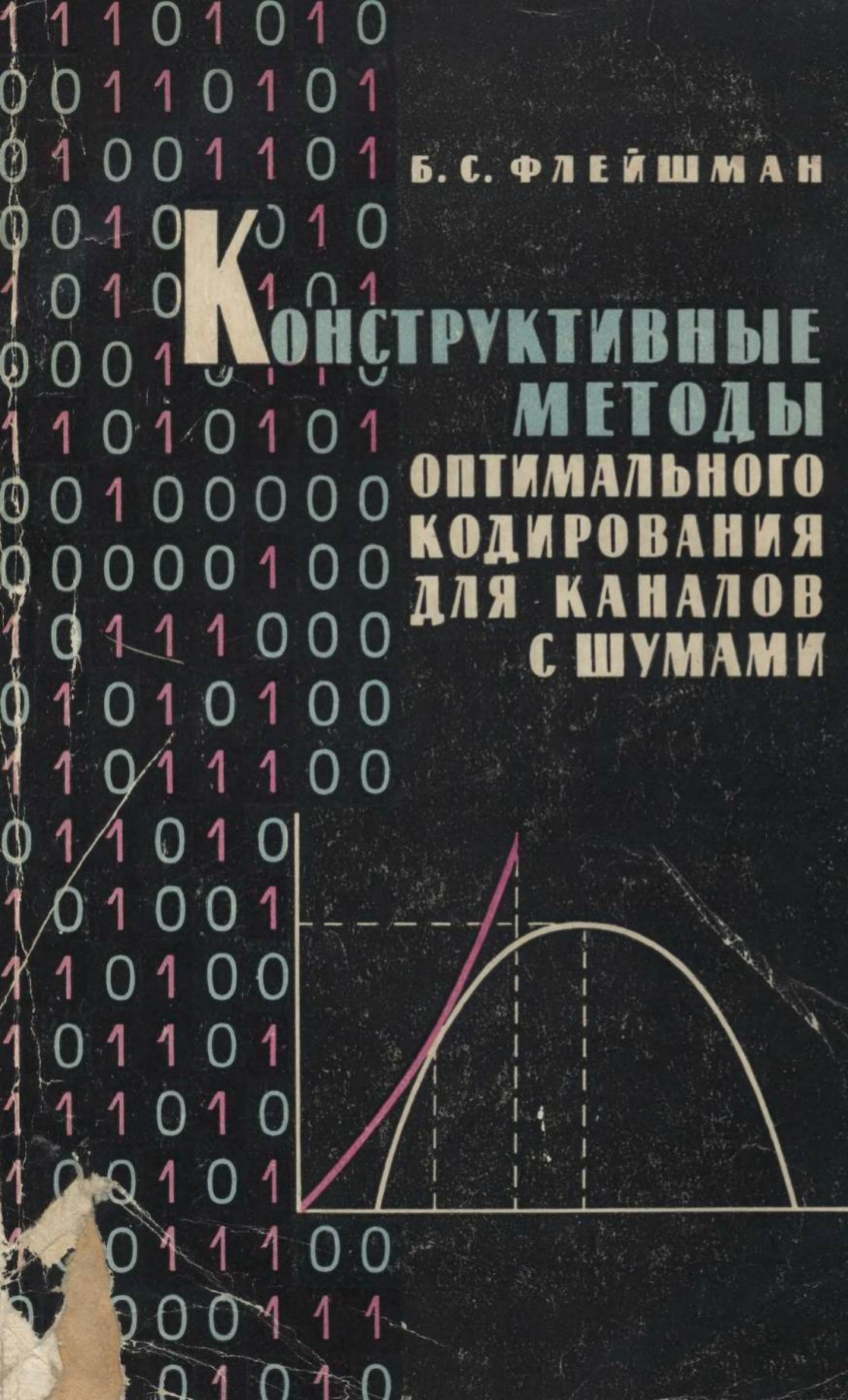
Б. С. ФЛЕЙШМАН

Конструктивные методы оптимального кодирования для каналов с шумами



АКАДЕМИЯ НАУК СССР
Институт радиотехники и электроники

Б. С. ФЛЕЙШМАН

КОНСТРУКТИВНЫЕ МЕТОДЫ ОПТИМАЛЬНОГО КОДИРОВАНИЯ ДЛЯ КАНАЛОВ С ШУМАМИ

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР
Москва 1963

В книге систематически излагаются конструктивные методы построения оптимальных кодов для каналов с шумами, в то время, как основная теорема Шеннона для каналов с шумами и дальнейшие ее обобщения лишь доказывают существование таких кодов, но не указывают их конструкции. Для построения оптимальных кодов развивается новый комбинаторный аппарат. Оцениваются технические возможности реализации оптимального кодирования, связанные с объемом памяти и быстродействием электронных машин. Формально вводимые в теории информации понятия энтропии, скорости передачи, пропускной способности и др. возникают в книге по ходу решения задач, как некоторые «характерные» константы.

Книга предназначена для научных работников и инженеров, работающих в областях, использующих идеи и методы общей теории связи.

ОТВЕТСТВЕННЫЙ РЕДАКТОР

член-корреспондент АН СССР В. И. Сифоров

Флейшман Бенцион Семенович

Конструктивные методы оптимального кодирования для каналов с шумами

*

Утверждено к печати
Институтом радиотехники и электроники АН СССР

Редактор Издательства *В. Ф. Ржевский.* Технический редактор *Т. В. Полякова*

РИСО АН СССР № 69—82В. Сдано в набор 29/XII 1962 г. Подписано к печати 9/IV 1963 г. Формат 70×108 1/16. Печ. л. 14=19,18 усл. печ. л. Уч.-изд. л. 15,1 Тираж 4200 экз. Т-02374. Изд. № 1432. Тип. зак. № 5493

Цена 1 р. 06 к.

Издательство Академии наук СССР. Москва, К-62, Подсосенский пер., 21

2-я типография Издательства АН СССР. Москва, Г-99, Шубинский пер., 10

Содержание

Предисловие

В настоящее время происходит бурное развитие радиоэлектроники, новых средств связи и управления, создаются большие электронные машины дискретного действия, а также сложные радиотехнические системы, элементы которых отличаются большой сложностью. Вместе с прогрессом техники преобразуются старые и возникают новые математические теории и дисциплины, которые составляют пока очень разрозненный математический аппарат кибернетики.

Темп жизни середины XX века, когда родилась кибернетика, существенно отличается от сравнительно размеренного и спокойного ритма XVII века, когда произошла другая научная революция, связанная с рождением ньютоновской механики и ее математического аппарата дифференциального и интегрального исчисления. Тогда рождение новой науки произошло после формулировки основных ее законов и после создания основ адэкватного математического аппарата. В настоящее время еще не сформулированы общие законы кибернетики, и она не имеет адэкватного математического аппарата. Речь идет о формулировке общих математических законов кибернетики, аналогичных законам механики и физики, а не о тех разрозненных результатах и общефилософских высказываниях, которыми изобилует современная кибернетическая литература.

Прообразом общих законов кибернетики могут служить общие законы физики. Несводимость первых ко вторым полностью осознана в настоящее время. Понимаемая в собирательном смысле слова как точная наука о природе физика в настоящее время является единственной точной естественной наукой. Ее поистине триумфальное шествие, начиная с создания ньютоновской механики и до величайших успехов современной физики, привело к безраздельному господству физических представлений и методов во всем современном естествознании. Проникновение точных количественных методов физики в заповедники качественного естествознания — химию и биологию еще не завершено, но для него здесь нет каких-либо принципиальных преград.

Однако в настоящее время все более осознается двуединость широкой области бурно прогрессирующей новой машинной техники и проблем «загадок» жизни сложных биологических автоматов, где физические представления и методы недостаточны. Это обстоятельство вполне естественно. В самом деле для физики характерно постепенное все более глубокое проникновение в скрытый механизм явлений индеферентной природы. Основы физики не заключают в себе возможности изучения и регулирования целенаправленной деятельности сложных автоматов, тем более изучения антагонистических ситуаций, когда проявляются ощущения индивидуальности (чувства «я») сторон.

Точная математическая формулировка и решение всех этих новых нефизических проблем составляют предмет кибернетики. Последняя

станет наукой в современном смысле слова, когда подобно физике она сформулирует свои общие законы на языке математики.

Один и тот же сложный технический или биологический объект можно изучать с физической и кибернетической точки зрения. Физика сосредоточивает внимание на вещественных и энергетических носителях объекта, а кибернетика — на чисто функциональной его схеме, связанной с целенаправленным поведением. В настоящее время имеют место оба указанных рассмотрения. Лишь их совместное развитие может открыть целостную картину мира.

Однако, по мнению автора, в настоящее время для общего прогресса исследований сложных автоматов было бы важным выделение и обособление чисто кибернетической проблематики и даже противопоставление ее физической проблематике. В самом деле, упоминавшееся безраздельное господство физических представлений и методов, не адэкватных кибернетике, пагубно сказывается в последнее время на развитии последней.

Характерным примером этому может служить современное состояние теории информации. Возникшая благодаря гениальной интуиции Шеннона, исходившего из физических предпосылок измерения величины «информации» — этого «флогистона» кибернетики, теория информации в настоящее время испытывает кризис неадэкватности ее физических представлений и аппарата ее кибернетической сущности.

Современное состояние теории информации неудовлетворительно в двух пунктах. Во-первых, основные ее результаты, относящиеся к оптимальному кодированию для каналов с шумами, носят характер теорем существования, не указывая способов построения оптимальных кодов. Во-вторых, существует неестественный разрыв между сравнительно недавно возникшей теорией информации и более ранними эффективными статистическими методами оптимального выбора между гипотезами.

Терминология шеноновской теории информации засорена физическими терминами — «энтропия», «скорость передачи», «пропускная способность», которые создают неверные физические ассоциации и запутывают и без того сложную ситуацию, рассматриваемую в теории информации[1]. Но, конечно, основным тормозом развития теории информации является не адэкватный ей физический аппарат, вследствие чего все ее основные результаты оказываются неконструктивными.

В настоящей книге выделяется чисто кибернетическая часть теории информации и для нее строится адэкватный комбинаторный аппарат. При этом результаты носят статический характер теории выбора между гипотезами в отличие от современного динамического характера теории информации с обязательными «*бит/сек*».

В первой части книги развивается новый математический аппарат, с помощью которого во второй части проводится построение оптимального кода для общего случая дискретного канала с независимыми шумами. Выясняется статистический смысл основных понятий теории информации. Подробно рассматриваются варианты общего алгоритма построения оптимального кода для канала с высоким и низким уровнями шумов.

Построение ведется случайным выбором слов фиксированной длины из определенной генеральной совокупности. Определяется вероят-

[1] Русская терминология теории информации усугублена еще и неточным переводом английского термина «rate» как «скорость», что придает ему неоправданную физическую окраску.

ность получения при этом оптимального кода. Показано, что последняя с ростом длины выбираемых слов столь быстро стремится к единице, что практически предлагаемый алгоритм всегда приводит к цели. Известные регулярные методы получения случайных чисел приводят к регулярному варианту предлагаемого алгоритма. Показано, что использование последнего позволяет существенно сократить объем памяти на входе канала. Это обстоятельство делает более реальными перспективы технического осуществления оптимального кодирования.

Для овладения материалом книги не требуется каких-либо математических знаний, выходящих за рамки радиотехнического вуза. Книга имеет целью привлечь внимание научных работников и инженеров к новым возможностям осуществления оптимального кодирования для каналов с шумами.

В зависимости от склонностей читателя можно рекомендовать различное чтение книги: для теоретиков — в порядке, принятом в книге, а для практиков — сразу гл. 8, где подробно развита теория для простейшего случая бинарного симметричного канала, с последующим ознакомлением со второй частью, пользуясь по желанию ссылками на первую часть.

В целях облегчения пользования формульным текстом в конце книги дан список основных математических обозначений, принятых в книге.

В заключение не могу не выразить особой благодарности доктору физико-математических наук А. С. Монину, который после ознакомления с первыми набросками моей работы оценил значение комбинаторного подхода в разработке теории оптимального кодирования. Автор благодарен члену-корреспонденту АН СССР В. И. Сифорову, доктору технических наук В. И. Бунимовичу и кандидату технических наук В. Д. Зубакову за ценные замечания в процессе работы над книгой.

