)^2 = \mathbf{Y}^t \mathbf{Y},$$

где \mathbf{Y} — вектор-столбец с $s^{(I+1)}$ компонентами (s — число состояний цепи)

$$\mathbf{Y}_{I, J, \dots, R; L} = \frac{\check{n}_{I, J, \dots, R; L} - \tilde{n}_{I, J, \dots, R; L}}{\sqrt{\tilde{n}_{I, J, \dots, R; L}}}$$

или

$$\check{n}_{I, J, \dots, R; L} = \tilde{n}_{I, J, \dots, R; L} + \sqrt{\tilde{n}_{I, J, \dots, R; L}} \mathbf{Y}_{I, J, \dots, R; L}. \quad (\text{VI. 3. 5})$$

Необходимо найти минимум $\mathbf{Y}^t \mathbf{Y}$, подчиняющийся определенным линейным ограничениям, наложенным на величины \check{n} . Наложив эти ограничения на величины $\check{n}_{I, J, \dots, R; L}$, необходимо выразить их через компоненты вектора $\mathbf{Y}_{I, J, \dots, R; L}$.

Система указанных ограничений заключается в том, что

$$\left. \begin{aligned} \check{n}_{I, J, \dots, R; L} - \check{n}_{L, R, \dots, J; I} &= 0, \\ \sum_{I, J, \dots, R; L} \check{n}_{I, J, \dots, R; L} &= N, \quad I, J, \dots, L \in \mathcal{S} \end{aligned} \right\}, \quad (\text{VI. 3. 6})$$

где N — общее число наблюдений.

Исходя из (VI. 3. 5) и (VI. 3. 6), получаем следующую систему линейных уравнений:

$$\begin{aligned} \check{n}_{I, J, \dots, R; L} - \sqrt{\tilde{n}_{I, J, \dots, R; L}} \mathbf{Y}_{I, J, \dots, R; L} - \check{n}_{L, R, \dots, J; I} + \\ + \sqrt{\tilde{n}_{L, R, \dots, J; I}} \mathbf{Y}_{L, R, \dots, J; I} &= 0 \end{aligned}$$

или

$$\begin{aligned} \sqrt{\bar{n}_{I, J, \dots, R; L}} \mathbf{Y}_{I, J, \dots, R; L} - \sqrt{\bar{n}_{L, R, \dots, J; I}} \mathbf{Y}_{I, J, \dots, R; L} = \\ = \bar{n}_{L, R, \dots, J; I} - \bar{n}_{I, J, \dots, R; L} = 0 \end{aligned} \quad (\text{VI. 3. 7})$$

и

$$\begin{aligned} \sum_{I, J, \dots, R; L \in \mathcal{S}} \bar{n}_{I, J, \dots, R; L} - \\ - \sum_{I, J, \dots, R; L \in \mathcal{S}} \sqrt{\bar{n}_{I, J, \dots, R; L}} \mathbf{Y}_{I, J, \dots, R; L} = N. \end{aligned} \quad (\text{VI. 3. 8})$$

Поскольку первая сумма в (VI. 3. 8) равна N , (VI. 3. 8) можно переписать в виде

$$\sum_{I, J, \dots, R; L \in \mathcal{S}} \sqrt{\bar{n}_{I, J, \dots, R; L}} \mathbf{Y}_{I, J, \dots, R; L} = 0. \quad (\text{VI. 3. 9})$$

Таким образом, мы получили систему линейных ограничений на $\mathbf{Y}_{I, J, \dots, R; L}$, определенных уравнениями (VI. 3. 7) и (VI. 3. 9).

Пусть матрица коэффициентов этой системы есть \mathbf{A} , а вектор свободных членов есть \mathbf{b} . Нужно найти условный минимум $\mathbf{Y}^T \mathbf{Y}$ при

$$\mathbf{A} \mathbf{Y} = \mathbf{b}. \quad (\text{VI. 3. 10})$$

Этот условный минимум суммы квадратов вычисляется по формуле (Беллман, 1969, гл. 5, § 6)

$$\mathbf{Y}^T \mathbf{Y} = \mathbf{b}^T (\mathbf{A} \mathbf{A}^{-1})^{-1} \mathbf{b}. \quad (\text{VI. 3. 11})$$

Вычисляя \mathbf{b} по (VI. 3. 11), получаем

$$\mathbf{b} = \begin{pmatrix} 0 \\ \vdots \\ \bar{n}_{L, R, \dots, J; I} - \bar{n}_{I, J, \dots, R; L} \\ \vdots \\ 0 \end{pmatrix}.$$

Из выражения для вектора \mathbf{b} очевидно, что для того, чтобы найти квадратичную форму $\mathbf{b}^T \mathbf{A} \mathbf{b}$, нужно воспользоваться квадратичной формой $\beta^T \mathbf{R} \beta$, где β — вектор \mathbf{b} без последнего компонента, равного нулю, а \mathbf{R} — матрица $(\mathbf{A} \mathbf{A})^{-1}$ без последней строки и последнего столбца.

Рассчитаем теперь элементы матрицы \mathbf{R} . Найдем сначала $\mathbf{A} \mathbf{A}^T$. Линейные ограничения в (VI. 3. 7) показывают, что элемент на главной диагонали $(\mathbf{A} \mathbf{A}^T)_{ii}$ матрицы

$$\begin{aligned} (\mathbf{A} \mathbf{A}^T)_{ii} &= (\sqrt{\bar{n}_{I, J, \dots, R; L}})^2 + (-\sqrt{\bar{n}_{L, R, \dots, J; I}})^2 = \\ &= \bar{n}_{I, J, \dots, R; L} + \bar{n}_{L, R, \dots, J; I}, \end{aligned} \quad (\text{VI. 3. 12})$$

а общий элемент вне главной диагонали

$$(\mathbf{A} \mathbf{A}^T)_{ij} = 0. \quad (\text{VI. 3. 13})$$

Таким образом

$$(AA^T)^{-1} = \begin{pmatrix} 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & \dots & \frac{1}{\bar{n}_{I, J, \dots, R; L} + \bar{n}_{L, R, \dots, J; I}} & \dots & 0 & \dots \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix},$$

а квадратичная форма

$$\beta^T R \beta = (\dots; \bar{n}_{L, R, \dots, J; I} - \bar{n}_{I, J, \dots, R; L}; \dots) \times$$

$$\times \begin{pmatrix} 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & \dots & \frac{1}{\bar{n}_{I, J, \dots, R; L} + \bar{n}_{L, R, \dots, J; I}} & \dots & 0 & \dots \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix} \begin{pmatrix} \vdots \\ \vdots \\ \frac{1}{\bar{n}_{I, R, \dots, J; I} - \bar{n}_{I, J, \dots, R; L}} \\ \vdots \\ \vdots \end{pmatrix} =$$

$$= \sum_{I, J, \dots, R; L \in \mathcal{O}} \frac{(\bar{n}_{I, J, \dots, R; L} - \bar{n}_{L, R, \dots, J; I})^2}{\bar{n}_{I, J, \dots, R; L} + \bar{n}_{L, R, \dots, J; I}}. \quad (\text{VI. 3. 14})$$

Суммируя по всем I, J, \dots, R, L , мы дважды суммируем разность между одними и теми же клетками. Таким образом, (VI. 3. 14) нужно разделить на два. В итоге получаем

$$\chi_x^2 = \frac{1}{2} \sum_{I, J, \dots, R; L \in \mathcal{O}} \frac{(\bar{n}_{I, J, \dots, R; L} - \bar{n}_{L, R, \dots, J; I})^2}{\bar{n}_{I, J, \dots, R; L} + \bar{n}_{L, R, \dots, J; I}}. \quad (\text{VI. 3. 15})$$

Если обратимость имеет место, то статистика (VI.3.15) асимптотически распределена как центральное χ^2 . Для определения числа степеней свободы вычтем число независимых параметров для H_0 , использованное в (VI.3.4), из числа независимых параметров для H_1 в том же (VI.3.4). Число независимых параметров для H_1 в нашем случае

$$\nu_1 = \mu - \mu^{(d)} - 1,$$

где μ — число клеток в прямоугольной матрице, т. е.

$$\mu = s^{m+1}, \quad (\text{VI. 3. 16})$$

s — число состояний, m — принятый порядок марковости, а $\mu^{(d)}$ — число диагональных клеток в прямоугольной матрице. Нижние индексы у этих клеток симметричны (т. е. вида IJI или $RSTTSR$ и т. п.). Число этих клеток вычитается потому, что параметры $\pi_{I, J, \dots, R; L}$, отвечающие им, не используются в уравнении (VI.3.15), так как вклад этих параметров в общую сумму

равен нулю. Выражение для ν_1 содержит минус единицу, так как существует ограничение

$$\sum_{I, J, \dots, R, L \in \mathcal{S}} \pi_{I, J, \dots, R; L} = 1.$$

Все ограничения для параметров полиномиального распределения, имеющиеся при H_1 , имеют место также при H_0 , но к ним добавляются еще ограничения на симметрию (VI. 3. 1).

Число ограничений на симметрию при H_0 равно

$$\frac{\mu - \mu^{(d)}}{2}.$$

Таким образом, в случае H_0 мы имеем число независимых параметров

$$\nu_0 = \mu - \mu^{(d)} - 1 - \frac{\mu - \mu^{(d)}}{2},$$

т. е.

$$\nu = \nu_1 - \nu_0 = \frac{\mu - \mu^{(d)}}{2}, \quad (\text{VI. 3. 17})$$

где ν — число степеней свободы.

Вычисление $\mu^{(d)}$ основано на следующих соображениях. Назовем сочетание $\underbrace{I, J, \dots, R; L}_m$ «словом» длины $m+1$ при алфа-

вите, содержащем s букв (равном числу состояний; если мы изучаем в граните Og , Q и Ab , то наш алфавит содержит три буквы). Число диагональных клеток равно числу таких слов в алфавите из s букв длиной $m+1$, которые читаются одинаково слева направо и справа налево. Здесь возможны два случая. Если m нечетное число, мы выписываем первую половину слова произвольно, а вторую получаем симметричным отражением. Таким образом,

$$\mu^{(d)} = s^{\frac{m+1}{2}}. \quad (\text{VI. 3. 18})$$

Если m четное, то $(m+1)$ — буквенные слова — содержат в середине произвольные буквы. В этом случае левая часть слова выбирается произвольно, в то время как правая получается симметричным отражением. Тогда

$$\mu^{(d)} = s^{(m/2)+1}. \quad (\text{VI. 3. 19})$$

Подстановка в (VI. 3. 17) выражений для μ из (VI. 3. 16) и выражений для $\mu^{(d)}$ из (VI. 3. 18) и (VI. 3. 19) дает

$$\nu = \begin{cases} \frac{s^{m+1} - s^{\frac{(m+1)}{2}}}{2} & \text{для нечетного } m, \\ \frac{s^{m+1} - s^{(m/2)+1}}{2} & \text{для четного } m, \end{cases} \quad (\text{VI. 3. 20})$$

где s — число состояний, а m — принятый порядок марковости. Например, при исследовании гранитов мы будем проверять обратимость простой цепи Маркова на трех состояниях ($s = 3, m = 1$), т. е.

$$v = \frac{3^2 - 3}{2} = 3; \quad (\text{VI. 3. 21})$$

если бы цепь была второго порядка, то

$$v = \frac{3^3 - 3^2}{2} = 9. \quad (\text{VI. 3. 22})$$

Заметим, что в частном случае, когда мы имеем простую цепь Маркова, указанный только что критерий совпадает с информационным критерием симметрии Кульбака (Кульбак, 1967, с. 191). Если $m > 1$, то такого совпадения нет. Если порядок марковости m неясен, то кроме порядка m полезно проверить все порядки для $l \leq m$ ($l = 1, 2, \dots, m$).

Обратимость тривиальна для последовательности независимых испытаний. Она также всегда имеет место при числе состояний, равном двум.