Введение

Формированию современных представлений общей теории связи предшествовал период физических исследований прохождения сигналов по каналам с шумами. При этом предлагались разнообразные способы подавления шумов, однако не были ясны общее направление возможных усовершенствований и их пределы.

Впервые общие вероятностные принципы приема сигналов при наличии шумов были сформулированы в 1946 г. В. А. Котельниковым в его докторской диссертации [1]. В ней существенно использовалось представление шумов и сигналов n-мерными векторами для построения так называемого идеального приемника, осуществляющего правильный прием $M \leqslant n$ равновероятных, ортогональных сигналов с максимальной вероятностью P.

В 1947 г. Шенноном [2] был рассмотрен важный предельный случай экспоненциального растущего при $n \to \infty$ числа сигналов $M = e^{Cn}$. Для широкого класса каналов найдено выражение максимально большого значения C (пропускная способность), при котором вероятность P еще стремится к единице. Доказано существование оптимальных кодов, обеспечивающих достижение предельно возможного C.

Однако в отличие от работы В. А. Котельникова [1], где давалось построение идеального приемника, работа Шеннона [2] и ряд работ, последовавших за ней, носили неконструктивный характер, когда шла речь об эффективном построении оптимальных кодов. Это направление общей теории связи получило название теории информации[1].

Живой интерес к теории информации и интенсивное развитие ее в первые годы после опубликования работы Шеннона [2] сменились в последние годы заметным охлаждением интереса к этому направлению общей теории связи. Такое изменение отношения к теории информации объяснялось двумя обстоятельствами.

Во-первых, многолетние попытки получения эффективных способов построения оптимальных кодов были неудачны. Это давало основание предполагать, что получение оптимальных кодов математическими средствами, которыми до того обходилась теория информации, бесперспективно[2].

Во-вторых, оказалось возможным доказывать[3] существование оптимальных кодов, основываясь не на новых понятиях теории информации (энтропии, скорости передачи, пропускной способности и т. д.), а на ранее известных понятиях математической статистики (см. далее, гл. 5).

---

[1] В американской литературе теорией информации называется широкая область теории связи, в которой преобладают вероятностные методы исследования.

[2] Следует отметить, что для построения неоптимальных корректирующих кодов использовался весьма развитой математический аппарат теории чисел, теории групп и в последнее время теории конечных полей Галуа.

[3] Такого рода доказательство для канала с аддитивным белым шумом, основанное на геометрических рассмотрениях, содержится в одной из первых работ Шеннона [3].

Это обстоятельство ставило под сомнение целесообразность введения новых понятий теории информации.

Конечно, трудно переоценить значение пионерских работ Шеннона [2, 3], содержащих фундаментальные результаты, полученные благодаря редкой интуиции их автора.

Однако описанное состояние теории информации заставляло вернуться к исходным конструктивным идеям теории потенциальной помехоустойчивости В. А. Котельникова и развивать новые математические методы для построения оптимальных кодов.

Ввиду трудности такого рода рассмотрений естественно было начинать их с предельно простого случая бинарного симметричного канала. Построение автором [4—6] кода, оптимального в смысле Шеннона, в этом простейшем случае действительно потребовало развития нового математического аппарата, который оказался комбинаторным [7]. Дальнейшие исследования велись в направлении построения оптимального по Шеннону кода для общего случая дискретного канала с шумами. Это требовало дальнейшего развития комбинаторного аппарата. Данная книга посвящена систематическому изложению указанных результатов автора с полными доказательствами.

В первой части развивается комбинаторно-вероятностный математический аппарат, используемый во второй части.

В главе 1 подробно изложены результаты комбинаторики расположений. Основным из них является вычисление числа элементов в пересечении «комбинаторных сфер» (основная лемма о пересечениях, п. 1, 7, 2).

Глава 2 посвящена развитию методов теории стохастических функций, первоначально исследованных А. А. Марковым [8] и В. И. Романовским [9]. Матричная символика позволяет получить компактные соотношения связи производящих функций и моментов стохастического аргумента и функции. Рассмотрения ограничиваются случаем дискретной стохастической зависимости с независимыми переходами.

Глава 3 содержит в основном известные результаты математической статистики (теории оптимального выбора между гипотезами), необходимые для дальнейшего изложения.

Вторая часть книги начинается главой 4, содержащей основные понятия теории оптимального кодирования и формулировки задач.

В главе 5 доказываются теоремы существования оптимальных кодов для дискретного канала с независимыми шумами с помощью статистических результатов гл. 3. При этом приводится эффективное построение процедуры оптимального декодирования.

Основные результаты книги сосредоточены в гл. 6 и 7. Здесь существенно используются результаты гл. 1 и 2 для построения процедур оптимального кодирования и декодирования. Используя основную лемму о пересечениях, в гл. 6 формулируются условия, достаточные для оптимальности соответствующего кода.

В главе 7 дана оценка вероятности того, что этим условиям удовлетворяет код, являющийся случайной выборкой из определенной генеральной совокупности. Рассматривается процедура получения оптимального кода случайным выбором с последующим выбрасыванием части «плохих» кодовых слов.

Показано, что вероятность P' получения такими способами оптимальных кодов с ростом длины кодовых слов n существенно быстрее стремится к единице, чем, например, вероятность P правильного декодирования. Таким образом, практически указанные процедуры всегда приводят к цели. Далее сопоставляются полученные результаты с тео-

ремами существования оптимальных кодов теории информации для соответствующего дискретного канала [10].

В главе 8 рассмотрены особенности оптимального кодирования для простейшего случая бинарного симметричного канала, а в главе 9 — технические возможности осуществления оптимального кодирования.

В дополнениях I и II рассмотрены вопросы, возникающие при решении задачи осуществления оптимального кодирования и в ряде родственных ей задач.

В целом полученные результаты относятся к направлению, которое можно назвать конструктивной теорией информации. Дальнейшее развитие этого направления, по мнению автора, связано с распространением полученных результатов на случай каналов с зависимыми шумами.

Теория информации является наиболее развитой и в своем неконструктивном варианте вполне сложившейся областью кибернетики. Описанное ее состояние характерно и для других областей кибернетики.

В самом деле, неэффективность большинства результатов имеет место и в таких областях, как теория игр, динамическое и линейное программирование и др. В связи с этим в указанных новых областях отказываются от поиска «формульных» решений, ограничиваясь алгоритмическими решениями, доставляющими численные результаты с помощью машинного счета.

Однако никакие численные результаты не могут заменить формульных решений, наиболее полно концентрирующих в себе количественную картину явления. Поэтому успехи использования вычислительных машин не снижают актуальности поисков адекватного кибернетике математического аппарата. Примечателен тот факт, что и в других областях кибернетики, например линейном программировании, ощущается необходимость использования именно комбинаторных методов.

По мнению автора, в связи с возникновением кибернетики создание новых дискретных, комбинаторных методов имело бы такое же значение, как и возникновение анализа бесконечно малых в XVII веке в связи с развитием механики и физики.

Часть I

АППАРАТ КОНЕЧНОЙ МАТЕМАТИКИ

Глава 1

КОМБИНАТОРИКА РАСПОЛОЖЕНИЙ

§ 1.1. Вводные замечания

Комбинаторика, в отличие от других областей математики, не выделилась до сих пор в самостоятельную ветвь этой науки в современном смысле слова. Однако, комбинаторные факты накапливались и накапливаются достаточно быстро как в математике, так и в ее физических и, особенно, кибернетических приложениях. Современная комбинаторика состоит из большого числа разнородных задач, решаемых различными комбинаторными приемами, в которых трудно уловить руководящие идеи и методы. Это относится и к трем известным в мировой литературе монографиям по комбинаторике [11—13], из которых последняя является наиболее удачной.

В этой главе выделен сравнительно меньший, чем в указанных монографиях[1], круг комбинаторных задач, связанных с расположением предметов на определенных местах [7]. При этом удается построить указанную часть комбинаторики на аксиоматической основе. Следует сразу же оговорить, что без введения теоретико-множественной аксиоматики было бы затруднительно получить необходимые для дальнейшего изложения новые комбинаторные факты. Последние связаны с определением пересечений «комбинаторных сфер», что имеет решающее значение для всей теории оптимального кодирования.

Большинство комбинаторных построений этой главы развито специально для теории оптимального кодирования. Однако формулируемые по необходимости в общей форме эти построения могут оказаться полезными и для других областей кибернетики (см. § 9.2).

Перейдем к описанию содержания гл. 1. В § 1.2 приводится аксиоматика комбинаторики расположений предметов $A=(A_1,\ldots, A_\alpha,\ldots, A_a)$ на s местах, дающая стандартные методы подсчета числа элементов сложных множеств расположений по числам элементов простых множеств, через которые они выражаются. В первых же примерах использования введенной аксиоматики в § 1.3 обнаруживается неэффективность элементарных методов подсчета числа элементов сложных множеств расположений. Здесь же вводятся основные для всего дальнейшего изложения элементарные множества расположений $\mathscr{C}_\sigma^{\overline{m}}$, означающие все расположения с фиксированной частотой $\overline{m}=(m_\alpha)$ предметов A_α.

В § 1.4 результаты § 1.3 используются для классического определения ряда вероятностей, важных для дальнейшего изложения.

§ 1.5 посвящен описанию общих свойств и конкретной структуры производящих функций чисел и вероятностей множеств расположений. Особое внимание уделяется производящим функциям производящих функций (коротко вторым, третьим и т. д. производящим). Важная роль последних обнаруживается и в дальнейшем изложении в гл. 2.

[1] Здесь не рассматриваются, например, комбинаторные задачи теории графов [14], комбинаторной топологии [15] и линейного программирования [16].

В § 1.6 проводятся асимптотические при $s \to \infty$ рассмотрения чисел и вероятностей множеств расположений. При этом сами собой возникают характерные энтропийные выражения, связанные с формально вводимой в теории информации h-функцией. Здесь же описаны малоизвестные (не нормальные) асимптотические приближения полиномиального распределения.

Основной результат содержится в § 1.7, где подробно изучаются пересечения «комбинаторных сфер» — множеств расположений, отстоящих от данного расположения на фиксированном «матричном» расстоянии. Этот результат по сути дела является ключом к решению всей проблемы построения оптимального по Шеннону кода для рассматриваемого канала.

§ 1.2. Аксиоматика

1.2.1. *τ-проекции и расположения.* Рассмотрим два конечных множества

$$A = \{A_1, \ldots, A_\alpha, \ldots, A_a\} \text{ и } \sigma = \{1, 2, \ldots, t, \ldots, s\},$$

состоящих из различных элементов.

Элементы $A_\alpha (\alpha = \overline{1,a})$ будем называть предметами и обозначать[1] соответствующими индексами α. Элементы $t (t = \overline{1,s})$ будем называть местами и считать упорядоченными. Соответственно множество A назовем множеством предметов, а множество σ — множеством мест.

Наряду с упорядоченным множеством мест σ рассмотрим множество Σ всевозможных его подмножеств $\tau = \{i_1, \ldots, i_k, \ldots, i_t\} \subseteq \sigma$, в которых элементы i_k упорядочены так же, как и в σ. В Σ удобно включить и пустое множество, обозначаемое $\varnothing$. Очевидно, что множество Σ снова конечно. Во всех дальнейших рассмотрениях множества A и σ произвольны, но фиксированы.

Отобразим некоторые (не обязательно все) элементы $A_\alpha \in A$ в элементы множества τ. От этого отображения $\rho(A) = \tau$ будем требовать однозначности лишь в одну сторону так, чтобы каждому из элементов $i_k \in \tau$ соответствовал один и только один элемент A_α, но каждому из отображаемых элементов A_α может, вообще говоря, соответствовать несколько элементов $i_k \in \tau$. Отображение $\rho(A) = \tau$ будем называть τ-проекций и коротко обозначать ρ_τ.

Используя пустое множество $\varnothing \in \Sigma$, можно всегда формально считать, что все множество A отображается в множества $\varnothing$ и $\tau \subseteq \Sigma$, так что та часть элементов A, которая не отображается в τ, отображается в $\varnothing$. Ясно, что отображение A в τ есть вместе с тем отображение A в σ, ибо $\tau \subseteq \sigma$.