Пример VI.2. Рассмотрим аплитовидный гранит из карьера Данбиты (Шотландия). Предполагая, что последовательность есть марковская цепь второго порядка, изучим с точки зрения обратимости выборочную матрицу для переходов через два шага:

	Or	Q	Ab	
Or Or	0	7	15	}
Or Q	40	10	22	
Or Ab	61	29	22	
Q Or	9	32	38	
Q Q	21	8	8	
Q Ab	29	18	14	
Ab Or	13	38	57	
Ab Q	17	19	30	
Ab Ab	19	19	14	

(VI. 3. 23)

Получим статистику χ_x^2 для проверки обратимости по формуле (VI.3.15). Подсчет дает

$$\begin{aligned} \chi_x^2 = & \frac{1}{2} \left[\frac{(7-9)^2}{7+9} + \frac{(15-13)^2}{15+13} + \frac{(10-21)^2}{10+21} + \frac{(22-17)^2}{22+17} + \frac{(29-29)^2}{29+29} + \right. \\ & + \frac{(22-19)^2}{22+19} + \frac{(9-7)^2}{9+7} + \frac{(33-38)^2}{38+38} + \frac{(21-10)^2}{21+10} + \frac{(8-19)^2}{8+19} + \frac{(29-29)^2}{29+29} + \\ & + \frac{(14-19)^2}{14+19} + \frac{(13-15)^2}{13+15} + \frac{(38-38)^2}{38+38} + \frac{(17-22)^2}{17+22} + \frac{(19-8)^2}{19+8} + \frac{(19-22)^2}{19+22} + \\ & \left. + \frac{(19-14)^2}{19+14} \right] = 10.39. \end{aligned}$$

Формула (VI.3.22) приводит к $v=9$.

Таким образом, по таблице распределения χ^2 имеем $0.500 > P(\chi_9^2 \geq 10.39) > 0.250$, т. е. матрица (VI.3.23) не противоречит предположению о том, что последовательность обратима, если она является марковской цепью второго порядка.

VI.4. КРИТЕРИЙ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ λ ДЛЯ ПРОВЕРКИ ПРОПОРЦИОНАЛЬНОСТИ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ

Исследуем простейший случай, который будет важен при исследовании гранитов и решении задач, связанных с изучением марковских цепей на трех состояниях. В рассматриваемом варианте мы имеем две строки из вероятностей

$$\begin{array}{ccc} p_{11} & p_{12} & p_{13}, \\ p_{21} & p_{22} & p_{23} \end{array} \quad (\text{VI. 4. 1})$$

и хотим проверить гипотезу H_0 :

$$\frac{p_{11}}{p_{21}} = \frac{p_{12}}{p_{22}}. \quad (\text{VI. 4. 2})$$

Процедура получения оценок максимального правдоподобия \hat{p}_{ij} ($i=1, 2; j=1, 2, 3$) для вероятностей (VI.4.1) в деталях изложена в примере III.1. При получении этих оценок мы предполагаем, что гипотеза H_0 истинна, и налагаем на вероятности ограничения, характерные для марковских матриц, т. е.

$$\begin{array}{l} p_{11} + p_{12} + p_{13} = 1, \\ p_{21} + p_{22} + p_{23} = 1, \end{array} \quad (\text{VI. 4. 3})$$

и (VI.4.2). Если гипотеза H_1 истинна, то обязательно имеет место только (VI.4.3).

Оценка максимального правдоподобия при H_1 есть просто

$$\hat{p}_{ij}(H_1) = \bar{p}_{ij} = \frac{\bar{n}_{ij}}{N}. \quad (\text{VI. 4. 4})$$

Оценка максимального правдоподобия для H_0 (см. пример III.1) есть

$$\hat{p}_{ij}(H_0) = \frac{\bar{s}_{i \cdot} \cdot \bar{s}_{\cdot j}}{N_i \bar{s}} \quad (i=1, 2; j=1, 2)$$

и

$$\hat{p}_{i3}(H_0) = \hat{p}_{i3}(H_1) = \bar{p}_{i3} = 1 - \bar{p}_{i1} - \bar{p}_{i2},$$

где

$$\begin{array}{l} \bar{s}_{i \cdot} = \bar{n}_{i1} + \bar{n}_{i2}, \quad \bar{s}_{\cdot j} = \bar{n}_{1j} + \bar{n}_{2j}, \quad \bar{s} = \bar{n}_{11} + \bar{n}_{12} + \bar{n}_{21} + \bar{n}_{22}; \\ N_i = \bar{n}_{i1} + \bar{n}_{i2} + \bar{n}_{i3}; \quad N = N_1 + N_2. \end{array}$$

В нашей задаче функция правдоподобия

$$L(\bar{n}_{11}, \bar{n}_{12}, \bar{n}_{13}, \bar{n}_{21}, \bar{n}_{22}, \bar{n}_{23}) = \frac{N!}{\prod_{\substack{i=1,2 \\ j=1,2,3}} n_{ij}!} \prod_{\substack{i=1,2 \\ j=1,2,3}} (\pi_{ij}^{n_{ij}}) \quad (VI. 4. 5)$$

где π_{ij} — вероятность попадания в ij -тую клетку.

Поскольку вероятности попадания в строку фиксированы (см. гл. III) и равны

$$p_1 = \frac{N_1}{N}, \quad p_2 = \frac{N_2}{N},$$

то, подставляя их, получаем

$$\pi_{ij} = p_i p_{ij} = \frac{N_i}{N} p_{ij}$$

или (для (6. 4. 5))

$$L(\bar{n}_{11}, \dots, \bar{n}_{23}) = \frac{N!}{\prod_{\substack{i=1,2 \\ j=1,2,3}} \bar{n}_{ij}!} \frac{N_1^{N_1} N_2^{N_2}}{N^N} \prod_{\substack{i=1,2 \\ j=1,2,3}} (p_{ij}^{\bar{n}_{ij}}). \quad (VI. 4. 6)$$

Подставим теперь вместо p_{ij} в (VI. 4. 6) первую оценку $p_{ij}(H_0) = \hat{p}_{ij}$. Затем в качестве p_{ij} подставим в (VI. 4. 6) оценку $p_{ij}(H_1) = \bar{p}_{ij}$. Разделим первое выражение на второе. После сокращений получим отношение правдоподобия

$$R = \frac{\prod_{\substack{i=1,2 \\ j=1,2,3}} (\hat{p}_{ij}(H_0)^{\bar{n}_{ij}})}{\prod_{\substack{i=1,2 \\ j=1,2,3}} (\bar{p}_{ij}^{\bar{n}_{ij}})}. \quad (VI. 4. 7)$$

Поскольку оценки для непропорциональных вероятностей p_{13} и p_{23} в обеих гипотезах совпадают:

$$\hat{p}_{13}(H_1) = \hat{p}_{13}(H_0) = \hat{p}_{13}$$

и

$$\hat{p}_{23}(H_1) = \hat{p}_{23}(H_0) = \hat{p}_{23},$$

то из (VI. 4. 7) получаем

$$R = \prod_{\substack{i=1,2 \\ j=1,2}} \left[\frac{\hat{p}_{ij}(H_0)}{\bar{p}_{ij}} \right]^{\bar{n}_{ij}}. \quad (VI. 4. 8)$$

В приложениях используется статистика

$$\Lambda = -2 \ln R. \quad (VI. 4. 9)$$

В нашем случае

$$\Lambda = 2 \left[\bar{n}_{11} \ln \frac{\hat{p}_{11}}{\hat{p}_{11}(H_0)} + \bar{n}_{12} \ln \frac{\hat{p}_{12}}{\hat{p}_{12}(H_0)} + \bar{n}_{21} \ln \frac{\hat{p}_{21}}{\hat{p}_{21}(H_0)} + \bar{n}_{22} \ln \frac{\hat{p}_{22}}{\hat{p}_{22}(H_0)} \right]. \quad (\text{VI. 4. 10})$$

Известно (Уилкс, 1967, гл. 13), что если гипотеза H_0 верна (если исследуемые вероятности пропорциональны), то статистика Λ имеет центральное распределение χ^2 с числом степеней свободы, равным разности между размерностью параметрического пространства для H_1 и для H_0 (размерность параметрического пространства равна минимальному числу координат, которые определяют положение параметрической точки, или числу независимых определяемых параметров). В нашем случае параметрическая точка есть $(p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23})$, а p_{11}, p_{12}, p_{21} и p_{22} могут быть приняты как независимые параметры при H_1 .

Таким образом, для H_1 мы имеем $\nu_1 = 4$. Параметры p_{13} и p_{23} можно выбрать зависимыми как для H_1 , так и для H_0 . При H_0 отношение

$$\frac{p_{11}}{p_{21}} = \frac{p_{12}}{p_{22}}$$

позволяет исключить еще один зависимый параметр.

Итак, для H_0 имеется три независимых параметра ($\nu_0 = 3$). Другими словами, число степеней свободы для статистики Λ в нашем случае равно

$$\nu = \nu_1 - \nu_0 = 4 - 3 = 1.$$

Если пропорциональность имеет место, то Λ асимптотически распределяется как центральное χ^2 с одной степенью свободы. Если же H_0 неверна, то статистика Λ асимптотически распределяется как нецентральное χ^2 с параметром нецентральности ρ . Для вычисления ρ необходимо знать истинное положение параметрической точки, которое обычно неизвестно, что очень осложняет задачу. В связи со сказанным мощность критерия мы не исследуем. Критерий, основанный на статистике Λ , в нашей задаче состоятелен. Он основан на точечных оценках максимального правдоподобия. Как можно показать, эти оценки в нашей задаче состоятельны (Рао, 5-е-1, 1968). Критерий, основанный на состоятельных оценках, состоятелен (см. гл. III, с. 194, 195).

Пример VI.3. Проверим гипотезу о пропорциональности переходных вероятностей для последовательностей зерен, слагающих гранит. В образце АБХ-216 из Якутии (коллекция М. А. Романовой) примем, что переходные вероятности $p_{I, J; K}$ для марковской цепи второго порядка

$$(p(O_{\Gamma}(\hat{h}) | A_{\text{B}}^{(h-1)} O_{\Gamma}^{(h-2)}) \text{ с } p(O_{\Gamma}(\hat{h}) | A_{\text{B}}^{(h-1)} Q^{(h-2)})$$

и

$$p(Q^{(h)} | Ab^{(h-1)}Or^{(h-2)}) \text{ с } p(Q^{(h)} | Ab^{(h-1)}Q^{(h-2)})$$

пропорциональны друг другу.

Проверим эту гипотезу с помощью критерия отношения правдоподобия. Оценки переходных вероятностей, полученные в предположении, что они пропорциональны для H_0 (см. гл. III, пример III. 1), имеют следующие значения:

$$\hat{p}(Or^{(h)} | Ab^{(h-1)}Or^{(h-2)}) = 0.2880, \quad \hat{p}(Q^{(h)} | Ab^{(h-1)}Or^{(h-2)}) = 0.3247, \\ \hat{p}(Ab^{(h)} | Ab^{(h-1)}Or^{(h-2)}) = 0.3873,$$

а

$$\hat{p}(Or^{(h)} | Ab^{(h-1)}Q^{(h-2)}) = 0.3258, \quad \hat{p}(Q^{(h)} | Ab^{(h-1)}Q^{(h-2)}) = 0.3674, \\ \hat{p}(Ab^{(h)} | Ab^{(h-1)}Q^{(h-2)}) = 0.3068.$$

Без предположения о пропорциональности (H_1) оценки переходных вероятностей следующие:

$$\bar{p}(Or^{(h)} | Ab^{(h-2)}Or^{(h-2)}) = 0.2958, \quad \bar{p}(Q^{(h)} | Ab^{(h-1)}Or^{(h-2)}) = 0.3169, \\ \bar{p}(Ab^{(h)} | Ab^{(h-1)}Or^{(h-2)}) = 0.3873,$$

и

$$\bar{p}(Or^{(h)} | Ab^{(h-1)}Q^{(h-2)}) = 0.3119, \quad \bar{p}(Q^{(h)} | Ab^{(h-1)}Q^{(h-2)}) = 0.3743, \\ \bar{p}(Ab^{(h)} | Ab^{(h-1)}Q^{(h-2)}) = 0.3067.$$

Подстановка в (VI.4.10) дает

$$\Lambda = 2 \left(42 \ln \frac{0.2958}{0.2880} + 45 \ln \frac{0.3169}{0.3247} + 52 \ln \frac{0.3191}{0.3258} + 61 \ln \frac{0.3743}{0.3674} \right) = 0.174.$$

Из таблиц распределения χ^2 при одной степени свободы ($\nu = 1$) видно, что предположение о пропорциональности сравниваемых вероятностей не может быть забраковано, так как $0.750 > p(\chi^2 \geq 0.174) > 0.500$.

VI.5. МАРКОВСКИЕ ГИПОТЕЗЫ

Параграфы VI.2, VI.3 и VI.4 были посвящены таким характеристикам последовательностей (однородность, обратимость, пропорциональность), которые не имеют прямого отношения к марковским. Эти характеристики могут иметь место как у марковских, так и у немарковских последовательностей. Параграфы VI.5 и VI.6 посвящены марковским цепям. В настоящем параграфе мы исследуем критерий для статистической проверки гипотезы H_m , что случайная последовательность есть однородная цепь Маркова порядка m против гипотезы H_{m-1} о том, что цепь имеет порядок $m-1$ ($m=1, 2, \dots$ может быть любым целым положительным числом).