Отображение множества A в множество σ назовем расположением и будем обозначать ρ. Ясно, что расположение является частным случаем τ-проекции, когда $\tau = \sigma$.

Множество всех расположений ρ назовем пространством расположений и будем обозначать R. Множество всех τ-проекций ρ_τ назовем пространством τ-проекций и будем обозначать R_τ. Ясно, что R и R_τ являются конечными множествами, которые можно отобразить в s-мерный и t-мерный целочисленные кубы со стороною a соответственно (последние содержат a^s и a^t целочисленных точек).

В качестве примера рассмотрим случай $a = 1$ и $s = 3$. Здесь $A = \{A_1\}$ и $\sigma = \{1, 2, 3\}$.

[1] Обозначение $\alpha = \overline{1, a}$ означает $\alpha = 1, 2, \ldots, a$.

На рис. 1.1 схематически стрелками показаны всевозможные отображения A в σ. Из них лишь последнее является расположением, а все остальные — τ-проекции.

Заметим, что если добавить к множеству предметов $A = \{A_1\}$ еще один элемент A_2, т. е. рассматривать множество предметов $A = \{A_1, A_2\}$, и

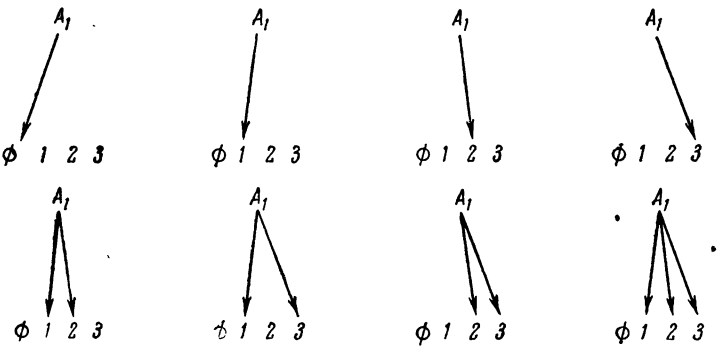


Рис. 1.1. Множество всех τ-проекций для случая s-3

отобразить элемент A_2 во все элементы множества $\sigma - \tau$ дополнения τ (рис. 1.1), то получим (рис. 1.2) множество всевозможных расположений $A = \{A_1, A_2\}$ на σ. Таким образом, множество Σ всевозможных τ-проекций $A = \{A_1\}$ на σ взаимно однозначно отображается в пространство R всех

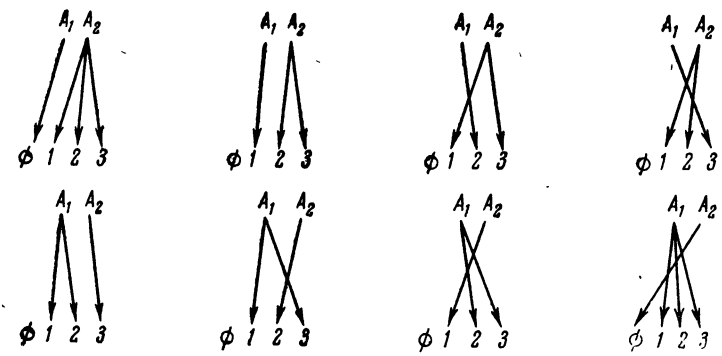


Рис. 1.2. Отображение множества всех τ-проекций на множество всех расположений с добавлением одного предмета

расположений $A = \{A_1, A_2\}$ на $\sigma = (1, 2, 3)$, отсюда следует их равночисленность.

Очевидное обобщение приведенных выше рассмотрений на случай произвольных конечных a и s устанавливает, что множество Σ всевозможных τ-проекций $A = \{A_1, \ldots, A_a\}$ на $\sigma = (1, \ldots, s)$, взаимно однозначно отображается в пространство R всех расположений $A = (A_1, \ldots, A_{a+1})$ на $\sigma = (1, \ldots, s)$.

Отсюда следует, что множество Σ состоит из $(a+1)^s$ элементов. Если $\tau = \{i\}$ состоит из одного места, на которое отображен предмет A_α, то $\{i\}$-проекцию $\rho_{\{i\}} = \rho^{\alpha}_{\{i\}}$ будем называть элементарной τ-проекцией. τ-проекцию ρ_τ можно записать в виде таблицы

$$\rho_\tau = \begin{Bmatrix} \alpha_{i_1}\alpha_{i_2} \ldots \alpha_{i_k} \ldots \alpha_{i_t} \\ i_1\ i_2 \ldots i_k \ \ldots\ i_t \end{Bmatrix} = \begin{Bmatrix} \alpha_{i_k} \\ i_k \end{Bmatrix}, \tag{1.1}$$

столбцы которой, следуя порядку мест, указывают соответствующие предметы. Другая запись ρ_τ упорядочена по предметам

$$\rho_\tau = \{\tau_1, \tau_2, \ldots, \tau_\alpha, \ldots, \tau_a\} \left(\bigcup_{\alpha=1}^{a} \tau_\alpha = \tau\right), \qquad (1.1')$$

где τ_α означает подмножество мест, в которые отображается один и тот же предмет A_α.

В частности, расположение ρ имеет двойное представление:

$$\rho = \begin{Bmatrix} \alpha_1 \ldots \alpha_t \ldots \alpha_s \\ 1 \ldots t \ldots s \end{Bmatrix} = \begin{Bmatrix} \alpha_t \\ t \end{Bmatrix}$$

и

$$\rho = \{\tau_1, \tau_2, \ldots, \tau_\alpha, \ldots, \tau_a\} \left(\bigcup_{\alpha=1}^{a} \tau_\alpha = \sigma\right). \qquad (1.2)$$

1.2.2. *Операции над множествами τ-проекций.* Пусть $\mathscr{E}_\tau \subset R_\tau$ произвольное множество τ-проекций ρ_τ. В дальнейшем изложении используются следующие основные общеупотребительные операции над множествами $\mathscr{E}'_\tau$, $\mathscr{E}''_\tau \subset R_\tau$.

Объединение (сумма) $\mathscr{E}'_\tau$ и $\mathscr{E}''_\tau$

$$\mathscr{E}'_\tau \bigcup \mathscr{E}''_\tau = \mathscr{E}_\tau \subset R_\tau$$

и пересечение (взятие общей части) $\mathscr{E}'_\tau$ и $\mathscr{E}''_\tau$

$$\mathscr{E}'_\tau \bigcap \mathscr{E}''_\tau = \mathscr{E}'''_\tau \subset R_\tau.$$

Если $\mathscr{E}'_\tau$ и $\mathscr{E}''_\tau$ не содержат общих элементов, то мы будем говорить, что их пересечение пусто, и обозначать это так

$$\mathscr{E}'_\tau \bigcap \mathscr{E}''_\tau = \varnothing.$$

Далее, для $\mathscr{E}_{\tau_1} \subset R_{\tau_1}$ и $\mathscr{E}_{\tau_2} \subset R_{\tau_2}$, где $\tau_1 \bigcap \tau_2 = \varnothing$, определим операцию прямого произведения

$$\mathscr{E}_{\tau_1} \times \mathscr{E}_{\tau_2} = \mathscr{E}_{\tau_1 \bigcup \tau_2}.$$

Эта операция заключается в образовании из каждой пары τ-проекций $(\rho_{\tau_1}, \rho_{\tau_2})$, где $\rho_{\tau_1} \in \mathscr{E}_{\tau_1}$ и $\rho_{\tau_2} \in \mathscr{E}_{\tau_2}$, новой «составной» $\tau_1 \bigcup \tau_2$-проекции $\rho_{\tau_1 \bigcup \tau_2} \in \mathscr{E}_{\tau_1 \bigcup \tau_2}$ с общим упорядочением мест $\tau_1 \bigcup \tau_2$ согласно упорядочению в σ. Например, если $\rho_{\tau_1} = \begin{pmatrix} A_3 A_3 A_2 \\ 2\ 8\ 9 \end{pmatrix}$ и $\rho_{\tau_2} = \begin{pmatrix} A_3 A_1 A_1 \\ 1\ 5\ 6 \end{pmatrix}$, то

$$\rho_{\tau_1 \bigcup \tau_2} = \begin{pmatrix} A_3 A_3 A_1 A_1 A_3 A_2 \\ 1\ 2\ 5\ 6\ 8\ 9 \end{pmatrix}.$$

Разность $\mathscr{E}_3$ между двумя множествами произвольной структуры $\mathscr{E}_1 - \mathscr{E}_2 = \mathscr{E}_3$ понимается как множество, которое при объединении с $\mathscr{E}_2$ дает $\mathscr{E}_1$ ($\mathscr{E}_2 \bigcup \mathscr{E}_3 = \mathscr{E}_1$).

Рассмотренные операции объединения, пересечения и прямого произведения для двух множеств $\mathscr{E}_\tau$ легко обобщаются на k таких множеств. Удобны

следующие обозначения:

$$\bigcup_{l=1}^{k} \mathscr{E}_\tau^{(l)} = \mathscr{E}_\tau^{(1)} \cup \mathscr{E}_\tau^{(2)} \cup \ldots \cup \mathscr{E}_\tau^{(k)} = \mathscr{E}_\tau;$$

$$\bigcap_{l=1}^{k} \mathscr{E}_\tau^{(l)} = \mathscr{E}_\tau^{(1)} \cap \mathscr{E}_\tau^{(2)} \cap \ldots \cap \mathscr{E}_\tau^{(k)} = \mathscr{E}'_\tau;$$

$$\times_{l=1}^{k} \mathscr{E}_{\tau_l} = \mathscr{E}_{\tau_1} \times \mathscr{E}_{\tau_2} \times \ldots \times \mathscr{E}_{\tau_k} = \mathscr{E}''_\tau,$$

где $\tau = \bigcup_{l=1}^{k} \tau_l$ и $\tau_{l_1} \cap \tau_{l_2} = \varnothing$ для любых $l_1 \neq l_2 (1 \leqslant l_1, l_2 \leqslant k)$.

Используя запись (1.1) для ρ_τ, легко получим соотношение

$$\rho_\tau = \times_{k=1}^{t} \rho_{\{i_k\}}^{a_{i_k}}. \tag{1.3}$$

В дальнейшем нам понадобятся следующие соотношения между множествами произвольной структуры. Пусть R — произвольное множество; $\mathscr{E}_1$ и $\mathscr{E}_2 \subseteq R$ — его подмножества. Определим дополнение $\overline{\mathscr{E}} = R - \mathscr{E}$ подмножества $\mathscr{E} \subseteq R$ как множество, дающее при объединении с $\mathscr{E}$ все R ($\mathscr{E} \cup \overline{\mathscr{E}} = R$). Легко проверить, что $\overline{\mathscr{E}_1} \cup \overline{\mathscr{E}_2} \subseteq \mathscr{E}_1 \cap \mathscr{E}_2 \subseteq \mathscr{E}_1, \mathscr{E}_2$, где символ $\mathscr{E}' \subseteq \mathscr{E}$ означает, что $\overline{\mathscr{E}}'$ входит в $\overline{\mathscr{E}}$ и может совпадать с ним. По индукции легко доказать, что для любого M

$$\bigcup_{i=1}^{M} \overline{\mathscr{E}} \subseteq \bigcap_{i=1}^{M} \mathscr{E}_i \subseteq \mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_M. \tag{1.4}$$

1.2.3. *Аксиоматическое введение числа элементов.* Зададим на подмножествах $\mathscr{E}_\tau \subseteq R_\tau$, где $\tau \subseteq \Sigma$, числовую функцию $N(\mathscr{E}_\tau)$, определяемую следующими постулируемыми свойствами [7].

Аксиома 1 (аддитивности)

$$N(\mathscr{E}'_\tau \cup \mathscr{E}''_\tau) = N(\mathscr{E}'_\tau) + N(\mathscr{E}''_\tau) - N(\mathscr{E}'_\tau \cap \mathscr{E}''_\tau). \tag{1.5}$$

Аксиома 2 (мультипликативности)

$$N(\mathscr{E}_{\tau_1} \times \mathscr{E}_{\tau_2}) = N(\mathscr{E}_{\tau_1}) \cdot N(\mathscr{E}_{\tau_2}) (\tau_1 \cap \tau_2 = \varnothing). \tag{1.6}$$

Аксиома 3 (начальных условий).