Индексация гипотез с помощью номера, указывающего порядок марковости, удобна в рассматриваемой задаче. В общей терминологии, приведенной в гл. III, гипотезе H_{m-1} отвечает H_0 , а гипотезе $H_m - H_1$.

Мы будем рассматривать критерий, основанный на отношении правдоподобия (Anderson, Goodman, 1957; Billingsley, 1961).

При исследовании марковских альтернатив мы будем полагать, что гипотезе H_m отвечает полиномиальное распределение переходных частот

$$\tilde{n}_{\underbrace{I, J, \dots, R}_m; L},$$

где I, J, \dots, R — обозначение строки прямоугольной матрицы; L — обозначение столбца, а $\tilde{n}_{\underbrace{I, J, \dots, R}_m; L}$ — случайное число объектов,

попавших в клетку, указанную нижними индексами; в выборочной последовательности длины N эти переходные частоты распределены по клеткам прямоугольной матрицы с s^m строками и s столбцами (здесь s — число состояний в исследуемой цепи Маркова, а m — ее порядок).

В такой постановке вопроса функция правдоподобия

$$L_m(\tilde{n}_{I, I, \dots, I; I}, \dots, \tilde{n}_{I, J, \dots, R; L}, \dots, \tilde{n}_{L, L, \dots, L; L}) = \\ = \frac{N!}{\prod_{I, J, \dots, R, L \in \mathcal{S}} \tilde{n}_{I, J, \dots, R; L}!} \prod_{\underbrace{I, J, \dots, R}_m, L \in \mathcal{S}} \pi_{I, J, \dots, R; L}^{\tilde{n}_{I, J, \dots, R; L}} \quad (\text{VI. 5. 4})$$

где $\pi_{I, J, \dots, R; L}$ — параметры полиномиального распределения, являющиеся вероятностями попадания в клетки прямоугольной матрицы; \mathcal{S} — множество состояний последовательности (скажем, кварц, калиевый полевой шпат и плагиоклаз при исследовании гранитов). Эти параметры ограничены соотношениями

$$0 < \pi_{I, J, \dots, R; L} < 1 \text{ и } \sum_{\underbrace{I, J, \dots, R}_m, L \in \mathcal{S}} \pi_{I, J, \dots, R; L} = 1. \quad \text{VI. 5. 2.}$$

Прямоугольная матрица переходных вероятностей ($P_{I, J, \dots, R; L}$) может быть получена путем приведения к единице строк матрицы, отвечающей (VI.5.2), по следующей формуле:

$$P_{\underbrace{I, J, \dots, R}_m; L} = \frac{\pi_{I, J, \dots, R; L}}{\sum_{L \in \mathcal{S}} \pi_{I, J, \dots, R; L}} = \frac{\pi_{I, J, \dots, R; L}}{\pi_{I, J, \dots, R; \cdot}}.$$

Здесь и в дальнейшем точка означает суммирование по всем значениям индекса, замененного точкой.

В случае H_{m-1} параметры полиномиального распределения должны удовлетворять (VI.5.2) и кроме того обладать тем свойством, что переходные вероятности $\underbrace{P_{I, J, \dots, R; L}}_m$ подчиняются дополнительному ограничению, заключающемуся в том, что они независимы от первого индекса I .

Другими словами, величины $\underbrace{P_{I, J, \dots, R; L}}_m$ не меняются при любом выборе $I \in \mathcal{S}$. Таким образом, в случае H_{m-1} матрица переходных вероятностей с элементами $p_{I, J, \dots, R; L}^{m-1}$ распадается на s одинаковых подматриц с s^{m-1} строками и s столбцами каждая. Оценки максимального правдоподобия вероятностей в случае H_m имеют вид:

$$\hat{p}_{\underbrace{I, J, \dots, R; L}}^{(m)} = \frac{\bar{n}_{I, J, \dots, R; L}}{\bar{n}_{I, J, \dots, R; \cdot}} \quad \text{VI. 5. 3}$$

Для получения оценок при H_{m-1} мы должны принять тождественность некоторых строк переходных вероятностей. Оценка максимального правдоподобия вероятностей в этих строках

$$\hat{p}_{\underbrace{I, J, \dots, R; L}}^{(m-1)} = \frac{\bar{n}_{\cdot, J, \dots, R; L}}{\bar{n}_{\cdot, J, \dots, R; \cdot}} \quad \text{VI. 5. 4}$$

Оценим теперь параметры полиномиального распределения

$$\underbrace{\pi_{I, J, \dots, R; L}}_m^{(m-1)} \text{ и } \underbrace{\pi_{I, J, \dots, R; L}}_m^{(m)}$$

Как известно,

$$\pi_{I, J, \dots, R; L} = P_{I, J, \dots, R} p_{I, J, \dots, R; L}, \quad \text{VI. 5. 5}$$

где $P_{I, J, \dots, R}$ — вероятность попадания в строку I, J, \dots, R , а $p_{I, J, \dots, R; L}$ — переходная вероятность. Оценки $p_{I, J, \dots, R}$ одни и те же для H_m и H_{m-1} , т. е.

$$\hat{p}_{I, J, \dots, R} = \frac{\bar{n}_{I, J, \dots, R; \cdot}}{N} \quad \text{VI. 5. 6}$$

Таким образом, из (VI. 5. 3) и (VI. 5. 4) получаем

$$\frac{\hat{\pi}_{I, J, \dots, R; L}^{(m-1)}}{\hat{\pi}_{J, J, \dots, R; L}^{(m)}} = \frac{\hat{p}_{I, J, \dots, R; L}^{(m-1)}}{\hat{p}_{J, J, \dots, R; L}^{(m)}} = \frac{\bar{n}_{\cdot, J, \dots, R; L} \bar{n}_{I, J, \dots, R; \cdot}}{\bar{n}_{I, J, \dots, R; \cdot} \bar{n}_{\cdot, J, \dots, R; L}} \quad \text{VI. 5. 7}$$

Используя выражение для R (об отношении правдоподобия R см. гл. III), а также (VI. 5. 1) и (VI. 5. 7), находим

$$R = \frac{N!}{\prod_{I, J, \dots, R, L \in \mathcal{S}} \tilde{n}_{I, J, \dots, R; L}!} \cdot \frac{\prod_{I, J, \dots, R, L \in \mathcal{S}} \hat{\pi}_{I, J, \dots, R; L}^{(m-1)}}{N!} = \frac{\prod_{I, J, \dots, R, L \in \mathcal{S}} \tilde{n}_{I, J, \dots, R; L}}{\prod_{I, J, \dots, R, L \in \mathcal{S}} \hat{\pi}_{I, J, \dots, R; L}^{(m)}} = \prod_{I, J, \dots, R, L \in \mathcal{S}} \left(\frac{\tilde{n}_{I, J, \dots, R; L}^{(m-1)}}{\hat{\pi}_{I, J, \dots, R; L}^{(m)}} \right)$$

Логарифмируя в соответствии с (VI.4.9), окончательно получаем

$$\Lambda = 2 \sum_{I, J, \dots, R, L \in \mathcal{S}} \tilde{n}_{I, J, \dots, R; L} \ln \frac{\tilde{n}_{I, J, \dots, R; L} \tilde{n}_{\cdot, J, \dots, R; \cdot}}{\tilde{n}_{\cdot, J, \dots, R; \cdot} \tilde{n}_{I, J, \dots, R; \cdot}}. \quad (\text{VI. 5. 8})$$

При изучении наблюдаемых последовательностей на первом этапе работы следует проверять гипотезу о последовательности независимых испытаний H_0 против предположения о том, что она является простой (первого порядка) цепью Маркова. В этом случае формула (VI.5.8) преобразуется к виду

$$\Lambda = 2 \sum_{I; L} \tilde{n}_{I; L} \ln \left(\frac{\tilde{n}_{\cdot; L} \tilde{n}_{I; \cdot}}{\tilde{n}_{\cdot; \cdot} \tilde{n}_{I; L}} \right) = 2 \sum_{I; L} \tilde{n}_{I; L} \ln \left(\frac{N \tilde{n}_{I; L}}{\tilde{n}_{\cdot; L} \tilde{n}_{I; \cdot}} \right). \quad (\text{VI. 5. 9})$$

Следующий шаг заключается в проверке марковости первого порядка против марковости второго порядка. В этом случае из (VI. 5. 8) следует формула

$$\Lambda = 2 \sum_{I, J; L} \tilde{n}_{I, J; L} \ln \left(\frac{\tilde{n}_{\cdot, J; L} \tilde{n}_{I, J; \cdot}}{\tilde{n}_{\cdot, J; \cdot} \tilde{n}_{I, J; L}} \right). \quad (\text{VI. 5. 10})$$

Далее, альтернативные гипотезы для более высоких порядков марковости проверяются по формулам, получаемым тем же способом из (VI.5.8).

Отыщем теперь число степеней свободы для Λ (Леман, 1964, с. 426; Уилкс, 1967, с. 416; глава III и § VI.4 этой книги). Единственными параметрами, чьи оценки входят в статистику Λ , являются переходные вероятности $p_{I, J, \dots, R; L}$. В случае H_m имеется

$$v_m = s^m \times s - s^m$$

независимых параметров, так как мы имеем s^m ограничений типа

$$\sum_{L \in \{s\}} p_{I, J, \dots, R; L} = 1.$$

В случае H_{m-1} число независимых параметров меньше, так как в матрице переходных вероятностей имеется s идентичных

подматриц с s^{m-1} строками и с s столбцами в каждой. Для общей подматрицы имеется только s^{m-1} ограничений в соответствии с числом строк в подматрице. Таким образом, мы имеем

$$\nu_{m-1} = s^{(m-1)} \times s - s^{(m-1)}$$

независимых параметров.

Как указано на с. 361, если H_{m-1} истинна, то статистика Λ асимптотически распределена как центральное χ^2 с $\nu = \nu_m - \nu_{m-1}$ степенями свободы, где ν — число степеней свободы; ν_m — число независимых параметров из H_m ; ν_{m-1} — число независимых параметров из H_{m-1} . Обе эти гипотезы отражены в Λ . В итоге имеем:

$$\nu = s^{(m-1)} (s - 1)^2. \quad (\text{VI. 5. 11})$$

При проверке бернуллиевской гипотезы против марковской первого порядка $m=1$ и $\nu=4$, в случае проверки цепи первого порядка против цепи второго порядка $m=2$ и $\nu=12$ (если $s=3$).

В частном случае, когда $m=1$, критерий, основанный на отношении правдоподобия, совпадает с информационным критерием Кульбака (Кульбак, 1967) для проверки независимости по таблице сопряженности с двумя входами. Однако при $m=2$ и для более высоких порядков критерий, используемый нами, не совпадает с кульбаковским для трехвходовой таблицы сопряженности и для подобных же таблиц с большим числом входов (Кульбак, 1967, с. 167—174), (если $s=3$).

Пример VI.4. Проверим непротиворечивость марковских гипотез последовательности зерен кварца (Q), калиевого полевого шпата (Or) и плагиоклаза (Ab) в аплитовидном граните из карьера близ Данбиты в юго-западной Шотландии.

Всего было исследовано q переходов с зерна на зерно. Они образовывали k кусочных подпоследовательностей, которые составляли изучаемую последовательность. Таким образом, при построении матрицы переходных вероятностей для проверки марковости первого порядка мы теряем $k-1$ наблюдение, при проверке марковости второго порядка — $2(k-1)$ наблюдений и т. д. Другими словами, для проверки марковости первого порядка сохраняется $N-k+1$ наблюдение, для второго порядка марковости $N-2(k-1)$ наблюдений и т. д.

Матрица переходных вероятностей на один шаг для исследованного гранита следующая:

$$\tilde{N}_{I;L} = \begin{matrix} & \text{Or} & \text{Q} & \text{Ab} \\ \text{Or} & 227 & 542 & 1019 \\ \text{Q} & 521 & 291 & 411 \\ \text{Ab} & 1041 & 392 & 559 \end{matrix}.$$

Подсчитывая Λ по формуле (VI. 5. 9), проверяем, что представляет собой последовательность с точки зрения альтернативы — бер-

нуллиевская последовательность против простой марковской цепи. Частоты и общее число наблюдений следующие:

$$\bar{n}_{11} = 227, \bar{n}_{12} = 542, \bar{n}_{13} = 1019, \dots, \bar{n}_{33} = 559, \bar{n}_{.1} = 1788, \bar{n}_{.2} = 1223, \\ \bar{n}_{.3} = 1992, \bar{n}_{.1} = 1789, \bar{n}_{.2} = 1225, \bar{n}_{.3} = 1989, N = 5003;$$

$$\Lambda = 2 \left(227 \ln \frac{227 \times 5003}{1788 \times 1789} + 542 \ln \frac{542 \times 5003}{1788 \times 1225} + 1019 \ln \frac{1019 \times 5003}{1788 \times 1789} + \right. \\ \left. + 521 \ln \frac{521 \times 5003}{1223 \times 1789} + 291 \ln \frac{291 \times 5003}{1223 \times 1225} + 411 \ln \frac{411 \times 5003}{1223 \times 1989} + \right. \\ \left. + 1041 \ln \frac{1041 \times 5003}{1992 \times 1789} + 392 \ln \frac{392 \times 5003}{1992 \times 1225} + 559 \ln \frac{559 \times 5003}{1992 \times 1989} \right) = \\ = 2 [227 \times (-1.0357) + 542 \times 0.2135 + 1019 \times 0.4662 + 521 \times 0.1748 + \\ + 291 \times (-0.0284) + 411 \times (-0.1684) + 1041 \times 0.3791 + \\ + 392 \times (-0.2182) + 559 \times (-0.3482)] = 967.46.$$

Если бы мы имели последовательность независимых испытаний, то число степеней свободы для распределения Λ при трех состояниях определялось бы по формуле (VI.5.11) и было бы равно 4. Из таблицы распределения χ^2 мы находим, что для 5%-го уровня значимости $\chi_{0.05}^2 = 9.5$ при $\nu = 4$, а для 1%-го уровня $\chi_{0.01}^2 = 13.3$. Поскольку наблюдаемое значение $\chi^2 = 967.46$, очевидно, что гипотеза о независимых испытаниях должна быть отброшена.

Переходим к следующей альтернативе.

Теперь мы должны проверить гипотезу о марковости первого порядка против гипотезы о марковости второго порядка. Выборочная матрица, по которой проводится проверка, дана ниже:

	Or	Q	Ab
Or Or	47	69	110
Or Q	253	116	117
Or Ab	618	156	243
Q Or	69	176	275
$\tilde{N}_{I, J; L} =$ Q Q	112	86	92
Q Ab	171	110	129
Ab Or	111	297	630
Ab Q	153	88	149
Ab Ab	247	123	184

Используя формулу (VI.5.10), получаем статистику Λ . В этом случае необходимо подсчитать 27 членов по формуле (VI.5.10). Первый член таков:

$$47 \ln \frac{47 \times (226 + 520 + 1038)}{226 \times (47 + 69 + 111)} = 47 \ln \frac{47 \times 1784}{226 \times 227} = 47 \ln 1.634.$$

Получен этот член так. В клетке $OgOg$ матрицы $N_{I, J; L}$ имеется 47 наблюдений; n_{Og} равно сумме частот во всех клетках, где второй индекс Og , а первый и третий произвольны. Таким образом, 226 есть сумма $47+69+110$ и равна частоте в строке $OgOg$. Аналогично 520 есть сумма $69+176+275$, а число 1038 образовано из $111+297+630$. В знаменателе 226 отвечает $n_{I, J;}$ из (VI.5.10). Здесь I и J означают Og и, таким образом, это число совпадает с $n_{Og, Og}$ и равно сумме частот в строке $OgOg$ или $47+69+110$. Часть знаменателя в скобках отвечает $n_{, J; L}$ в (VI.5.10) или в нашем случае $n_{, Og; Og}$. Это сумма частот во всех клетках с последними двумя индексами $OgOg$ и произвольными первыми индексами (т. е. $Og, Og; Og; Q, Og; Og$ и $Ab, Og; Og$ с частотами $47+69+111$). По указанному способу находим

$$\begin{aligned} \Lambda = & 2 \left(47 \ln \frac{47 \times 1784}{226 \times 227} + 69 \ln \frac{69 \times 1784}{226 \times 542} + 110 \ln \frac{110 \times 1784}{226 \times 1015} + \right. \\ & + 253 \ln \frac{253 \times 1219}{539 \times 518} + 116 \ln \frac{116 \times 1219}{539 \times 290} + 170 \ln \frac{170 \times 1219}{539 \times 411} + \\ & + 618 \ln \frac{618 \times 1981}{1017 \times 1036} + 156 \ln \frac{156 \times 1981}{1017 \times 389} + 243 \ln \frac{243 \times 1981}{1017 \times 556} + \\ & + 69 \ln \frac{69 \times 1784}{520 \times 227} + 176 \ln \frac{176 \times 1784}{520 \times 542} + 275 \ln \frac{275 \times 1784}{520 \times 1015} + \\ & + 112 \ln \frac{112 \times 1219}{290 \times 518} + 86 \ln \frac{86 \times 1219}{290 \times 290} + 92 \ln \frac{92 \times 1219}{290 \times 411} + \\ & + 171 \ln \frac{171 \times 1981}{410 \times 1096} + 110 \ln \frac{110 \times 1981}{410 \times 389} + 129 \ln \frac{129 \times 1981}{410 \times 556} + \\ & + 111 \ln \frac{111 \times 1784}{1038 \times 227} + 297 \ln \frac{297 \times 1784}{1038 \times 542} + 630 \ln \frac{630 \times 1784}{1038 \times 1015} + \\ & + 153 \ln \frac{153 \times 1219}{320 \times 518} + 88 \ln \frac{88 \times 1219}{390 \times 290} + 149 \ln \frac{149 \times 1219}{390 \times 411} + \\ & \left. + 247 \ln \frac{247 \times 1981}{554 \times 1036} + 123 \ln \frac{123 \times 1981}{554 \times 389} + 184 \ln \frac{184 \times 1981}{554 \times 556} \right) = \\ & = 2(47 \times 0.4910 + \dots + 184 \times 0.1681) = 100.57. \end{aligned}$$

Формула (VI.5.13) для подсчета числа степеней свободы дает

$$\nu = 3(3-1)^2 = 12.$$

Из таблицы распределения χ^2 видно, что для 1%-го уровня при 12 степенях свободы $\chi_{0.01}^2 = 28.3$, но мы имеем 100.57. Таким образом, гипотеза о тождестве наблюдаемой последовательности с простой цепью Маркова может быть уверенно забракована.

Переходим теперь к выяснению вопроса о том, насколько наблюдаемая последовательность не противоречит предположению о том, что она отвечает цепи Маркова либо второго, либо третьего порядка. Для этого вводится альтернатива о цепи второго порядка

против цепи третьего порядка.¹ Повторяя вычисления для матрицы $\tilde{N}_{I, J, K, L}$, содержащей 81 клетку, мы получили

$$\Lambda = 29.4, \nu = 36.$$

На основании данных таблиц распределения χ^2 можно убедиться, что полученное значение не бракует гипотезы о том, что изучаемая последовательность является цепью Маркова второго порядка.

VI.6. ЧАСТНАЯ МАРКОВСТЬ

Как отмечалось в гл. IV, изучение геологических явлений привело к тому, что были обнаружены такие случайные последовательности марковского типа, в которых через разные состояния или сочетания состояний наблюдаются переходы разного порядка марковости, равно как через отдельные состояния возможны существенно немарковские переходы. Такие структуры обсуждались в гл. IV, где они были названы частными марковскими переходами (частная марковость); там же отмечалась их характерность для геологических объектов.

Ситуации, возникающие при наличии частных марковских переходов, очень разнообразны. Вследствие этого нет критерия, охватывающего все случаи; для каждой задачи необходима собственная проверочная процедура. Поэтому и расчет числа степеней свободы для критериев Λ и χ^2 в разных задачах различен. В таких обстоятельствах кажется разумным изложить методику статистической проверки частной марковости на материале конкретных примеров.

Пример VI. 5. Простейший случай проверки частной марковости сводится к следующему. Мы хотим проверить гипотезу, что переход через некоторое фиксированное состояние является простым марковским без определения марковского типа такого перехода через другие состояния (H_0). H_1 в этом случае отвечает любой случайной последовательности (мы изучаем только однородные последовательности).

При численной проверке отмеченной альтернативы мы заменяем исследование простой марковости, как было отмечено в гл. IV и как мы только что поступили в предыдущем параграфе, исследованием марковского признака первого порядка. В этом случае мы берем матрицы $P_{I, L}$ и $P_{I, J, L}$ и сравниваем переходные вероятности через фиксированное состояние x (переход через которое мы изучаем). Эта процедура аналогична процедуре, рассмотренной в § VI.5, где мы изучали вопрос о том, как отличить марковские свойства цепи первого порядка от соответствующих

¹ Поскольку вычисления с матрицей $\tilde{N}_{I, J, K, L}$ в принципе ничем не отличаются от работы с матрицей $\tilde{N}_{I, J, L}$, матрицу $N_{I, J, K, L}$ мы не приводим.

свойств цепи второго порядка. Однако вместо сравнения всей матрицы $P_{I;L}$ ($I, L \in \mathcal{S}$) со всей матрицей $P_{I,J;L}$ ($I, J, L \in \mathcal{S}$), с которыми мы имели дело в § IV.5, здесь мы берем подматрицы

$$P_{x;L} (L \in \mathcal{S}) \text{ и } P_{I,x;L} (I, L \in \mathcal{S}), \quad (\text{IV.6.1})$$

в то время как состояние x фиксировано.

Легко заметить, что оценка максимального правдоподобия для переходных вероятностей в этом случае аналогична рассмотренным в § VI.5. Однако статистика отношения правдоподобия Λ строится с учетом данных, вошедших только в матрицы (VI.6.1), и имеет соответственно иное число степеней свободы, чем статистика Λ , рассмотренная в § VI.5, хотя вид асимптотического распределения сохраняется. Это распределение является центральным χ^2 , если H_0 истинно, и нецентральным, если H_0 неверно. В случае H_1 имеется $s(s-1)$ независимых параметров (переходных вероятностей), где s — число состояний в \mathcal{S} . При H_0 имеется $s-1$ независимых переходных вероятностей. Таким образом,

$$v = s(s-1) - (s-1) = (s-1)^2.$$

В частности, если $s=3$, как это имеет место во многих наших исследованиях, $v=4$.

Проверим теперь, что переходы через x и y являются простыми марковскими одновременно. Поскольку центральное χ^2 обладает аддитивностью, равно как аддитивно число степеней свободы, мы имеем (для трех состояний в \mathcal{S})

$$\Lambda_{x,y} = \Lambda_x + \Lambda_y$$

с восемью степенями свободы. В то же время статистика

$$\Lambda_{x,y,z} = \Lambda_x + \Lambda_y + \Lambda_z$$

имеет 12 степеней свободы. Эта статистика является одновременно статистикой Λ в § VI.5, которую мы использовали для проверки гипотезы о марковской цепи первого порядка против альтернативы о цепи второго порядка при трех состояниях.

В тех случаях, когда необходимо проверить марковские гипотезы относительно нескольких переходов, имеющих различный марковский порядок, оценки максимального правдоподобия для переходных вероятностей довольно сложны. В этих случаях мы использовали оценки минимума χ^2 и соответственно статистику χ^2 .

Если необходимо проверить гипотезу о том, что переход через некоторое сочетание состояний является марковским порядка m_1 , а через другое сочетание состояний — марковским порядка $m_2 > m_1$ (H_0), то в качестве альтернативы H_1 следует использовать цепь с марковским порядком выше m_2 . Наименьший порядок цепи в этом случае будет m_2+1 . Поэтому полезно принять в качестве альтернативы марковскую цепь с порядком перехода на

единицу выше, чем марковский порядок сочетания с наивысшим порядком марковского перехода.

Методика оценивания переходных вероятностей и расчета числа степеней свободы разъясняется следующим примером.

Пусть проверяется предположение о том, что переход через сочетание $a_L^{(h+1)} \dots a_L^{(h+m)}$ является марковским порядка m против альтернативы, что этот переход марковский порядка $r > m$. Тогда статистика, основанная на минимуме χ^2 , отыскивается следующим образом.