Пусть $N(\varnothing) \equiv 0$ и для элементарных τ-проекций

$$\rho_{\{t\}}^{\alpha} (\alpha = \overline{1, a};\ t = \overline{1, s})\, N(\rho_{\{t\}}^{\alpha}) \equiv 1.$$

Последовательно используя соотношения (1.5) и (1.6), получим обобщение на случай k слагаемых и сомножителей:

$$N\left(\bigcup_{l=1}^{k} \mathscr{E}_\tau^{(l)}\right) = \sum_{l=1} N(\mathscr{E}_\tau^{(l)}) - \sum_{l_1 \neq l_2} N\, \mathscr{E}_\tau^{(l_1)} \cap \mathscr{E}_\tau^{(l_2)} + \ldots + (-1)^{k-1} N\left(\bigcap_{l=1}^{k} \mathscr{E}_\tau^{(l)}\right); \tag{1.7}$$

$$N\left(\times_{l=1}^{k} \mathscr{E}_{\tau_l}\right) = \prod_{l=1}^{k} N(\mathscr{E}_{\tau_l}) (\tau_{l_1} \cap \tau_{l_2} = \varnothing \text{ для } l_1 \neq l_2,\ 1 \leqslant l_1, l_2 \leqslant k).$$

Если $\mathscr{E}_{\tau_{l_1}} \cap \mathscr{E}_{\tau_{l_2}} = \varnothing$ для $l_1 \neq l_2 (1 \leqslant l_1;\ l_2 \leqslant k)$, то соотношение (1.7) упрощается

$$N\left(\bigcup_{l=1}^{k} \mathscr{E}_{\tau}^{(l)}\right) = \sum_{l=1}^{k} N(\mathscr{E}_{\tau}^{(l)}). \tag{1.8}$$

Теорема 1.1 [7]. $N(\mathscr{E}_\tau)$ равно числу элементов множества $\mathscr{E}_\tau$.

Доказательство. В самом деле, пусть множество

$$\mathscr{E}_\tau = \{\rho_\tau^{(1)}, \rho_\tau^{(2)}, \ldots, \rho_\tau^{(m)}, \ldots, \rho_\tau^{(M)}\}$$

состоит из M различных τ-проекций. Представим согласно (1.3) каждую τ-проекцию $\rho_\tau^{(m)}$ в виде произведения элементарных τ-проекций.

$$\rho_\tau^{(m)} = \begin{Bmatrix} \alpha_{i_k}^{(m)} \\ i_k \end{Bmatrix} = \bigtimes_{k=1}^{t} \rho_{\{i_k\}}^{\alpha_{i_k}^{(m)}},$$

тогда

$$\mathscr{E}_\tau = \bigcup_{m=1}^{M} \bigtimes_{k=1}^{t} \rho_{\{i_k\}}^{\alpha_{i_k}^{(m)}}. \tag{1.9}$$

Беря от обеих частей соотношения (1.9) функцию N, будем иметь, используя аксиомы аддитивности и мультипликативности

$$N(\mathscr{E}_\tau) = N\left(\bigcup_{m=1}^{M} \bigtimes_{k=1}^{t} \rho_{\{i_k\}}^{\alpha_{i_k}^{(m)}}\right) = \sum_{m=1}^{M} \prod_{k=1}^{t} N(\rho_{\{i_k\}}^{\alpha_{i_k}^{(m)}}), \tag{1.10}$$

где первая аксиома используется в форме (1.8) из-за того, что $\rho_\tau^{(m)}$ все различные (не пересекаются). Далее используя аксиому начальных условий, согласно которой $N(\rho_{\{i_k\}}^{\alpha_k^{(m)}}) \equiv 1$, получим из (1.10)

$$N(\mathscr{E}_\tau) = \sum_{m=1}^{M} \prod_{k=1}^{t} 1 = \sum_{m=1}^{M} 1 = M,$$

что и завершает доказательство теоремы.

Вычислим теперь величину $N(R_\tau)$; имеем

$$R = \bigtimes_{k=1}^{t} \bigcup_{\alpha=1}^{a} \rho_{\{i_k\}}^{\alpha},$$

откуда

$$N(R_\tau) = N\left(\bigtimes_{k=1}^{t} \bigcup_{\alpha=1}^{a} \rho_{\{i_k\}}^{\alpha}\right) = \prod_{k=1}^{t} \sum_{\alpha=1}^{a} N(\rho_{\{i_k\}}^{\alpha}) = \prod_{k=1}^{t} \sum_{\alpha=1}^{a} 1 = a^t$$

в соответствии с ранее найденным в п. 1.2.1 числом элементов R_τ.

1.2.4. *Обсуждение результатов.* На первый взгляд может показаться, что приведенное в п. 1.2.3 операционное аксиоматическое определение числа элементов множеств расположений (как бы просто оно ни было) является искусственным. Однако последующее изложение не должно оставить сомнений на этот счет. В самом деле, если необходимо вычислить число элементов простого множества расположений (такого, например, как R_τ),

то использование для этого аксиоматических соотношений не обязательно. Но в дальнейшем изложении по мере усложнения рассматриваемых множеств расположений использование аксиоматических соотношений для подсчета числа их элементов делается все более оправданным. Более того, подсчет ряда чисел элементов сложных множеств расположений кажется невозможным без использования аксиоматических соотношений (например, чисел элементов пересечений «комбинаторных сфер»; см. п.1.7.2).

Приведенная аксиоматика является разновидностью известной теоретико-множественной аксиоматики теории вероятностей А. Н. Колмогорова [17]. Ее основное отличие от последней состоит во «внутреннем» определении «независимости» подмножеств $\mathscr{E}_\tau \subset R_\tau$ аксиомой 2. В теоретико-вероятностной аксиоматике независимыми считаются множества $\mathscr{E}'$ и $\mathscr{E}''$, для которых вероятность совмещения имеет вид

$$P(\mathscr{E}' \cap \mathscr{E}'') = P(\mathscr{E}')\,P(\mathscr{E}'').$$

Основной целью комбинаторики расположений является подсчет числа τ-проекций ρ_τ, обладающих определенным признаком. Другими словами, необходимо уметь подсчитывать число различных τ-проекций, принадлежащих некоторому подмножеству $\mathscr{E}_\tau \subset R_\tau$.

§ 1.3. Основные множества и соотношения

1.3.1. *Множества* $\mathscr{E}_\tau^{\bar{t}}$. Перейдем к рассмотрению основных для всего дальнейшего изложения множеств $\mathscr{E}_\tau^{\bar{t}}$. Для определения последних примем запись τ-проекций в форме (1.1′)

$$\rho_\tau = (\tau_1, \ldots, \tau_\alpha, \ldots, \tau_a)$$

и зададим на них векторную функцию $F(\rho_\tau)$ вида

$$F(\rho_\tau) = \bar{t} = (t_1, \ldots, t_\alpha, \ldots, t_a),$$

где $t_\alpha = N(\tau_\alpha)$ и из соотношения $\bigcup_{\alpha=1}^{a} \tau_\alpha = \tau$ следует, что целые неотрицательные числа $t_\alpha \geqslant 0\ (\alpha = \overline{1, a})$ удовлетворяют соотношению

$$t_1 + \ldots + t_\alpha + \ldots + t_a = t. \tag{1.11}$$

Зафиксируем некоторый вектор $\bar{t} = (t_1, \ldots, t_\alpha, \ldots, t_a)$ с целочисленными неотрицательными компонентами, удовлетворяющими соотношению (1.11). Определим множество $\mathscr{E}_\tau^{\bar{t}}$ как совокупность всех τ-проекций ρ_τ, при которых $F(\rho_\tau) = \bar{t}$, т. е.

$$\mathscr{E}_\tau^{\bar{t}} = \bigcup_{F(\rho_\tau)=\bar{t}} \rho_\tau.$$

Другими словами, множество $\mathscr{E}_\tau^{\bar{t}}$ состоит из всех τ-проекций ρ_τ, у которых на $\tau = \{i_k\}$ местах стоят предметы $A = \{A_\alpha\}$ в числе $\bar{t} = \{t_\alpha\}$ штук соответственно, безразлично в каком порядке.

Множества $\mathscr{E}_\tau^{\bar{t}}$ содержат в себе как частный случай элементарные τ-проекции. В самом деле

$$\rho_{\{i\}}^{a} = \mathscr{E}_{\{i\}}^{\bar{e}_a},$$

где

$$\overline{e_\alpha} = (\underbrace{0, \ldots, 0}_{\alpha}, \overbrace{1, 0, \ldots, 0}^{a}).$$

Все последующие соотношения выводятся стандартным приемом, состоящим в доказательстве принадлежности произвольной τ-проекции левой части соотношения при допущении о ее принадлежности правой части, и наоборот.

Имеем

$$R = \bigcup_{\overline{m} \in \mathfrak{M}_{a,s}} \mathscr{E}_\sigma^{\overline{m}}, \tag{1.12}$$

где объединение (суммирование) ведется по множеству $\mathfrak{M}_{a,s}$ всех векторов $\overline{m}$ с неотрицательными целочисленными компонентами, удовлетворяющими соотношению

$$m_1 + \ldots + m_\alpha + \ldots + m_a = s.$$

Легко показать, что при различных $\overline{m}_1 \neq \overline{m}_2$

$$\mathscr{E}_\sigma^{\overline{m}_1} \cap \mathscr{E}_\sigma^{\overline{m}_2} = \varnothing. \tag{1.13}$$

Таким образом, соотношение (1.12) является разбиением пространства расположений на сумму непересекающихся множеств $\mathscr{E}_\sigma^{\overline{m}}$.

Далее имеет место соотношение

$$\mathscr{E}_\sigma^{\overline{m}} = \bigcup_{\overline{t}} \mathscr{E}_\tau^{\overline{t}} \times \mathscr{E}_{\sigma-\tau}^{\overline{m}-\overline{t}}, \tag{1.14}$$

где $\tau \subset \sigma$ произвольное подмножество σ и суммирование ведется по всем векторам $\overline{t}$ с неотрицательными целочисленными компонентами, удовлетворяющими условиям $\sum_{\alpha=1}^{a} t_\alpha = t$ и $0 \leqslant t_\alpha \leqslant m_\alpha (\alpha = \overline{1, a})$.

Вычислим число элементов $N(\mathscr{E}_\sigma^{\overline{m}})$. Для этого покажем, что число $N(\mathscr{E}_\sigma^{\overline{m}}) = C_s^{\overline{m}}$ является симметричной функцией координат $m_\alpha (\alpha = \overline{1, a})$ вектора $\overline{m}$, зависящей от числа элементов $N(\sigma) = s$.

В самом деле, пусть заданы два множества $\mathscr{E}_\sigma^{\overline{m}}$ и $\mathscr{E}_{\sigma'}^{\overline{m}'}$, где $N(\sigma) = N(\sigma') = s$, и вектор $\overline{m}'$ получается перестановкой $L = \begin{pmatrix} 1, & 2, \ldots, & a \\ i_1, & i_2, \ldots, & i_a \end{pmatrix}$ координат вектора $\overline{m}$. Установим произвольное взаимно-однозначное соответствие между элементами множеств мест σ и σ′ и осуществим над предметами расположений $\mathscr{E}_\sigma^{\overline{m}}$ преобразование согласно подстановке L, размещая их на соответствующих местах τ′. Этим осуществится взаимно однозначное соответствие между расположениями $\rho \in \mathscr{E}_\sigma^{\overline{m}}$ и $\rho' \in \mathscr{E}_{\sigma'}^{\overline{m}'}$, что обеспечивает равночисленность элементов $\mathscr{E}_\sigma^{\overline{m}}$ и $\mathscr{E}_{\sigma'}^{\overline{m}'}$. Поэтому обозначим $N(\mathscr{E}_\sigma^{\overline{m}}) = C_s^{\overline{m}}$, где $C_s^{\overline{m}}$ симметричная функция координат вектора $\overline{m}$. Найдем ее явное выражение через $\overline{m}$ и s. Для этого рассмотрим частный случай соотношения (1.14) при $\tau = \{i\}$. Будем иметь

$$\mathscr{E}_\sigma^{\overline{m}} = \bigcup_{\alpha=1}^{a} \mathscr{E}_{\{i\}}^{\overline{e}_\alpha} \times \mathscr{E}_{\sigma-\{i\}}^{\overline{m}-\overline{e}_\alpha}.$$

Отсюда

$$N(\mathscr{C}_{\sigma}^{\overline{m}}) = \sum_{\alpha=1}^{a} N(\mathscr{C}_{\{i\}}^{\overline{e_\alpha}}) \cdot N(\mathscr{C}_{\sigma-\{i\}}^{\overline{m-e_\alpha}}) = \sum_{\alpha=1}^{a} N(\mathscr{C}_{\sigma-\{i\}}^{\overline{m-e_\alpha}})$$

или

$$C_s^{\overline{m}} = \sum_{\alpha=1}^{a} C_{s-1}^{\overline{m-e_\alpha}}. \tag{1.15}$$

Для решения конечно-разностного уравнения (1.15) введем производящую функцию чисел $C_s^{\overline{m}}$

$$g_s(u) = g_s(u_1, \ldots, u_a) = \sum_{\overline{m}} C_s^{m_1, \ldots, m_a} u_1^{m_1} \ldots u_a^{m_a}, \tag{1.16}$$

где $u_1 \ldots u_a$ формальные действительные аргументы.