Берем матрицу переходных частот размерности $r \times s$. В этой матрице мы рассматриваем только те строки, которые оканчиваются изучаемым сочетанием, т. е. строки вида

$$\overbrace{*** \dots * a_I \dots a_L}^r; \quad (\text{VI. 6. 2})$$

$\underbrace{\hspace{10em}}_{r-m} \quad \underbrace{\hspace{5em}}_m$

первые $r-m$ состояний произвольны (что мы обозначили *). При H_0 переходные вероятности во всех строках с такими входами, как в (VI.6.2), одни и те же, независимо от того, какое сочетание состояний располагается на местах, указанных *. В этом случае оценка переходных вероятностей с помощью минимума χ^2 (Пао, 1968) получается суммированием соответствующих частот во всех идентичных строках и последующим делением на общую сумму по всем этим строкам. Например:

$$\check{p}(a_x^{(h+r+1)} | a_L^{(h+r)} \dots a_L^{(h+m)} \underbrace{***}_{r-m}) = \frac{\tilde{n}_{\dots I \dots L; x}}{\sum_{x \in \mathcal{P}} \tilde{n}_{\dots I \dots L; x}}$$

Для вычисления числа степеней свободы необходимо взять число независимых параметров в H_1 и вычесть из него число независимо оцениваемых параметров для H_0 . В нашем примере независимыми параметрами являются переходные вероятности.

Для H_1 , если мы имеем s состояний, имеется $s-1$ независимых переходных вероятностей в одной строке; общее число строк равно s^{r-m} ; таким образом, $v_1 = (s-1) s^{r-m}$.

Для H_0 мы предполагаем, что все эти s^{r-m} строк идентичны. Таким образом, фактически имеется одна строка с $s-1$ независимым параметром, т. е.

$$v_0 = s - 1.$$

Итак,

$$v = v_1 - v_0 = (s-1) s^{r-m} - (s-1) = (s-1) (s^{r-m} - 1).$$

Если есть необходимость, то можно проверить одновременное существование нескольких марковских переходов для различных сочетаний против альтернативы об одном определенном марковском

порядке в цепи в целом. Для этой цели достаточно построить критерий для проверки отдельных сочетаний и сложить полученные статистики, объединяя величины с центральным χ^2 -распределением и складывая соответствующие числа степеней свободы.

Пример VI.6. Для шлифа АБХ-44 из Якутии (материалы М. А. Романовой) необходимо было проверить следующие альтернативы. H_0 — переход через Q является марковским первого порядка, а переход через сочетание зерен $Or^{(h-1)}Ab^{(h)}$ — марковский второго порядка против H_1 , когда оба эти перехода являются марковскими третьего порядка.

Обратимся к матрицам частот наблюдаемых переходов для проверки соответствующих альтернатив.

Сначала образуем статистику χ^2 для переходов через зерно кварца (Q). Для этого отберем из матрицы $\tilde{N}_{I, J, K; L}$ все строки, оканчивающиеся переходом через Q. В итоге получим матрицу и вектор

$$\tilde{N}_{I, J, Q; L} = \begin{matrix} & Or & Q & Ab & \Sigma \\ \begin{matrix} Or & Or & Q \\ Or & Q & Q \\ Or & Ab & Q \\ Q & Or & Q \\ Q & Q & Q \\ Q & Ab & Q \\ Ab & Or & Q \\ Ab & Q & Q \\ Ab & Ab & Q \end{matrix} & \begin{bmatrix} 0 & 1 & 1 \\ 1 & 3 & 3 \\ 5 & 2 & 16 \\ 0 & 2 & 3 \\ 1 & 3 & 9 \\ 3 & 11 & 27 \\ 7 & 4 & 18 \\ 4 & 8 & 19 \\ 7 & 19 & 45 \end{bmatrix} & , & \begin{bmatrix} 2 \\ 7 \\ 23 \\ 5 \\ 13 \\ 41 \\ 29 \\ 31 \\ 71 \end{bmatrix} \end{matrix} \quad (VI. 6. 3)$$

Если H_0 истинна, то существует только одна строка теоретических переходных вероятностей для всей матрицы (VI. 6. 3). Эта строка имеет вид

$$** Q \begin{pmatrix} Or & Q & Ab \\ \dots & \dots & \dots \end{pmatrix}.$$

В этом случае мы имеем следующие оценки:

$$\begin{aligned} \bar{p}(Or^{(h)}/Q^{(h-1)}) &= \\ &= \frac{0 + 1 + 5 + 0 + 1 + 3 + 7 + 4 + 7}{2 + 7 + 23 + 5 + 13 + 41 + 29 + 31 + 17} = \frac{28}{222} = 0.128, \\ \bar{p}(Q^{(h)}/Q^{(h-1)}) &= \frac{53}{222} = 0.239, \\ \bar{p}(Ab^{(h)}/Q^{(h-1)}) &= \frac{141}{222} = 0.635. \end{aligned} \quad (VI. 6. 4)$$

Найдем теперь теоретические частоты для матрицы $N_{I, J, Q; L}$. Они могут быть получены умножением сумм, стоящих справа в (VI. 6. 3), на соответствующие значения \bar{p} , что дает

$$\tilde{N}_{I, J, Q; L} = \begin{matrix} & & \text{Or} & \text{Q} & \text{Ab} \\ & \text{Or Or Q} & \left[\begin{array}{ccc} 0.3 & 0.5 & 1.3 \\ 0.9 & 1.7 & 4.4 \\ 2.9 & 5.5 & 14.6 \\ 0.6 & 1.7 & 3.2 \\ 1.6 & 3.1 & 8.3 \\ 5.3 & 9.8 & 26.0 \\ 3.7 & 6.9 & 18.4 \\ 3.9 & 7.4 & 19.7 \\ 8.9 & 17.0 & 45.1 \end{array} \right. \\ & \text{Or Q Q} & \\ & \text{Or Ab Q} & \\ & \text{Q Or Q} & \\ & \text{Q Q Q} & \\ & \text{Q Ab Q} & \\ & \text{Ab Or Q} & \\ & \text{Ab Q Q} & \\ & \text{Ab Ab Q} & \end{matrix} .$$

Подсчет показывает, что $\chi_{16}^2(Q) = 13.63$.

16 степеней свободы были найдены так. В матрице $\tilde{N}_{I, J, Q; L}$ при $I, J, L \in \{\text{Or}, \text{Q}, \text{Ab}\}$ имеются 18 независимых параметров (переходных вероятностей), которые должны быть оценены. Согласно H_0 , в этой матрице имеются только две переходные вероятности, которые необходимо оценить и которые могут быть найдены из (VI.6.4).

Таким образом,

$$v(Q) = 18 - 2 = 16.$$

Вычислим теперь статистику χ^2 для случая перехода через $\text{Or}^{(h-1)}\text{Ab}^{(h)}$. Согласно H_1 , этот переход является марковским третьего порядка. Этому переходу отвечает:

$$\tilde{N}_{I, (\text{Or}, \text{Ab}); L} = \begin{matrix} & & \text{Or} & \text{Q} & \text{Ab} & \Sigma \\ & \text{Or Or Ab} & \left(\begin{array}{ccc} 5 & 4 & 7 \\ 8 & 3 & 10 \\ 37 & 14 & 40 \end{array} \right) & \left(\begin{array}{c} 16 \\ 21 \\ 91 \end{array} \right) \\ & \text{Q Or Ab} & \\ & \text{Ab Or Ab} & \end{matrix} .$$

Для гипотезы H_0 этим трем строкам отвечает одна строка теоретических переходных вероятностей

$$* \text{Or Ab} \left(\begin{array}{ccc} \text{Or} & \text{Q} & \text{Ab} \\ \dots & \dots & \dots \end{array} \right).$$

Вероятности для этой строки оцениваются ниже:

$$\hat{p}(\text{Or}^{(h)} | \text{Ab}^{(h-1)}\text{Or}^{(h-2)}x^{(h-3)}) = \frac{5 + 8 + 37}{16 + 21 + 91} = \frac{50}{128} = 0.391,$$

$$\hat{p}(\text{Q}^{(h)} | \text{Ab}^{(h-1)}\text{Or}^{(h-2)}x^{(h-3)}) = \frac{21}{128} = 0.164,$$

$$\hat{p}(\text{Ab}^{(h)} | \text{Ab}^{(h-1)}\text{Or}^{(h-2)}x^{(h-3)}) = \frac{57}{128} = 0.446.$$

По этим данным рассчитываем теоретическую матрицу

$$N_{I, (\text{Or}, \text{Ab}); L} = \begin{matrix} & & \text{Or} & \text{Q} & \text{Ab} \\ & \text{Or Or Ab} & \left(\begin{array}{ccc} 6.3 & 2.6 & 7.1 \\ 8.2 & 3.4 & 9.3 \\ 35.6 & 14.9 & 40.5 \end{array} \right) \\ & \text{Q Or Ab} & \\ & \text{Ab Or Ab} & \end{matrix} .$$

Сравнение теоретической матрицы с наблюдениями дает $\chi^2_4(Og, Ab) = 1.23$. Число степеней свободы $\nu = 6 - 2 = 4$. Теперь мы можем проверить сложную гипотезу о том, что переход через Q — марковский первого порядка, а переход через Og Ab — марковский второго порядка. Для этой гипотезы мы получаем $\chi^2_{20}(Q \text{ и } Og, Ab) = 14.86$, в то время как χ^2_{20} , отвечающее уровню 5%, равно 31.4.

Таким образом, наблюдения не противоречат гипотезе о том, что в изучаемой последовательности переход через Q имеет первый, а через сочетание (Og, Ab) — второй марковский порядок. Как мы видели, статистика $H_1 : H_2$ параграфа VI.5 образуется как сумма статистик для проверки простоты каждого частного перехода (соответствующее распределение χ^2 имеет суммарное число степеней свободы). В указанных условиях может случиться, что одно из этих слагаемых будет достаточно велико, чтобы простота соответствующего частного перехода была забракована, однако сумма слагаемых недостаточно велика для того, чтобы забраковать $H_1 : H_2$ в целом. Решение об истинной структуре цепи в этом случае принимается по частному переходу. Так, например, в вольфрамоносных гранитах Куроики (о-в Хонсю, Япония) в образце JYK-4 при состояниях Og, Q и Ab в однородной последовательности значение Δ для $H_1 : H_2$ было равно 18.89, т. е. решение было в пользу H_1 . В то же время частный переход через Ab дал значение χ^2 , равное 14.19, т. е. бракующее простоту перехода через Ab на 5%-м уровне.

VI.7. ОПЫТНАЯ ПРОВЕРКА СОСТОЯТЕЛЬНОСТИ СТАТИСТИЧЕСКИХ ТЕСТОВ О ПОРЯДКЕ МАРКОВОСТИ

Истинность заключения, полученного с помощью критериев из (VI.5) и (VI.6), проверяющих соответствие наблюдениям гипотез о марковском переходе в последовательности, зависит от соблюдения (как отмечалось во введении к этой главе) двух условий: последовательность должна быть однородна и ее марковский порядок должен быть конечен. Тогда, как следует из состоятельности критериев, при неограниченном увеличении длины последовательности правильное решение получается с вероятностью, стремящейся к единице. Однако, как было показано в гл. V, преобразование марковских цепей часто приводит к превращению их в существенно немарковские последовательности с бесконечным порядком марковости при сохранении однородности. При этом возникает вопрос, будет ли при последовательной проверке H_m против H_{m+1} ($m=0, 1, \dots$) всегда, начиная с достаточно большой выборки, приниматься решение в пользу H_{m+1} или возможны другие варианты. Вопрос этот пока не решен.

В настоящем параграфе мы рассмотрим опытным путем поведение статистических тестов, проверяющих порядок марковского

признака из § VI.5 для случая марковских и существенно немарковских последовательностей. Допустим, что мы исследуем конечное число пар альтернатив $H_0 : H_1, H_1 : H_2, \dots, H_{k-1} : H_k$ по одной выборке, соответствующей одной достаточно длинной реализации. Рассмотрим две ситуации. В первом случае выборка взята из последовательности, представляющей стационарную, эргодическую марковскую цепь порядка $r < k$. Во втором случае выборка взята из существенно немарковской последовательности. Эту последовательность мы полагаем стационарной (см. п. IV.2.2), а оценки матриц переходных вероятностей $P_{I;J}, P_{I;J;K}, \dots, P_{I;J,\dots;E;L}$, полученные по одной реализации, состоятельными, как это имело бы место для эргодической марковской цепи.

В обеих отмеченных ситуациях при проверке $H_{l-1} : H_l$ для всех l мы использовали одинаковый уровень значимости α (1 или 5%). В марковском случае мы говорим, что по выборке получено правильное решение, если гипотезы H_0, H_1, \dots, H_{r-1} бракуются, а гипотезы $H_{r+1}, H_{r+2}, \dots, H_{k-1}$ не бракуются. Как легко заметить, это соответствует корректному принятию решения, когда фактически порядок марковости есть r .

В существенно немарковском случае мы называем решение правильным, если все H_0, H_1, \dots, H_{k-1} последовательно бракуются. Это соответствует бесконечному порядку марковости, который «больше», чем $0, 1, \dots, k-1$.

При извлечении выборки из марковской цепи порядка r тест для альтернативы $H_{r-1} : H_r$ состоятелен (см. § VI.5). Это означает, что в альтернативе $H_{r-1} : H_r$ вероятность принятия правильного решения будет как угодно близка к единице, если обеспечено достаточно большое число наблюдений N . При этом возникает практический вопрос, можно ли обеспечить достаточно большое N в повседневной работе. N зависит прежде всего от специфики цепи. Очевидно также, что при прочих равных условиях N тем больше, чем выше порядок r цепи и чем больше в ней состояний.

В случае существенной немарковости нам неизвестно, будут ли вообще указанные тесты состоятельны. В этом случае неизвестно, гарантировано ли правильное решение даже тогда, когда объем выборки мог бы быть увеличен произвольно.

Излагаемые ниже результаты опытов показывают (для некоторых простейших случаев), каков должен быть объем выборки для обеспечения достаточно большой вероятности принятия правильного решения.

В качестве исходных были взяты две эргодические цепи Маркова первого порядка соответственно с матрицами переходных вероятностей и с векторами распределений начальных состояний:

$$P_{I,J} = \begin{matrix} & a & b \\ a & (0.75 & 0.25) \\ b & (0.50 & 0.50) \end{matrix}, P_I^{(0)} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \text{ и } P_{I,J} = \begin{matrix} & a & b \\ a & (0.75 & 0.25) \\ b & (0.25 & 0.75) \end{matrix}, P_I^{(0)} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}. \quad (\text{VI. 7. 1})$$

Для каждой из этих цепей моделировались по два варианта выборок объемом $N=1000$, $N=2000$, $N=4000$ и $N=6000$. Уже для выборок объемом $N=1000$ в каждом из всех четырех опытов принималось правильное решение. Иными словами, в последовательности альтернатив H_0 браковались, а H_1 , H_2 , H_3 и H_4 — не браковались. То же происходило для остальных выборок большого объема. Таким образом, в наших экспериментах одной тысячи наблюдений оказалось вполне достаточно для получения правильного решения при изучении выборок из простых цепей.

Полученные реализации использовались далее для перестройки их в реализации существенно немарковских последовательностей. Поскольку нельзя моделировать существенно немарковскую последовательность, задавая начальные распределения и последовательность матриц переходных вероятностей, мы провели это моделирование на основе теоремы V.6. Для этого брались все полученные ранее реализации и в каждую из них случайным образом, с постоянной вероятностью π добавлялись элементы состояния b , обозначавшиеся b^* ; b^* могли добавляться только между каждыми двумя смежными элементами b в первичной реализации. Получив указанным способом новую реализацию, мы укрупняли b и b^* , т. е. не различали эти элементы в окончательной последовательности, полагая $b' = b \cup b^*$. Из теоремы V.6 следует, что полученная реализация будет реализацией существенно немарковской последовательности независимо от того, каковы значения $0 < \pi < 1$ и матриц переходных вероятностей $P_{I, J}$ исходной стационарной эргодической простой цепи Маркова. Хотя существенная немарковость получается безотносительно к тому, большим или малым было взято π , следует ожидать, что для статистического теста отличить марковскую реализацию от немарковской тем труднее, чем меньше π . Соответственно в наших опытах каждая выборка преобразовывалась для двух значений π . При этом $\pi_1=0.1$ (малая интенсивность нарушения марковости) и $\pi_2=0.8$ (большая интенсивность нарушения марковости). Соответственно этому были получены 32 выборки из существенно немарковских последовательностей. Как для малого, так и для большого значения π ни одна выборка объема 1000, 2000 и 4000 не дала правильного решения (определение «правильного решения» см. на с. 375). Для выборок объема $N=6000$ при обоих значениях π было получено правильное решение. Таким образом, опыты показывают, что для корректной идентификации существенно немарковских последовательностей необходимо очень большое число наблюдений.

Оставалось неясным, вызывается ли необходимость очень большого числа наблюдений существенной немарковостью или обнаруженное явление характерно вообще для марковских цепей высоких порядков. Для выяснения вопроса мы промоделировали дополнительно четыре выборки из реализаций марковских цепей пятого порядка. Так как задание цепей пятого порядка рекур-

Существенно немарковские последовательности

Параметры исходной простой цепи	π	Номер n и объем исходной выборки	Значения тестовых статистик для проверки марковских гипотез				
			$H_0 : H_1$	$H_1 : H_2$	$H_2 : H_3$	$H_3 : H_4$	$H_4 : H_5$
			$\chi^2_{0.05(1)} = 3.84$	$\chi^2_{0.05(2)} = 5.99$	$\chi^2_{0.05(4)} = 9.49$	$\chi^2_{0.05(8)} = 15.51$	$\chi^2_{0.05(16)} = 26.36$
$P_{ij} = \begin{pmatrix} 0.75 & 0.25 \\ 0.50 & 0.50 \end{pmatrix}$	0.1	$n_1=1000$	65.8	0.5	3.7	3.3	14.3
		$n_2=1000$	127.9	6.0	0.0	0.9	0.7
		$n_1=2000$	87.8	0.5	4.6	6.8	30.1
		$n_2=2000$	105.3	4.0	6.9	16.2	29.3
		$n_1=4000$	271.0	2.7	6.0	6.9	19.6
		$n_2=4000$	202.9	2.5	9.5	6.9	20.0
		$n_1=6000$	290.6	28.3	33.5	100.1	98.4
		$n_2=6000$	242.8	54.6	59.7	165.0	131.7
	0.8	$n_1=1000$	153.1	12.9	18.7	5.7	22.9
		$n_2=1000$	149.7	16.5	28.3	10.8	24.0
		$n_1=2000$	291.8	55.7	41.9	11.9	23.9
		$n_2=2000$	102.0	1.7	4.3	7.6	34.9
		$n_1=4000$	656.8	88.9	89.6	24.6	25.9
		$n_2=4000$	648.7	96.0	133.7	35.7	20.3
$n_1=6000$		694.0	265.7	167.6	193.6	155.2	
$n_2=6000$		786.1	216.1	146.8	130.8	104.1	
$P_{ij} = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$	0.1	$n_1=1000$	502.8	0.4	5.2	9.8	8.2
		$n_2=1000$	269.5	1.35	2.85	3.96	9.60
		$n_1=2000$	457.0	0.5	9.1	2.4	10.7
		$n_2=2000$	457.9	1.6	8.5	4.4	24.6
		$n_1=4000$	1111.8	2.3	5.6	9.8	15.4
		$n_2=4000$	1077.4	3.1	5.1	7.2	32.3
		$n_1=6000$	16.3	27.4	21.9	112.8	57.2
		$n_2=6000$	12.0	13.1	17.7	68.4	36.6
	0.8	$n_1=1000$	423.8	4.2	16.0	9.8	15.4
		$n_2=1000$	435.1	8.4	9.5	3.9	10.9
		$n_1=2000$	573.8	0.3	8.8	1.5	11.0
		$n_2=2000$	836.1	16.3	35.1	9.1	22.3
		$n_1=4000$	182.1	39.4	40.4	31.7	25.7
		$n_2=4000$	108.5	99.6	48.7	21.3	22.8
$n_1=6000$		21.5	205.5	152.9	136.6	72.1	
$n_2=6000$		40.3	198.2	193.0	97.9	57.3	

Примечание. Здесь n в табл. VI-2 в скобках при χ^2 указано число степеней свободы, 0.05 — уровень значимости.

Цепи пятого порядка

Параметры исходной простой цепи	Номер и объем исходной выборки	Значения тестовых статистик для проверки марковских гипотез				
		$H_0 : H_1$	$H_1 : H_2$	$H_2 : H_3$	$H_3 : H_4$	$H_4 : H_5$
		$\chi^2_{0.05}(1) = 3.8$	$\chi^2_{0.05}(2) = 6.0$	$\chi^2_{0.05}(4) = 9.5$	$\chi^2_{0.05}(8) = 15.5$	$\chi^2_{0.05}(16) = 26.3$
$P_{ij} = \begin{pmatrix} 0.75 & 0.25 \\ 0.50 & 0.50 \end{pmatrix}$ $p_i^{(j)} = \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix}$	$n_1=1000$	585.2	69.4	87.7	118.3	185.7
	$n_2=1000$	574.6	71.3	90.1	122.5	192.7
	$n_1=2000$	1097.0	147.3	186.0	251.0	393.3
	$n_2=2000$	1104.7	144.6	182.0	244.8	380.5
$P_j = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$ $p_i^{(j)} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$	$n_1=1000$	590.9	86.1	113.7	163.5	313.3
	$n_2=1000$	571.9	88.9	117.9	169.6	326.2
	$n_1=2000$	1108.1	186.3	243.2	354.2	685.3
	$n_2=2000$	1116.1	184.1	241.3	350.7	676.5

рентным способом весьма громоздко, мы и здесь прибегали к преобразованию простых цепей. В данном случае была использована теорема V.8. В восьми выборках, из которых четыре имели объем 1000 и 2000 наблюдений, соответствовавших простым цепям Маркова с матрицами (VI.7.1), было произведено следующее преобразование. К каждой серии элемента b детерминировано добавлялось по 4 таких же элемента. В результате по теореме V.8 получалась реализация цепи Маркова пятого порядка. Опыты показали, что во всех восьми случаях с весьма убедительными значениями статистик принимаются правильные решения; корректная идентификация цепей высокого порядка достигалась уже при объеме выборки $N=1000$. Таким образом, создается впечатление, что для корректных выводов на основе изучения случайной последовательности опасность представляет существенная немарковость последовательности, а не ее высокая, но конечная марковость.

Изложенные выводы основываются на данных табл. VI-1 и VI-2.

Рассмотрены принципы проверки статистических гипотез однородности, обратимости и порядка марковости случайных последовательностей. Изложена техника проверки статистических гипотез о порядке марковости и однородности. Дан новый критерий для проверки гипотезы об обратимости марковских цепей, часто используемых в геологии. Показаны методы проверки гипотез

о частной марковости. Приведены новые результаты численных экспериментов на ЭВМ, показавшие возможные источники ошибок при работе с существенно немарковскими последовательностями.

Изложение иллюстрировано примерами из геологической практики. Примеры подробно разобраны с описанием деталей вычислений. Все это предостерегает от возможности ошибок в выводах, получаемых с помощью существующих критериев, и свидетельствует о необходимости проверки этих выводов различными методами. Подчеркивается важность использования концептуальных моделей.

Литература

- Беллман Р. Введение в теорию матриц. М., «Наука», 1969. 367 с.
Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М., «Наука», 1965. 464 с.
Ван-дер Варден. Математическая статистика. М., ИЛ, 1960. 437 с.
Кульбак С. Теория информации и статистика. М., «Наука» 1967. 408 с.
Леман Э. Проверка статистических гипотез. М., «Наука», 1964. 498 с.
Рао С. Р. Линейные статистические методы и их применение. М., «Наука», 1968. 547 с.
Уилкс С. Математическая статистика. М., «Наука», 1967. 632 с.
Anderson T. W., Goodman L. A. Statistical Inference about Markov Chains. — Ann. Math. Stat., 1957, v. 28, p. 89—110.
Billingsley P. Statistical Methods in Markov Chains. — Ann. Math. Stat., 1961, v. 32, p. 12—40.