Числа $C_s^{\overline{m}}$ однозначно определяют $g_s(\overline{u})$, и наоборот. Умножая обе части соотношения (1.15) на $u_1^{\overline{m_1}} \ldots u_a^{m_a}$ и суммируя по всем $\overline{m}$, получим конечно-разностное уравнение для $g_s(\overline{u})$

$$g_s(\overline{u}) = g_{s-1}(\overline{u}) \sum_{\alpha=1}^{a} u_\alpha. \tag{1.17}$$

Последовательное применение соотношения (1.17) приводит к соотношению

$$g_s(\overline{u}) = g_1(\overline{u}) \left(\sum_{\alpha=1}^{a} u_\alpha \right)^{s-1}. \tag{1.18}$$

Но из определения (1.16) $g_s(\overline{u})$ следует, что

$$g_1(\overline{u}) = \sum_{\alpha=1}^{a} C_1^{\overline{e_\alpha}} u_\alpha$$

и так как $C_1^{\overline{e_\alpha}} = N(\mathscr{C}_{\{i\}}^{\overline{e_\alpha}}) \equiv 1$, то имеем

$$g_1(\overline{u}) = \sum_{\alpha=1}^{a} u_\alpha.$$

Поэтому имеем из (1.18)

$$g_s(\overline{u}) = \left(\sum_{\alpha=1}^{a} u_\alpha \right)^s = \sum_{\overline{m}} \frac{s!}{m_1! \ldots m_a!} u_1^{m_1} \ldots u_a^{m_a}. \tag{1.19}$$

Сравнивая коэффициенты при одинаковых степенях $u_1^{m_1} \ldots u_a^{m_a}$ соотношений (1.16) и (1.19), заключаем, что

$$C_s^{\overline{m}} = \frac{s!}{m_1! \ldots m_a!}. \tag{1.20}$$

Числа $C_s^{\overline{m}}$ называются полиномиальными коэффициентами.

При $a = 2$, $m_1 = m$ и $m_2 = s - m$

$$C_s^{\overline{m}} = C_s^m = \frac{s!}{m!\,(s-m)!}$$

называется биномиальным коэффициентом.

Если взять функцию N от обеих частей теоретико-множественных соотношений (1.12) и (1.14), то с учетом соотношения (1.20) получим известные соотношения между полиномиальными коэффициентами

$$\sum_{\overline{m}} C_s^{\overline{m}} = a^s; \tag{1.21}$$

$$\sum_{\overline{t}} C_{\overline{t}}^{\overline{t}} C_{s-\overline{t}}^{\overline{m}-\overline{t}} = C_s^{\overline{m}}.$$

Обобщение соотношения (1.14) дает

$$\mathscr{C}_\sigma^{\overline{m}} = \bigcup_{\sum_{\varkappa=1}^{k} \overline{t}_\varkappa = \overline{m}} \times_{\varkappa=1}^{k} \mathscr{C}_{\tau_\varkappa}^{\overline{t}_\varkappa} \tag{1.22}$$

для произвольных фиксированных $\tau_\varkappa (\varkappa = \overline{1, k})$, таких, что $\bigcup_{\varkappa=1}^{k} \tau_\varkappa = \sigma$ и $\tau_\varkappa \cap \tau_{\varkappa'} = \varnothing$, если $\varkappa \neq \varkappa'$.

Далее из (1.22) получим соотношение

$$C_s^{\overline{m}} = \sum_{\sum_{\varkappa=1}^{k} \overline{t}_\varkappa = \overline{m}} \prod_{\varkappa=1}^{k} C_{t_\varkappa}^{\overline{t}_\varkappa},$$

где $s = N(\sigma)$; $t_\varkappa = N(\tau_\varkappa)$; $\sum_{\varkappa=1}^{k} t_\varkappa = s$.

Рассмотрим частный случай соотношения (1.22), когда $t_\varkappa = t (\varkappa = \overline{1,k})$, откуда $s = k \cdot t$. В этом случае будем иметь

$$\mathscr{C}_\sigma^{\overline{m}} = \bigcup_{\overline{t}_1 + \ldots + \overline{t}_k = \overline{m}} \mathscr{C}_{\tau_1}^{\overline{t}_1} \times \ldots \times \mathscr{C}_{\tau_k}^{\overline{t}_k}, \tag{1.23}$$

где компоненты всех векторов $\overline{t}_\varkappa$ $(\varkappa = \overline{1,k})$ удовлетворяют одному и тому же соотношению $t_1^\varkappa + t_2^\varkappa + \ldots + t_a^\varkappa = t$ и их можно рассматривать как значения вектора $\overline{t}$, удовлетворяющего тем же условиям.

Будем рассматривать множества $(\tau_1, \ldots, \tau_\varkappa, \ldots, \tau_k) = \sigma'$ как места, а различные значения вектора $\overline{t}$ как предметы $\{\overline{t}\} = A'$. Расположение ρ имеет вид $\rho = \{\overline{t}_{\tau_\varkappa}\}$. Одновременно рассмотрим ранее определенное множество $\mathscr{C}_{\sigma'}^{\overline{M}}$. Здесь целочисленные координаты вектора $\overline{M} = \{M_{\overline{t}}\}$, означающие число векторов $\overline{t}$ на местах $\sigma' = (\tau_1, \ldots, \tau_k)$, как легко видеть, должны удовлетворять двум соотношениям

$$\sum_{\overline{t}} M_{\overline{t}} = k; \quad \sum_{\overline{t}} \overline{t} M_{\overline{t}} = \overline{m}.$$

Тогда соотношение (1.23) эквивалентно соотношению

$$\mathscr{C}_\sigma^{\overline{m}} = \bigcup_{\sum_{\overline{t}} M_{\overline{t}} = k;\ \sum_{\overline{t}} \overline{t} M_{\overline{t}} = \overline{m}} \bigcup_{\overline{M}} \bigcup_{\rho \in \mathscr{C}_{\sigma'}^{\overline{M}}} \times_{\overline{t}} \times_{r=1}^{M_{\overline{t}}} \mathscr{C}_{\tau_r}^{\overline{t}},$$

откуда

$$C_s^{\bar m} = \sum_{\sum_{\bar t} M_{\bar t} = k;\ \sum_{\bar t} \bar t M_{\bar t} = \bar m} C_k^{\bar M} \prod_{\bar t} \left(C_{\bar t}^{\bar t}\right)^{M_{\bar t}} \quad (s = k \cdot t).$$

В частности при $a = 2$, когда a-мерные вектора $\bar m = (m,\ s - m)$ и $\bar t = (t_1,\ t - t_1)$ характеризуются лишь первыми координатами $m_1 = m$ и t_1, будем иметь

$$C_s^m = \sum_{\sum_{t_1=0}^{t} M_{t_1} = k;\ \sum_{t_1=0}^{t} t_1 M_{t_1} = m} C_k^M \prod_{t_1=0}^{t} \left(C_t^{t_1}\right)^{M_{t_1}} \quad (s = k \cdot t).$$

В бинарном же случае $a = 2$; тогда можно получить соотношение, двойственное в некотором смысле соотношению (1.14). Оно имеет вид

$$\mathscr{C}_\sigma^{\bar m} = \bigcup_{\bar\tau} \mathscr{C}_{\tau - (t)}^{\bar t - \bar e_\alpha} \times \mathscr{C}_{(t)}^{\bar e_\alpha} \times \mathscr{C}_{\sigma - \tau}^{\bar m - \bar t}, \tag{1.24}$$

где $\alpha = 1, 2$ и $\tau = (1, 2, \ldots, t-1, t)$ имеет специальное строение. При этом $\bar t = (t_1, t_2) = (t_1, t - t_1)$ произвольно, но фиксировано, а t пробегает значения $t_\alpha \leqslant t \leqslant s - (m_\alpha - t_\alpha)$.

Из соотношения (1.24) имеем

$$C_s^m = \sum_t C_{t-1}^{\bar t - \bar e_\alpha} C_{s-t}^{\bar m - \bar t}.$$

Двойственность соотношений (1.14) и (1.24) состоит в том, что если в первом суммирование ведется при произвольных, но фиксированных местах (τ) по переменному составу предметов $(\bar t)$, то во втором, наоборот, фиксируется число предметов $(t_\alpha = \text{const})$, а места (τ), на которых они впервые появятся, являются переменными.

1.3.2. *Множества* $\mathscr{C}_{A,\,\sigma}^{\bar k}$. Рассмотрим множество $\mathfrak{M}_{a,s}^*$ a-мерных векторов $\bar m = (m_1, \ldots, m_\alpha, \ldots, m_a)$ с целочисленными неотрицательными компонентами, принимающими значения от 0 до s. Вектора $\bar m$ можно рассматривать как расположения $s+1$-го предмета (чисел) $\sigma' = (0, 1, \ldots, t, \ldots, s)$ на $A = (1, \ldots, \alpha, \ldots, a)$ местах.

Рассмотрим подмножество $\mathfrak{M}_{a,s} \subset \mathfrak{M}_{a,s}^*$, состоящее из всех векторов $\bar m$, удовлетворяющих условию

$$m_1 + \ldots + m_\alpha + \ldots + m_a = s. \tag{1.25}$$

Далее рассмотрим введенное в предыдущем пункте множество $\mathscr{C}_A^{\bar k} \subset \mathfrak{M}_{a,s}$, где $\bar k$ есть $s+1$-мерный вектор. Здесь $\mathscr{C}_A^{\bar k}$ означает совокупность векторов $\bar m$, у которых значение компонент $m_\alpha = t$ встречается в точности k_t раз. Заметим, что компоненты вектора $\bar k$ должны удовлетворять двум соотношениям

$$\begin{aligned} k_0 + k_1 + k_2 + \ldots + k_t + \ldots + k_s &= a; \\ 1k_1 + 2k_2 + \ldots + tk_t + \ldots + sk_s &= s, \end{aligned} \tag{1.26}$$

из которых второе является следствием соотношения (1.25).