УКАЗАТЕЛЬ ИМЕН

- Андерсон Р. С. 149
 Андерсон Т. (Anderson T. W.) 363
 Аренс Л. (Ahrens L. N.) 30, 33
- Байес 156
 Барнет 17
 Барра Ж. Р. 199
 Барт 110
 Башарин Г. П. 165, 166
 Беллман Р. (Bellman R.) 355
 Берксон 197
 Бёрк (Burke С. К.) 306
 Бернулли 36, 234—236
 Биллингсли П. (Billingsley P.) 363
 Бойд Ф. (Boyd F. R.) 221
 Боуэн Н. (Bowen N. L.) 21
 Бруш С. Г. (Brush S. G.) 16, 36
 Бюффон, де (Buffon de) 17, 21, 34, 143, 144
- Ваделл 115, 240
 Ван-дер-Варден Б. Л. 351, 354
 Вернадский В. И. 15, 24
 Вернер (Werner) 18
 Викман Ф. (Wickman F. E.) 33, 138, 285
 Винер Н. 134
 Вистелиус А. Б. 15—17, 32, 35, 60, 115, 121, 127, 129, 139, 205, 236, 243, 254, 287, 293, 294, 316, 318
 Вистелиус Н. Л. 8
 Вихерт А. В. 285
- Гантмахер Ф. Р. 254
 Геер, де 31
 Гейки А. 18
 Гельфанд И. М. 142
 Геснер Й. 17
 Гиббс В. (Gibbs J. W.) 23
 Гихман И. И. 92, 97, 126
 Гведенко Б. В. 7, 61, 62, 83, 128, 133
- Голубев В. С. 32
 Гольдшмидт В. (Goldschmidt W.) 24
 Гудмен Л. (Goodman L. A.) 363
- Дарвин Ч. (Darwin Ch.) 19
 Дасей М. (Dacey M. F.) 33
 Двали М. Ф. 8
 Дели Р. О. 21
 Демина М. Е. 7, 236, 300
 Джобсон (Jobson H. E.) 32
 Дынкин Е. Б. 131, 133
 Дюроше Ж. (Durocher J.) 21
- Жемчужников Ю. А. 7
- Заварицкий 110
 Зандбергер Ф. (Sandberger F.) 24
- Ибрагимов И. А. 125, 130
 Иванов Д. Н. 7, 111, 236, 243, 335
 Ито К. 93, 134
- Калицкий К. П. 21
 Калленберг О. (Kallenberg O.) 136
 Кан Т. (Kuhn T. S.) 17
 Кантор Г. (Kantor G.) 21
 Карно С. 36
 Кемени Д. 307, 308, 313, 315, 337
 Кендалл М. 59, 167, 176, 180, 181, 196, 197
 Кенуиль 167
 Кирван Р. (Kirwan R.) 17—19
 Кокс Д. 139
 Колмогоров А. Н. 7, 32, 59, 77, 91, 118, 122
 Коржинский Д. С. 23
 Кочрен У. (Cochran W. C.) 154
 Крамбайн У. (Krumbein W.) 25, 32, 33
 Крамер Г. 84, 86, 205—207
 Кротокин П. А. 21
 Кудрявцев Н. А. 21

- Кульбак С. (Kullbach S.) 165, 358, 366
Кьюнен (Kuenen Ph. H.) 31, 240
Лайель Ч. (Lyll Ch.) 19, 21, 24
Ландес К. (Landes K. K.) 21
Лаплас П. С. (Laplace P. S.) 17, 22
Латышев К. П. 7
Левинсон-Лессинг Ф. Ю. 25
Лейбниц Г. (Leubnitz G.) 16
Леман Э. 199, 365
Ли Б. (Lee B. K.) 32
Линдгрэн В. (Lindgren W.) 21
Майран, де (Mairan, de) 17
Маккин Г. 134
Марчисон Р. (Murchison R.) 19, 29
Мекки (Masky Wm.) 31
Миддлтон Г. В. 32, 35
Мизес Р. 119
Митчел (Mitchell J. P.) 32
Михайлов Н. Н. 7
Невё Ж. 70
Нейман Ю. (Neuman J.) 180, 181, 183, 201, 216
Ниггли П. (Niggli P.) 21, 110
Ньютон И. (Newton I.) 17, 22
Петров В. В. 207
Пирсон Э. (Pearson E.) 180, 183, 201, 216
Пламли У. (Plumley W. I.) 32
Плейфейер Д. (Playfair J.) 16, 19
Пляп (Plapp I. C.) 32
Подковыров В. Н. 32
Подольский Ю. В. 111
Разумовский Н. К. 25
Рао С. Р. (Rao C. R.) 83, 160, 167, 361, 371
Рейер Э. (Reyer E.) 25
Реймент Р. 16
Ричардсон У. (Richardson W. A.) 25
Розанов Ю. А. 125, 130
Розенблат А. (Rosenblatt A.) 306
Романова М. А. 7, 115, 129, 164, 236, 243, 287, 324, 328, 330, 361, 372
Сантало Л. (Santalo L.) 149
Сапогов Н. А. 7
Сарманов О. В. 32, 110
Севастьянов А. 18
Седжвик А. (Sedgwick A.) 19, 29
Сирот И. Э. 7
Скитович В. П. 83
Скороход А. В. 92, 97, 126
Смит В. 139
Смит Д. (Smith D.) 221
Смит У. (Smith W.) 18, 19
Снелл Дж. 307, 308, 313, 315, 337
Снезби Д. (Sneesby J.) 25
Стьюарт А. (Stuart A.) 59, 167, 176, 180, 181, 196, 197
Сугрובה Н. Г. 343
Тейлор 167
Тейт П. (Tait P. G.) 20
Томсон (Кельвин) В. (Thomson W.) 20, 22, 36
Тутубалин В. Н. 73
Уилкс С. (Wilks S.) 361, 365
Уильямс Д. (Williams J.) 19
Уиттен Е. Х. Т. (Whitten E. H. T.) 9, 16, 33
Фаас А. В. 7, 24, 110
Фаддеев Д. К. 259
Фаддеева В. Н. 259
Феллер В. 85, 139, 142, 236, 255, 294
Фишер Р. (Fisher R.) 24
Фихтенгольд Г. М. 87, 98
Фомин С. В. 59, 77
Фурье Ж. (Fourier J. B. J.) 20, 22, 36
Хаксли Т. (Huxley T. H.) 20
Харкер А. (Harker A.) 25, 110
Харламов Б. П. 7, 139, 294
Харрис Т. (Harris T. R.) 294, 328
Хаттон Д. (Hutton J.) 16, 18, 19
Хеннан М. 129
Хинце (Hinze J. O.) 32
Холанд 33
Холл Д. (Hall J.) 18, 19
Шварцхер В. (Schwarzacher W.) 33, 285
Шеннон К. (Shannon C.) 164
Шилов Г. Е. 142
Эйнштейн А. (Einstein A.) 134
Эйнштейн (Einstein H. A.) 32
Эль-Хиннави Е. Е. 138
Эммонс В. (Emmons W. H.) 21

УКАЗАТЕЛЬ ГЕОГРАФИЧЕСКИХ НАЗВАНИЙ

- Авача, вулк. 206
 Австралия 206
 — Восточная 26
 Азорские острова 206
 Алеутские острова 206
 Арга-Ынных-Хая, массив 243, 322,
 324, 328, 330, 361, 372
 Армения 206
 Афганистан
 Африка Южная 206
 Безымянный, массив 236
 Бразилия 322
 «Великан», гейзер 343
 Венгрия 206
 Гатино, массив 241, 243
 Гренландия Западная 206
 Гульшад, месторожд. 241, 244, 279,
 281
 Данбити, карьер 358, 366
 Долина гейзеров 343
 Доната, сел. 178
 Дракенберг, гора 206
 Жилой, о-в 314
 Земля Франца-Иосифа, о-в 26
 Иланлы, хр. 178
 Индокитай, п-ов 206
 Иран
 Иультин, месторожд. 243
 Йосемити, долина 323
 Кавказ северо-западный 228, 271
 Казахстан центральный 27
 Калифорния 206
 Камчатка 206, 236
 Каракульджур, р. 254, 263
 Каракумы Центральные 115
 Карменеллис, массив 243
 Кения 206
 Килауза, вулк. 206
 Кирмакинская долина 173
 Китай Северный 206
 Колыма, р. 241
 Кохала 206
 Кракатау, вулк. 206
 Куросики, массив 374
 Кызыл-Тас, массив 236, 241, 243
 Мадагаскар, о-в 206
 Мальсбург, массив 236, 241, 243
 Мангарева, о-в 206
 Маскаренские острова 206
 Мауна-Кеа, вулк. 206
 Мауна-Лоа, вулк. 206
 Мичиган, оз. 177, 178
 Монреаль, г. 16
 Новороссийск, г. 271
 Оаху, о-в 206
 Ока, месторожд. 241
 Омсукчан, сел. 243, 249, 338, 351
 Парижский бассейн 19, 24
 Парикутин, вулк. 206
 Прага, г. 16
 Реддинг, г. 33
 Реюньон, о-в 206
 Родезия Южная 206
 Русская платформа 52
 Саратовская область 127
 Силезия Нижняя 206
 Словакия 206

Суматра, о-в 206
Сьерра-Невада 236

Таити, о-ва 206
Тепловка, сел. 127
Тристан-да-Кунья, о-ва 206
Тянь-Шань Центральный 254

Урал Южный 228, 271
Усури, р. 206

Фарерские острова 206
Фудзияма, вулк. 206

Хаулаилаи, о-в 206
Хиллок, р. 206
Хонсю, о-в 374
Хуан-Фернандес, о-в 206

Чалбинская интрузия 236
Челекен, п-ов 287, 288, 314
Чукотка, п-ов 241

Швейцария 173—175
Швеция 16
Шотландия 18, 358, 366
Шпицберген, арх. 206

Эванстон, г. 178
Эран, о-в 18

Ява, о-в 206
Якутия 164, 206
— северо-восточная 236
Ян-Майен, о-в 206
Япония 206

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ ¹

- Аддитивность счетная 50
 Аксиоматика Колмогорова 48, 50, 51
 Алгоритм получения существенно немарковской последовательности 325—328
 Алмазы 35
 Альтернатива 180, 182—190
 — левосторонняя 188
 — правосторонняя 189, 190
 — простая 181
 — сложная 181, 183—191, 197, 203
 Аляскиты 254, 267
 Анализ дискриминантный 35
 Асимметрия 57, 112
 — левая 108, 109, 111—113
 — правая 108, 109, 111—113
 —, возникновение при процентном пересчете 112

 Бассейн седиментационный 139—142
 Биоценоз 19, 242
 Блуждание случайное 140—142

 Варв 31
 Вектор стохастический 251
 Величины случайные 45, 51
 — — векторные 76
 — —, их числовые характеристики 77—80
 — —, их числовые характеристики 56—59
 Вероятность 11, 41—44, 118, 119
 — апостериорная 45
 — априорная 45
 — безусловная 226
 — геометрическая 142—149
 — онтологическая 25
 — переходная 159, 164, 227
 — условная 44
 — эпистемологическая 25
 Воды ювенильные 21
 Возраст Земли 21, 36
 Выбор простой случайный 154, 155
 Выборка, объем 154, 197
 — случайная 153, 154

 Габбро-диабазы 241, 243, 281
 Гейзеры, интервалы между извержениями 343
 —, схема выбросов 245, 246
 Геология аналитическая 6, 15
 — математическая 6, 9, 10, 15, 34
 Гипотеза статистическая 179
 — — марковская, ее проверка 362—369
 — — о смеси нормальных распределений 207—212
 — — простая 181
 — — сложная 181, 197, 203
 Граниты 323
 — аплитовидные 241, 249, 250, 338, 351, 358, 359, 366
 — белые 324

 Диаграмма коннекционная 31
 Дисперсия 46, 57
 — обобщенная 86
 — энтропии Шеннона 171
 Доломит 51
 Дополнение к множеству 43

 Задача интегральная 24
 — локальная 24

¹ Курсивом выделены номера страниц, на которых даны развернутые сведения о предмете.