Определим теперь множество $\mathscr{C}_{A,\,\sigma}^{\bar k}$ расположений $p \in R$ соотношением

$$\mathscr{C}_{A,\,\sigma}^{\bar k} = \bigcup_{\bar m \in \mathscr{C}_A^{\bar k} \subset \mathfrak{M}_{a,\,s}} \mathscr{C}_\sigma^{\bar m}. \tag{1.27}$$

Вычислим $N(\mathcal{E}^{\overline{k}}_{A,\sigma})$. Имеем из (1.27)

$$N(\mathcal{E}^{\overline{k}}_{A,\sigma}) = \sum_{\overline{m}\in\mathcal{E}^{\overline{k}}_{A}} N(\mathcal{E}^{\overline{m}}_{\sigma}) = \sum_{\overline{m}\in\mathcal{E}^{\overline{k}}_{A}} C^{\overline{m}}_{s} \tag{1.28}$$

Но $C^{\overline{m}}_{s}$ — симметричная функция координат вектора $\overline{m}$, поэтому для всех $\overline{m}\in\mathcal{E}^{\overline{k}}_{A}$

$$C^{\overline{m}}_{s} = C_s\left(\underbrace{0\ldots0}_{k_0},\ldots,\underbrace{t\ldots t}_{k_t},\ldots,\underbrace{s\ldots s}_{k_s}\right) = C^{\overline{m}_0}_{s} = \text{const}.$$

Поэтому из (1.28) имеем

$$N(\mathcal{E}^{\overline{k}}_{A,\sigma}) = \sum_{\overline{m}\in\mathcal{E}^{\overline{k}}_{A}} C^{\overline{m}}_{s} = C^{\overline{m}_0}_{s}\sum_{\overline{m}\in\mathcal{E}^{\overline{k}}_{A}} 1 = C^{\overline{m}_0}_{s} N(\mathcal{E}^{\overline{k}}_{A}) = C^{\overline{m}_0}_{s}\cdot C^{\overline{k}}_{a}.$$

Итак

$$N(\mathcal{E}^{\overline{k}}_{A,\sigma}) = C^{\overline{m}_0}_{a}\cdot C^{\overline{k}}_{a} = C^{\overline{k}}_{a}\, s!\Big/\prod_{t=0}^{s}(t!)^{k_t}.$$

Далее имеем

$$R = \bigcup_{\overline{k}\,(1.26)} \mathcal{E}^{\overline{k}}_{A,\sigma}, \tag{1.29}$$

где суммирование ведется по всем векторам, удовлетворяющим условиям (1.26). Кроме того

$$\mathcal{E}^{\overline{k}_1}_{A,\sigma}\cap\mathcal{E}^{\overline{k}_2}_{A,\sigma} = \varnothing,$$

при $\overline{k}_1\neq\overline{k}_2$. Поэтому, беря функцию N от обеих частей соотношения (1.29), получим

$$a^s = \sum_{\overline{k}\,(1.26)} C^{\overline{k}}_{a} C^{\overline{m}_0}_{s}.$$

Заметим, что множество $\mathcal{E}^{\overline{m}}_{\sigma}$ инвариантно относительно любого взаимно-однозначного преобразования множества мест σ самого в себя в том смысле, что при этом расположения $\rho\in\mathcal{E}^{\overline{m}}_{\sigma}$, переходя друг в друга, остаются в своих $\mathcal{E}^{\overline{m}}_{\sigma}$.

В этом же смысле множество $\mathcal{E}^{\overline{k}}_{A,\sigma}$ инвариантно не только при взаимно однозначном преобразовании множества мест σ в себя, но при аналогичном преобразовании множества предметов A. Эти свойства множеств $\mathcal{E}^{\overline{m}}_{\sigma}$ и $\mathcal{E}^{\overline{k}}_{A,\sigma}$ могут служить в качестве самостоятельного повода для их изучения.

§ 1.4. Комбинаторное определение вероятности

1.4.1. *Классическое определение вероятности множеств расположений.* Примем за множество единственно возможных равновероятных, «элементарных» событий пространство расположений R. Определим для каждого подмножества $\mathcal{E}\subseteq R$ число

$$\mathcal{P}(\mathcal{E}) = N(\mathcal{E})/N(R),$$

которое будем называть вероятностью того, что расположение ρ входит в $\mathcal{E}$, или просто вероятностью множества $\mathcal{E}$. Ясно, что число $P(\mathcal{E})$ удовлетворяет тем же аксиомам п. 1.2.3, что и число $N(\mathcal{E})$, кроме последней аксиомы 3 начальных условий. Здесь

$$\mathcal{P}(\rho) = 1/N(R) \leqslant \mathcal{P}(\mathcal{E}) \leqslant 1 = \mathcal{P}(R).$$

Если принять за множество единственно возможных равновероятных «элементарных» событий произвольное фиксированное подмножество $\mathscr{E} \subset R$, то аналогично определяется условная вероятность подмножества $\mathscr{E}' \subseteq \mathscr{E}$.

$$\mathscr{P}(\mathscr{E}'/\mathscr{E}) = N(\mathscr{E}')/N(\mathscr{E}) \tag{1.30}$$

при условии $\mathscr{E}$.

Используя введенные определения, из соотношений § 1.3 получим вероятности и условные вероятности некоторых из рассмотренных там множеств. Имеем распределение, которое мы будем называть распределением Мизеса[1].

$$P_{a,s}(\bar{k}) = \mathscr{P}(\mathscr{E}^{\bar{k}}_{A,\sigma}) = C^{\bar{k}}_{a} \cdot \frac{C^{\bar{m}_0}_{s}}{a^s}, \tag{1.31}$$

и так называемое гипергеометрическое распределение

$$P_{s,\bar{m},\bar{t}}(\bar{t}) = \mathscr{P}(\mathscr{E}^{\bar{t}}_{\tau} \times \mathscr{E}^{\bar{m}-\bar{t}}_{\sigma-\tau} / \mathscr{E}^{\bar{m}}_{\sigma}) = C^{\bar{t}}_{t} \cdot \frac{C^{\bar{m}-\bar{t}}_{s-t}}{C^{\bar{m}}_{s}}. \tag{1.32}$$

В бинарном случае имеем

$$P_{s,m,t}(t_1) = \mathscr{P}(\mathscr{E}^{t_1}_{\tau} \times \mathscr{E}^{m-t_1}_{\sigma-\tau}/\mathscr{E}^{m}_{\sigma}) = C^{t_1}_{t} \frac{C^{m-t_1}_{s-t}}{C^{m}_{s}}, \tag{1.33}$$

$$Q_{s,m,t_1}(t) = \mathscr{P}(\mathscr{E}^{t_1-1}_{\tau-(t)} \times \mathscr{E}^{1}_{(t)} \times \mathscr{E}^{m-t_1}_{\sigma-\tau}/\mathscr{E}^{m}_{\sigma}) = C^{t_1-1}_{t-1} \frac{C^{m-t_1}_{s-t}}{C^{m}_{s}}. \tag{1.34}$$

Условные вероятности (1.32), (1.33) и (1.34) имеют простую урновую интерпретацию. В самом деле, фиксацию множества $\mathscr{E}^{\bar{m}}_{\sigma}$ можно интерпретировать как наличие урны с s шарами, из которых m_α шаров α-го цвета ($\alpha = \overline{1,a}$). Расположение $\rho \in \mathscr{E}^{\bar{t}}_{\tau} \times \mathscr{E}^{\bar{m}-\bar{t}}_{\sigma-\tau}$ можно рассматривать как выборку объема t из урны $\mathscr{E}^{\bar{m}}_{\sigma}$ «без возвращения» с заданными частотами $\bar{t} = (t_1 \ldots, t_a, \ldots, t_a)$ шаров различных цветов.

Расположение $\rho \subset \mathscr{E}^{t_1-1}_{\tau-(t)} \times \mathscr{E}^{1}_{(t)} \times \mathscr{E}^{m-t}_{s-t}$ можно рассматривать как выборку из урны $\mathscr{E}^{m}_{\sigma}$ «без возвращения», у которой t_1-й по счету шар 1-го цвета впервые появится при t-м вынимании.

1.4.2. *Расширение классического определения вероятности.* Рассмотрим асимптотический случай условных вероятностей (1.31), (1.33) и (1.34), при существовании пределов

$$\lim_{s\to\infty} \frac{m_\alpha}{s} = p_\alpha (\alpha = \overline{1,a}) \left(\sum_{\alpha=1}^{a} p_\alpha = 1, \text{ как следствие } \sum_{\alpha=1}^{a} m_\alpha = s \right)$$

и произвольных фиксированных t_α и $t (0 \leqslant t_\alpha \leqslant t)$.

Легко показать, что

$$\lim_{s\to\infty} \frac{C^{\bar{m}-\bar{t}}_{s-t}}{C^{\bar{m}}_{s}} = p_1^{t_1} \ldots p_\alpha^{t_\alpha} \ldots p_a^{t_a}.$$

Отсюда имеем из (1.32), (1.33) и (1.34)

$$P_t(\bar{t}) = \lim_{s\to\infty} P_{s,m,t}(\bar{t}) = C^{\bar{t}}_{t} p_1^{t_1} \ldots p_\alpha^{t_\alpha} \ldots p_a^{t_a}; \tag{1.35}$$

$$P_t(t_1) = \lim_{s\to\infty} P_{s,m,t}(t_1) = C^{t_1}_{t} p^{t_1} (1-p)^{t-t_1}; \tag{1.36}$$

$$Q_{t_1}(t) = \lim_{s\to\infty} Q_{s,m,t_1}(t) = C^{t_1-1}_{t-1} p^{t_1} (1-p)^{t-t_1}. \tag{1.37}$$

[1] Распределение (1.31), по-видимому, впервые встречается в работе Р. Мизеса [18].

Распределение (1.35) называется полиномиальным распределением с параметрами $\bar{p} = (p_1, \ldots, p_\alpha, \ldots, p_a)$. Распределение (1.36) является частным бинарным случаем ($a = 2$) полиномиального распределения. Оно называется биномиальным распределением или распределением Бернулли. Распределение (1.37) называется распределением Паскаля.

Рассмотрим частный случай полиномиального распределения (1.35), при $t = 1$. В этом случае a расположений $\rho^\alpha \in \mathscr{E}_{(1)}^{\bar{e}_\alpha} \times \mathscr{E}_{\sigma-(1)}^{\bar{m}-\bar{e}_\alpha}$ ($\alpha = \overline{1,a}$), при $s \to \infty$ можно интерпретировать как a значений произвольной дискретной случайной схемы

$$\xi = \begin{pmatrix} A_1, & \ldots, & A_\alpha, & \ldots, & A_a \\ p_1, & \ldots, & p_\alpha, & \ldots, & p_a \end{pmatrix} \left(\sum_{\alpha=1}^{a} p_\alpha = 1 \right), \tag{1.38}$$

определенной на элементах конечного абстрактного множества

$$A = \{A_1, \ldots, A_\alpha, \ldots, A_a\}.$$

Таким образом, постулирование существования обычных пределов

$$\lim_{s\to\infty} \frac{m_\alpha}{s} = p_\alpha \quad (\alpha = \overline{1,\alpha})$$

для параметров $\bar{m}$ и s множества $\mathscr{E}_\sigma^{\bar{m}}$ достаточно для получения произвольной дискретной случайной схемы (1.38), как предела вероятности в классическом определении (1.30).

§ 1.5. Производящие функции

1.5.1. *Определение и основные свойства.* Часто при решении комбинаторных и вероятностных задач полезно использование производящих функций или, как мы будем для краткости говорить, производящих. С примером производящей чисел C_s^m мы уже сталкивались в § 1.3 (см. (1.19)).

Дадим общее определение производящей $g(\mathbf{u})$ произвольной действительной функции $a(\mathbf{m})$ матричного аргумента $\mathbf{m} = \| m_{ij} \|$ ($i = \overline{1,k_1}$, $j = 1, k_2$) с ограниченными целочисленными элементами

$$g(\mathbf{u}) = \sum_{\mathbf{m} \in \mathfrak{A}} a(\mathbf{m})\, \mathbf{u}^{\mathbf{m}}, \tag{1.39}$$

где $\mathbf{u} = \| u_{ij} \|$ — формальный матричный аргумент с действительными элементами; $\mathfrak{A}$ — некоторое множество матриц $\mathbf{m}$, на котором определена функция $a(\mathbf{m})$; символ $\mathbf{u}^{\mathbf{m}}$ означает $\mathbf{u}^{\mathbf{m}} = \prod_{i,j} u_{ij}^{m_{ij}}$.

В случае бесконечного множества $\mathfrak{A}$ сумма (1.39) предполагается сходящейся хотя бы для одного конкретного значения матрицы $\mathbf{u} = \mathbf{u}_0$, отличного от матрицы с хотя бы одним нулевым элементом. Отсюда следует однозначное определение $g(\mathbf{u})$ по $a(\mathbf{m})$, и наоборот.

Рассмотрим некоторый набор пар индексов $(i, j) \in I$. Множество элементов матрицы $\mathbf{m} = \| m_{ij} \|$, состоящее из соответствующих элементов, обозначим через $\mathbf{m}_I$, а множество остальных элементов обозначим через $\mathbf{m}_{\bar{I}}$.

Тогда из $\mathbf{m} \in \mathfrak{A}$ следует $\mathbf{m}_I \in \mathfrak{A}_I$ и

$$g(\mathbf{u})\Big|_{u_{ij}=1 \text{ для } (i,j)\in\bar{I}} = \sum_{\mathbf{m}_I \in \mathfrak{A}_I} a(\mathbf{m}_I)\, \mathbf{u}_I^{\mathbf{m}_I},$$

где

$$\mathbf{u}_I^{\mathbf{m}_I} = \prod_{(i,j)\in I} u_{ij}^{m_{ij}} \quad \text{и} \quad a(\mathbf{m}_I) = \sum_{\mathbf{m}_{\overline{I}} \in \mathfrak{A}_{\overline{I}}} a(\mathbf{m}).$$

Легко показать, что если действительная функция $a(\mathbf{m}_1, \ldots, \mathbf{m}_l)$ зависит от l матричных аргументов указанного типа и имеет соответствующую производящую

$$g_a(\mathbf{u}_1, \ldots, \mathbf{u}_l) = \sum_{\mathbf{m}_1 \ldots \mathbf{m}_l \in \mathfrak{A}^{(l)}} a(\mathbf{m}_1, \ldots, \mathbf{m}_l)\, \mathbf{u}_1^{\mathbf{m}_1} \ldots \mathbf{u}_l^{\mathbf{m}_l},$$

то действительная функция

$$b(\mathbf{m}) = \sum_{\mathbf{m}_1 + \ldots + \mathbf{m}_l = \mathbf{m}} a(\mathbf{m}_1, \ldots, \mathbf{m}_l)$$

имеет производящую

$$g_b(\mathbf{u}) = \sum_{\mathbf{m}\in\mathfrak{A}} b(\mathbf{m})\, \mathbf{u}^{\mathbf{m}} = g_a(\underbrace{\mathbf{u}, \ldots, \mathbf{u}}_{l}).$$

В частности, если $a(\mathbf{m}_1, \ldots, \mathbf{m}_l) = a^{(1)}(\mathbf{m}_1) \ldots a^{(l)}(\mathbf{m}_l)$, то $g_a(\mathbf{u}_1, \ldots, \mathbf{u}_l) = g_a^{(1)}(\mathbf{u}_1) \ldots g_{a^{(l)}}(\mathbf{u}_l)$, $b(m)$ называется композицией $a^{(1)}(\mathbf{m}_1), \ldots, a^{(l)}(\mathbf{m}_l)$ и обозначается

$$b(\mathbf{m}) = a^{(1)}(\mathbf{m}_1) * \ldots * a^{(l)}(\mathbf{m}_l).$$

$g_b(\mathbf{u})$ имеет при этом следующее выражение

$$g_b(\mathbf{u}) = g_{a^{(1)}}(\mathbf{u}) \ldots g_{a^{(l)}}(\mathbf{u}). \tag{1.40}$$

Соотношение (1.40) называют соотношением мультипликативности производящих.

Если $g_{a^{(1)}}(\mathbf{u}) = \ldots = g_{a^{(l)}}(\mathbf{u}) = g_a(\mathbf{u})$, что имеет место, когда $a^{(1)}(\mathbf{m}) = \ldots = a^{(l)}(\mathbf{m}) = a(\mathbf{m})$, то имеем

$$g_b(\mathbf{u}) = g_a^l(\mathbf{u})$$

и в символической записи

$$b(\mathbf{m}) = a^{[l]}(\mathbf{m}).$$

В этом случае можно вывести важное свойство итерации производящих. В самом деле, пусть заданы действительная функция матричных аргументов $\mathbf{m}_1 \ldots \mathbf{m}_l$ и целочисленных параметров $n_1, \ldots, n_l$

$$a_{n_1, \ldots, n_l}(\mathbf{m}_1, \ldots, \mathbf{m}_l) = a_1^{[n_1]}(\mathbf{m}_1) * \ldots * a_l^{[n_l]}(\mathbf{m}_l) \tag{1.41}$$

и действительная функция целочисленных параметров

$$b(n_1, \ldots, n_l).$$

Найдем производящую действительной функции

$$a(\mathbf{m}_1, \ldots, \mathbf{m}_l) = \sum_{\mathbf{m}_1 \ldots \mathbf{m}_l} b(n_1, \ldots, n_l)\, a_{n_1, \ldots, n_l}(\mathbf{m}_1, \ldots, \mathbf{m}_l). \tag{1.42}$$

Для этого зададим производящие

$$g_1(\mathbf{u}_1) = \sum_{\mathbf{m}_1} a_1(\mathbf{m}_1)\, \mathbf{u}_1^{\mathbf{m}_1}, \ \ldots, \ g_l(\mathbf{u}_l) = \sum_{\mathbf{m}_l} a_l(\mathbf{m}_l)\, \mathbf{u}_l^{\mathbf{m}_l} \tag{1.43}$$

и

$$g_b(\mathbf{v}_1, \ldots, \mathbf{v}_l) = \sum_{n_1 \ldots n_l} b(n_1, \ldots, n_l)\mathbf{v}_1^{n_1} \ldots \mathbf{v}_l^{n_l}. \tag{1.44}$$

Тогда из (1.41) и (1.43) следует, что производящая

$$g_{n_1 \ldots n_l}(\mathbf{u}_1 \ldots \mathbf{u}_l) = \sum_{\mathbf{m}_1 \ldots \mathbf{m}_l} a_{n_1 \ldots n_l}(\mathbf{m}_1 \ldots \mathbf{m}_l)\ \mathbf{u}_1^{\mathbf{m}_1} \ldots \mathbf{u}_l^{\mathbf{m}_l} =$$

$$= g_1^{n_1}(\mathbf{u}_1) \ldots g_l^{n_l}(\mathbf{u}_l). \tag{1.45}$$

Далее из (1.42) и (1.45) имеем

$$g_a(\mathbf{u}_1 \ldots \mathbf{u}_l) = \sum_{\mathbf{m}_1 \ldots \mathbf{m}_l} a(\mathbf{m}_1 \ldots \mathbf{m}_l)\mathbf{u}_1^{\mathbf{m}_1} \ldots \mathbf{u}_l^{\mathbf{m}_l} =$$

$$= \sum_{n_1 \ldots n_2} b(n_1, \ldots, n_l) g_1^{n_1}(\mathbf{u}_1) \ldots g_l^{n_l}(\mathbf{u}_l). \tag{1.46}$$

Сравнивая (1.44) с (1.46) заключаем, что

$$g_a(\mathbf{u}_1, \ldots, \mathbf{u}_l) = g_b(g_1(\mathbf{u}_1), \ldots, g_l(\mathbf{u}_l)). \tag{1.47}$$

Соотношение (1.47) называют соотношением итерации производящих.

Пусть задана действительная функция матричных аргументов $\mathbf{m}_1 \ldots \mathbf{m}_l$ и матричных параметров $\mathbf{n}_1 \ldots \mathbf{n}_r$ с целочисленными компонентами

$$a_{\mathbf{n}_1 \ldots \mathbf{n}_r}(\mathbf{m}_1 \ldots \mathbf{m}_l), \tag{1.48}$$

где вообще говоря $l \neq r$.

Тогда производящая (1.48)

$$g(\mathbf{u}_1 \ldots \mathbf{u}_l) = g_{\mathbf{n}_1 \ldots \mathbf{n}_r}(\mathbf{u}_1 \ldots \mathbf{u}_l) = \sum_{\mathbf{m}_1 \ldots \mathbf{m}_l \in \mathfrak{A}} a_{\mathbf{n}_1 \ldots \mathbf{n}_r}(\mathbf{m}_1 \ldots \mathbf{m}_l)\mathbf{u}_1^{\mathbf{m}_1} \ldots \mathbf{u}_l^{\mathbf{m}_l} \tag{1.49}$$

зависит от матричных параметров $\mathbf{n}_1 \ldots \mathbf{n}_l$.

Умножая обе части соотношения (1.49) на действительную функцию параметров $b(\mathbf{n}_1, \ldots, \mathbf{n}_r)$ и степени соответствующих формальных матричных аргументов $\mathbf{v}_1^{\mathbf{n}_1} \ldots \mathbf{v}_r^{\mathbf{n}_r}$ и суммируя по всем $n_1 \ldots n_r \in \mathfrak{A}_2^{(r)}$, получим

$$g(\mathbf{u}_1 \ldots \mathbf{u}_l; \mathbf{v}_1 \ldots \mathbf{v}_r) = \sum_{\mathbf{n}_1 \ldots \mathbf{n}_r \in \mathfrak{A}_2^{(r)}} b(\mathbf{n}_1 \ldots \mathbf{n}_r) g_{n_1 \ldots n_r}(\mathbf{u}_1 \ldots \mathbf{u}_l)\mathbf{v}_1^{\mathbf{n}_1} \ldots \mathbf{v}_r^{\mathbf{n}_r}$$

$l + 1$-ю производящую с весами $b(\mathbf{n}_1 \ldots \mathbf{n}_r)$.

Часто (см. далее п. 1.5.2, а также пп. 2.6.2, 2.6.4) удается найти не саму производящую, а некоторую k-ю производящую с определенными весами. Также как и для первой производящей, для k-й производящей в случае бесконечности множества $\mathfrak{A}_2^{(r)}$ необходимо требовать сходимости соответствующих сумм, что приводит к взаимно-однозначному соответствию между $g(\mathbf{u}_1 \ldots \mathbf{u}_l; \mathbf{v}_1 \ldots \mathbf{v}_r)$ и $a_{\mathbf{n}_1 \ldots \mathbf{n}_l}(\mathbf{m}_1 \ldots \mathbf{m}_l)$.

Если матрицы вырождаются в вектор-строки или вектор-столбцы, то, например, производящая, соответствующая производящей (1.39), имеет вид

$$\bar{g}(\bar{u}) = \sum_{\bar{m} \in \mathfrak{A}} a(\bar{m})\bar{u}^{\bar{m}} \ldots$$

и т. д.

1.5.2. *Примеры конкретных производящих.* Вычислим производящие чисел элементов и вероятностей ранее рассмотренных множеств расположений.

Производящая чисел $C_s^{\overline{m}}$ [см. (1.19)] имеет вид

$$g_s(\overline{u}) = \left(\sum_{\alpha=1}^{a} u_\alpha\right)^s = \sum_{\overline{m}} C_s^{\overline{m}} \overline{u}^{\overline{m}}.$$

Вычислим теперь производящую чисел $N_{a,s} = N(\mathfrak{A}_{a,s})$ элементов множества $\mathfrak{A}_{a,s}$, т. е. числа решений в целых неотрицательных числах m_α уравнения $m_1 + \ldots + m_\alpha + \ldots + m_a = s$.

Легко видеть, что

$$N_{a,s} = \sum_{\overline{k}\ (1.26)} C_a^{\overline{k}}, \tag{1.50}$$

где суммирование ведется по всем векторам $\overline{k} = (k_0, \ldots, k_t, \ldots, k_s)$, удовлетворяющим соотношениям (1.26).

Воспользуемся формулой отрицательного бинома

$$(1-u)^{-a} = \sum_{s=0}^{\infty} C_{s+a-1}^{a-1} u^s, \tag{1.51}$$

легко получаемой дифференцированием обеих частей соотношения (1.51) по u.