- Независимость причинная 27
 — случайных величин 47
 — стохастическая 27, 97, 155
 Неравенство Иенсена 167, 172
 — Коши—Буняковского 77
 — Чебышева 116, 117
 Несмещенность 157
 — критерия 196
 — — асимптотическая 196
 — — оценки 157, 158
 — — асимптотическая 158, 165—171
 Ниша экологическая 242
 Нормальность асимптотическая 165
- Область критическая 181
 — — наилучшая 182
 — — — равномерно 183
 — — уровня α 182—191
 Обратимость в марковских цепях 272, 273
 — — — — —, проверка 353—359
 — случайной последовательности 227, 229
 Объединение множеств 43
 Объем элементарный 110, 111
 Однородность в случайной последовательности 227, 228
 — — — — —, проверка 350—353
 Ожидание математическое 45, 46, 56, 157
 — — условное 90, 91
 Окаиты 241
 Окаймление, стохастическая схема 318, 319, 345, 346
 Организация среды 23—27
 Ординарность 62, 133
 Ошибка второго рода 153, 180
 — первого рода 153, 180
 Оценивание 157
 — интервальное 156
 — точечное 156, 157
 Оценка достаточная 158, 159
 — — точки усечения 175—178
 — информационной статистики 164, 165
 — максимума правдоподобия 159, 202
 — — — в марковских цепях 364, 365
 — минимума χ^2 160
 — несмещенная 157
 — параметра 157
 — переходных вероятностей при пропорциональности 159—164
 — по методу моментов 157, 160
 — состоятельная 158
 — точечная 157
 — энтропии Шеннона 167
 — — — несмещенная 171
- эффективная 157
 Оценки метода максимального правдоподобия 159, 160
- Пакеты 244, 289, 290, 298
 Парадигмы в геологии 17—22
 Парагенезисы элементов 10
 Парадокс Берксона 197
 Пегматиты 21
 Перекристаллизация песчаника 300
 Пересечение множеств 43
 Переход бернуллиевский 286
 — марковский 286
 — ограниченно марковский 289—291
 — частный марковский 286
 Период активности вулкана 138
 — в марковской цепи 240, 264
 Петрология 9
 Пироксен 221
 Плотность вероятности 51, 176
 — — Коши 69
 — — равномерная 64, 65, 178, 183—190
 Поле случайное 129
 — — пуассоновское 136, 137
 — — геометрических объектов 145
 — — точечное 135
 Пористость 127
 Порода карбонатная 51, 52
 Последовательности пакетов 343—346
 Последовательность Бернулли 27, 123, 124
 — гауссовская 125
 — ковариационная 268—272
 — марковская 124, 226—229, 230—242
 — — квазиэргодическая 258—260, 264
 — — обращенная 229
 — — эргодическая 238, 239
 — независимых испытаний 234—237
 — — случайных величин 122, 123
 — однородная 228
 — случайная 121, 129, 219, 224—227
 — — эргодическая 125—129
 — стационарная 125—129, 227
 — существенно немарковская 234, 246, 325—328, 332, 377
 Правило фаз 23
 Преобразование Лапласа 58, 294
 — — векторной случайной величины 80
 — марковских цепей 296
 — случайных величин 98—110
 — — — линейное 99—102
 — — — нелинейное 102—110

- Фурье 59, 294
- Признак марковский 291, 292, 311, 312
- Проекция распределения 75
- Произведение вероятностных пространств 70, 71
 - скалярное 265
- Пространство вероятностное 41, 42, 50, 71, 72, 224, 225
 - выборочное 159
 - евклидово R_1 46
 - евклидово R_n 71—76
 - измеримое 50, 71
 - элементарных событий 42, 43
 - — — дискретное 42
- Профилирование 218
- Процентный пересчет 106—113, 205, 212
- Процесс Винера 134
 - восстановления 138
 - при вулканической деятельности 138
 - гауссовский стационарный 130
 - детерминированный 26, 219
 - диффузионный 134
 - марковский 130—132
 - — однородный 131
 - Пуассона 132—134
 - случайный 27, 129
 - — точечный 62, 135—139
 - — — изотропный 136
 - — — , интенсивность 136
 - — — Пуассона 62
 - — — стационарный 136
 - с независимыми приращениями 132
- Размах 157
- Разность множеств 43
- Разрежение в случайной последовательности 298, 319, 320, 336—339
- Рандомизация 114, 115
- Распределение Бернштейна 106, 108, 109, 112, 113, 205
 - биномиальное 59—61, 170
 - зерен в шлихе 60, 177
 - вероятностей концентраций 153, 205—212
 - — начальное 226
 - — предельное в марковских цепях 253—262
 - — стационарное 252
 - — усеченное 176—178
 - — условное 89—94
 - — частное 121, 122
 - гамма 67—69
 - геометрическое 62
 - — для марковских цепей 63, 247, 248
- Копи 69
- — n -мерное 87—89, 101
- коэффициента спектральной яркости 115
- логарифмически нормальное 30, 32
- — нормальное 65—67
- — n -мерное 82—87
- — — вырожденное 84, 109
- — полиномиальное 80—82, 162
- Пуассона 61, 62
- равномерное 64, 65, 191
- серий в марковских цепях 246, 247, 248—250, 287, 288
- случайной величины 45
- — — абсолютно непрерывное 51, 63—69
- — — векторной 69, 72—76
- — — — марковское 98
- — — дискретное 51, 59—63
- — — маргинальное 73, 75
- — — смешанное 52
- — — условное 89—94
- χ^2 центральное 67—69, 166
- экспоненциальное 63
- — для марковских процессов 64
- Регрессия 95
- Ритм флишевый 31
- Свертка 100
- Свойство марковское 230, 231
 - — строго 133
- Сгущение в случайной последовательности 298, 319, 320
- Седиментология 9, 24
- Секрция латеральная 21
- σ -алгебра событий 48, 49, 51
 - — — борелевская 49—51, 121
- Смещение в двухходовых таблицах сопряженности 172
 - — трехходовых таблицах сопряженности 172
 - — информационного коэффициента корреляции 172
 - — информационных статистик 164
 - — оценок 157, 158, 172—175
 - — энтропии Шеннона 171
 - — статистического критерия 195, 196
- Событие случайное 219
 - — восстанавливающее 283—286
 - — независимое 44
 - — простое 43
 - — сложное 43
 - — элементарное 43
- Состояние в случайной последовательности 219, 224
 - невозвратное 239, 240
 - недостижимое 239

- пакетное 344—346
- периодическое 240
- поглощающее 239
- Состоятельность оценок 158
- статистических тестов 194, 195
- — о порядке марковости 374—378
- Спессартин 139
- Стандарт случайной величины 57
- Статистика 157, 158
- достаточная 158
- Δ 349
- математическая 11, 153
- χ^2 349—353
- χ^2_{α} 349, 354—356, 358, 359
- Статистики информационные 153, 165
- Стационарность 227, 228
- Степени свободы в марковских цепях 274—276, 351, 357, 358
- — для распределения χ^2 67
- Структура 23—29
- осуммита 260
- силикатов 261
- слюды 222, 223
- Суперпозиция функций распределения 114, 115
- Схема замещения стохастическая 297, 298, 340—343
- окаймления стохастическая 345, 346
- очагового зерна стохастическая, 325
- разрежения стохастическая 297, 298, 336—339
- сгущения стохастическая 297, 298, 329
- слияния стохастическая 339, 340
- стирания границ стохастическая 297, 298, 301, 302, 343—346
- увеличения серий стохастическая 332—335
- удаления состояний стохастическая 336—338
- чужеродного зерна стохастическая 325
- Таблица сопряженности 165, 173—175
- — двухходовая 165
- — трехходовая 165
- Таблицы сопряженности, их оценивание 165
- Тектоника 10
- Теорема Биркгофа—Хинчина 126
- Радоны—Никодима 91
- Теоремы предельные центральные 65, 66, 119, 120
- Теория Неймана—Пирсона 181, 183, 216
- укрупнения 298
- Толща красноцветная 287, 301
- Точка выборочная 160
- усечения 175—178
- — левая 175
- — правая 175—178
- Турбидиты 32
- Укрупнение по времени 297, 298
- — множеству состояний 297, 298, 302
- сильное 303, 304
- слабое 304, 306
- Униформизм 17
- Упаковка марковская трехмерная 292—294
- Упорядоченность 26
- Уравнение Колмогорова—Чепмена 131
- Уровень значимости 182, 197
- — критерия 198
- — — гарантированный 198
- Условия слабого укрупнения достаточные 307—314
- Устойчивость 350
- для марковских гипотез 374—378
- Флиш 9
- Формула Байеса 45
- Гейлора 167, 168
- Функция взаимной корреляции 128, 129
- вогнутая 172
- выпуклая 167
- гамма 67
- дисперсионная 106
- индикаторная (множества) 47
- корреляционная 127
- правдоподобия 159, 162
- производящая 58
- распределения случайной величины 45, 154, 155
- характеристическая 58, 59
- — векторной случайной величины 79, 80
- Цепи марковские 12, 98, 124, 250—264
- — второго порядка 230, 242—244
- — квазиэргодические 254, 259, 260, 262—264
- — неприводимые 239, 263
- — нулевого порядка 230, 234, 235

- — обратимые 238
- — однородные 233, 234
- — первого порядка 230
- — периодические 254, 261—263, 266, 273, 274
- — приводимые 239, 263
- — простые 230, 234, 235—242
- — сложные 230, 242—246, 276—283
- — — пятого порядка 378
- — стационарные 233, 234
- — эргодические 239, 253, 264
- Цикличность стохастическая 284, 285
- — в седиментологии 285
- Член остаточный в форме Коши 168
- Шум белый 130
- Экссесс 57
- Энтропия Шеннона 58, 164—166
- Эффективность асимптотическая 153
- Якобиан 87, 88, 98, 147, 149

Андрей Борисович Вистелнус

ОСНОВЫ МАТЕМАТИЧЕСКОЙ ГЕОЛОГИИ

(определение предмета, изложение аппарата)

Утверждено к печати

*Ленинградским отделением ордена Ленина
Математического института им. В. А. Стеклова*

Редактор издательства *Е. А. Семенова*

Художник *Д. С. Данилов*

Технический редактор *А. П. Чистякова*

Корректоры *Н. И. Журавлева, Г. И. Суворова,
К. С. Фридлянд и Т. Г. Эдельман*

ИБ № 8812

Сдано в набор 17.07.79. Подписано к печати 05.02.80.
М-29527. Формат $60 \times 90^{1/16}$. Бумага типограф-
ская № 1. Гарнитура обыкновенная. Печать высокая.
Печ. бл. $24\frac{1}{2} + 1$ вкл. ($\frac{3}{8}$ печ. л.) = 24.87 усл. печ. л.
Уч.-изд. л. 25.78. Тираж 2900. Изд. № 7086. Тип.
зак. 527. Цена 2 р. 90 к.

Издательство «Наука», Ленинградское отделение
199164, Ленинград, В-164, Менделеевская линия, 1

Ордена Трудового Красного Знамени
Первая типография издательства «Наука»
199034, Ленинград, В-34, 9 линия, 12

КНИГИ ИЗДАТЕЛЬСТВА «НАУКА»

Можно предварительно заказать

в магазинах конторы «Академкнига»

Для получения книг почтой

заказы просим направлять по адресу:

117192 Москва В-192, Мичуринский пр., 12

магазин «Книга — почтой»

Центральной конторы «Академкнига»;

197110 Ленинград П-110, Петрозаводская ул., 7

магазин «Книга — почтой»

Северо-Западной конторы «Академкнига»

или в ближайший магазин «Академкнига»,

имеющий отдел «Книга — почтой»;

Адреса и почтовые индексы магазинов:

- 480091 Алма-Ата, ул. Фурманова, 91/97 («Книга — почтой»);
- 370005 Баку, ул. Джапаридзе, 13;
- 320005 Днепропетровск, пр. Гагарина, 24 («Книга — почтой»);
- 734001 Душанбе, пр. Ленина, 95 («Книга — почтой»);
- 335009 Ереван, ул. Туманяна, 31;
- 664033 Иркутск, ул. Лермонтова, 289;
- 252030 Киев, ул. Ленина, 42;
- 252030 Киев, ул. Пирогова, 2;
- 252142 Киев, пр. Вернадского, 79;
- 252030 Киев, ул. Пирогова, 4 («Книга — почтой»);
- 277001 Кишинев, ул. Пирогова, 28 («Книга — почтой»);
- 343900 Краматорск (Донецкой обл.), ул. Марата, 1;
- 660049 Красноярск, пр. Мира, 84;
- 443002 Куйбышев, пр. Ленина, 2 («Книга — почтой»);
- 192104 Ленинград, Д-120, Литейный пр., 57;
- 199164 Ленинград, Таможенный пер., 2;

- 199034 Ленинград, 9 линия, 16;
220012 Минск, Ленинский пр., 72 («Книга — почтой»);
103009 Москва, ул. Горького, 8;
117312 Москва, ул. Вавилова, 55/7;
630076 Новосибирск, Красный пр., 51;
630090 Новосибирск, Академгородок, Морской пр., 22 («Книга — почтой»);
142292 Пушкино (Московской обл.), «Академкнига»;
620151 Свердловск, ул. Мамина-Сибиряка, 137 («Книга — почтой»);
700029 Ташкент, ул. Ленина, 73;
700100 Ташкент, ул. Шота Руставели, 43;
700187 Ташкент, ул. Дружбы народов, 6 («Книга — почтой»);
634050 Томск, наб. реки Ушайки, 18;
450059 Уфа, ул. Р. Зорге, 10 («Книга — почтой»);
450025 Уфа, Коммунистическая ул., 49;
720001 Фрунзе, бульв. Дзержинского, 42 («Книга — почтой»);
310003 Харьков, ул. Чернышевского, 87 («Книга — почтой»).

2 р. 90 к.

3196



«НАУКА»
ЛЕНИНГРАДСКОЕ
ОТДЕЛЕНИЕ