Имеем, используя формулу геометрической прогрессии и обобщенного бинома,

$$(1-u)^{-a} = \left(\sum_{t=0}^{\infty} u^t\right)^a = \lim_{r\to\infty}\left(\sum_{t=0}^{r} u^t\right)^a = \lim_{r\to\infty}\left(\sum_{s=0}^{ar}\sum_{\overline{k}\ (3.27)} C_a^{\overline{k}} \cdot u^s\right).$$

Отсюда, учитывая соотношения (1.50) и (1.51), получим

$$(1-u)^{-a} = \sum_{s=0}^{\infty} C_{s+a-1}^{a-1} u^s = \sum_{s=0}^{\infty} N_{a,s}\, u^s.$$

Сравнивая коэффициенты при одинаковых степенях, имеем [19]

$$N_{a,s} = C_{s+a-1}^{a-1}.$$

Аналогичными предельными переходами легко получим вторую производящую вероятностей $P_{a,t}(\overline{k})$ (1.31) с весами $a^t/t!$

$$g_a(\overline{u}, v) = \sum_{t=0}^{\infty} \frac{a^t}{t!} g_{a,t}(\overline{x})\, y^t = \left(\sum_{r=0}^{\infty} \frac{1}{r!} x_r y^r\right)^a,$$

где

$$g_{a,t}(\overline{x}) = \sum_{\overline{k}\ (1.26)} P_{a,t}(\overline{k})\, \overline{u}^{\overline{k}}.$$

Далее имеем вторую производящую вероятностей $P_{s,\overline{m},t}(\overline{l})$ (1.32) с весами $C_s^{\overline{m}}$

$$g_{s,t}(\overline{u}, \overline{v}) = \left(\sum_{\alpha=1}^{a} u_\alpha v_\alpha\right)^t \left(\sum_{\alpha=1}^{a} v_\alpha\right)^{s-t} = \sum_{\overline{m}} C_s^{\overline{m}} g_{s,\overline{m},t}(\overline{u})\, \overline{v}^{\overline{m}},$$

где

$$g_{s,\bar{m},t}(\bar{u}) = \sum_{\bar{t}} P_{s,m,t}(\bar{t})\,\bar{u}^{\bar{t}}.$$

Для вероятностей $P_t(t_1)$ (1.36) и $Q_{t_1}(t)$ (1.37) имеем первые производящие

$$g_t(u) = (pu+q)^t = \sum_{t_1=0}^{t} P_t(t_1)\,u^{t_1}$$

и

$$\tilde{g}_{t_1}(u) = \left(\frac{pu}{1-qu}\right)^{t_1} = \sum_{t=t_1}^{\infty} Q_{t_1}(t)\,u^t, \tag{1.52}$$

где $q = 1-p$.

В ряде случаев не удается найти компактного выражения для числа элементов или вероятности множества, в то время как соответствующая производящая имеет простой вид. Рассмотрим в качестве примера число N_a различных разбиений на подмножества конечного неупорядоченного множества $A = (A_1, \ldots, A_a, \ldots, A_a)$. Пусть множество A разбито на c частей, среди которых k_t подмножеств содержат в точности t элементов. Тогда числа $\bar{k} = (k_1, k_2, \ldots, k_c)$ удовлетворяют соотношениям

$$\begin{aligned} k_1 + k_2 + \ldots + k_t + \ldots + k_a &= c; \\ k_1 + 2k_2 + \ldots + tk_t + \ldots + ak_a &= a. \end{aligned} \tag{1.53}$$

Легко показать, что при фиксации вектора $\bar{k}$ из-за неупорядоченности множества A при всех $a!$ перестановках его элементов к различным разбиениям приведет лишь $N(\bar{k})$-я доля, где

$$N(\bar{k}) = \frac{a!}{\prod_{t=1}^{c} (t!)^{k_t} k_t!} = C_a^{\bar{t}_0} C_c^{\bar{k}} \frac{1}{c!}, \tag{1.54}$$

причем $\bar{t}_0 = (\underbrace{1 \ldots 1}_{k_1}, \underbrace{2 \ldots 2}_{k_2}, \ldots, \underbrace{a \ldots a}_{k_a})$.

Тогда искомое число N_a имеет вид

$$N_a = \sum_{\bar{k}\,(1.53)} N(\bar{k}), \tag{1.55}$$

где суммирование ведется по всем числам $\bar{k}$, удовлетворяющим системе (1.53) при фиксированном a и $1 \leqslant c \leqslant a$ (формально можно считать, что $1 \leqslant c < \infty$, причем при $c > a$ полагать соответствующее $N(\bar{k}) = 0$).

Покажем, что производящая функция чисел N_a с весами $1/a!$ имеет вид [16]

$$g(u) = \sum_{a=0}^{\infty} \frac{N_a}{a!} u^a = e^{e^u - 1}. \tag{1.56}$$

В самом деле, имеем, учитывая (1.54) и (1.55),

$$e^{e^u-1} = e^{\sum_{t=1}^{\infty} \frac{u^t}{t!}} = \prod_{t=1}^{\infty} e^{\frac{u^t}{t!}} = \prod_{t=1}^{\infty} \sum_{k_t=0}^{\infty} \left(\frac{u^t}{t!}\right)^{k_t} \frac{1}{k_t!} = \sum_{a=0}^{\infty} \frac{u^a}{a!} \sum_{\bar{k}\,(1.53)} N(\bar{k}).$$

Отсюда следует соотношение (1.56).

1.5.3. *Факториальные моменты.* Производящие $g(\mathbf{u})$ функции $a(\mathbf{m})$ матричного аргумента $\mathbf{m} = \| m_{ij} \|$ можно использовать для получения так называемых факториальных моментов.

Факториальным моментом $a(\mathbf{m})$ порядка $\mathbf{r} = \| r_{ij} \|$ (r_{ij} — целые неотрицательные числа) назовем сходящуюся сумму

$$\sum_{m_{ij} \geqslant r_{ij}} \prod_{ij} m_{ij} (m_{ij} - 1) \ldots (m_{ij} - r_{ij} + 1)\, a(\mathbf{m}),$$

которую будем обозначать $\mathrm{E}\mathbf{m}^{[\mathbf{r}]} = \mathrm{E} \prod_{ij} m_{ij} (m_{ij} - 1) \ldots (m_{ij} - r_{ij} + 1)$.

Легко видеть, что

$$\mathrm{E}\mathbf{m}^{[\mathbf{r}]} = \prod_{ij} \frac{\partial^{k_{ij}}}{\partial u_{ij}^{k_{ij}}} g(\mathbf{u}) \Big|_{\mathbf{u}=\mathbf{e}},$$

где $\mathbf{e}$ матрица со всеми элементами, равными единице.

Факториальные моменты и обычные и центральные моменты, связанные с ними, полезны как некоторые усредненные характеристики функции $a(\mathbf{m})$ (см. далее гл. 2).

§ 1.6. Асимптотические соотношения

1.6.1. *Асимптотика полиномиальных коэффициентов и h-функция.* Используя формулу Стирлинга [19]

$$n! \approx \sqrt{2\pi n}\,(ne^{-1})^n = e^{n \ln n + O(\ln n)},$$

легко получим асимптотическое представление

$$C_s^{\bar{m}} = e^{\mathrm{sh}(\bar{\mu}) + O(\ln s)} \approx e^{\mathrm{sh}(\bar{\mu})*}, \tag{1.57}$$

где $\bar{\mu} = (\mu_1, \ldots, \mu_2, \ldots, \mu_a)$, $\mu_\alpha = \lim\limits_{s \to \infty} \dfrac{m_\alpha}{s} \geqslant 0$

$$\sum_{\alpha=1}^{a} \mu_\alpha = 1 \tag{1.58}$$

и

$$h(\bar{\mu}) = -\sum_{\alpha=1}^{a} \mu_\alpha \ln \mu_\alpha,$$

причем полагаем $0 \ln 0 = 0$.

Таким образом, функция $h(\bar{\mu})$ (h-функция) определена на точках a-мерного единичного куба, лежащих на гиперплоскости (1.58). Это множество точек с неотрицательными координатами $\bar{\mu}$ будем обозначать $\mathfrak{A}_a$. Умножая координаты точек $\bar{\mu} \in \mathfrak{A}_a$ на s и беря ближайшие целые числа к ним (произвольно разрешая неоднозначности), получим совокупность целочисленных векторов $\bar{m} \in \mathfrak{A}_{a,s}$, что символически запишем так

$$[s\,\mathfrak{A}_a] = \mathfrak{A}_{a,s}. \tag{1.59}$$

Заметим, что к векторам $\bar{m} \in \mathfrak{A}_{a,s}$ приводят целые ε-окрестности (например, в Евклидовой метрике) векторов $\mu = \dfrac{1}{s}\bar{m} \in \mathfrak{A}_a$, где $\varepsilon \leqslant \dfrac{1}{2s}$. Поэтому всюду далее для ε-окрестностей векторов $\mu \in \mathfrak{A}_a$ стремление $\varepsilon \to 0$ считается

* Всюду в дальнейшем изложении приближенное равенство $K(s) \approx e^{sk}$ понимается в смысле $K(s) = e^{sk} + O(\ln s)$.

более медленным, чем $1/s \to 0$, при $s \to 0$. Сделанные оговорки разъясняют описание целочисленных векторов $\bar{m}$ непрерывными векторами $\bar{\mu}$.

Заметим также, что далее во всех асимптотических соотношениях типа (1.57), *где слева должно стоять целое число, предполагается взятие ближайшего целого числа от правой части.*

Приступим теперь к изучению свойств h-функции, используя ее комбинаторное происхождение [см. (1.57)].

Из (1.20) и (1.21) имеем для $\bar{m} \in \mathfrak{A}_{a,s}$

$$1 \leqslant C_s^{\bar{m}} \leqslant a^s,$$

откуда следует, что для $\bar{\mu} \in \mathfrak{A}_a$

$$0 \leqslant h(\bar{\mu}) \leqslant \ln a.$$

Далее, из-за того что число слагаемых в сумме (1.21)

$$N_{a,s} = N(\mathfrak{A}_{a,s}) = C_{s+a-1}^{a-1} = e^{a \ln s + O(s^{-1})}$$

с ростом s растет медленнее, чем по экспоненциальному закону, $\max\limits_{\mu} h(\bar{\mu}) = \ln a$, причем, как легко показать, максимум $h(\bar{\mu})$ достигается при $\bar{\mu} = (\overbrace{1/a, \ldots, 1/a}^{a})$.

Из (1.21) имеем для произвольных $0 \leqslant t \leqslant s$ и $\bar{0} \leqslant \bar{t} \leqslant \bar{m}$, где неравенство между векторами понимается как неравенство между их соответствующими компонентами

$$1 \leqslant C_t^{\bar{t}} \cdot C_{s-t}^{\bar{m}-\bar{t}} \leqslant C_s^{\bar{m}},$$

откуда следует, что при произвольных $0 \leqslant \Theta = \lim\limits_{s\to\infty} \frac{t}{s} \leqslant 1$ $\bar{\Theta}_1 = \lim\limits_{t\to\infty} \frac{1}{t}\bar{t}$ и $\bar{\Theta}_2 = \lim\limits_{s-t\to\infty} \frac{1}{s-t}(\bar{m} - \bar{t})$, связанных соотношением $\Theta\bar{\Theta}_1 + (1-\Theta)\bar{\Theta}_2 = \bar{\mu}$, имеем

$$\Theta h(\bar{\Theta}_1) + (1-\Theta) h(\bar{\Theta}_2) \leqslant h(\Theta\bar{\Theta}_1 + (1-\Theta)\bar{\Theta}_2).$$

При этом знак равенства достигается для $\bar{\Theta}_1 = \bar{\Theta}_2 = \bar{\mu}$.

Все приведенные соотношения известны из теории выпуклых функций, каковой является h-функция, однако специально для h-функции эти свойства и последующие являются следствием ее комбинаторного происхождения (1.57).

1.6.2. *Асимптотика полиномиального распределения и k-функция.* Общеизвестно асимптотическое приближение полиномиального распределения к многомерному нормальному распределению [19].

Однако для дальнейшего изложения необходима асимптотика полиномиального распределения, связанная с так называемыми большими уклонениями и исследованная сравнительно недавно [20, 21, 22].

Прежде всего из соотношений (1.35) и (1.57) при $t \to \infty$ имеем

$$C_t^{\bar{t}} \bar{p}^{\bar{t}} \approx e^{-t[h(\bar{\Theta}, \bar{p}) - h(\bar{\Theta})]} \leqslant 1, \tag{1.60}$$

где

$$h(\bar{\Theta}, \bar{p}) = -\sum_{\alpha=1}^{a} \Theta_\alpha \ln p_\alpha; \quad (h(\Theta, \Theta) = h(\Theta)); \quad \bar{\Theta}, \bar{p} \in \mathfrak{A}_a,$$

откуда

$$h(\overline{\Theta}, \overline{p}) - h(\overline{\Theta}) \geqslant 0,$$

причем знак равенства достигается при $\overline{\Theta} = \overline{p}$.

Рассмотрим некоторое подмножество $\mathfrak{M} \subset \mathfrak{A}_a$ векторов Θ, которому соответствует в смысле (1.59) подмножество $\mathfrak{M}_t = [t\,\mathfrak{M}] \subset \mathfrak{A}_{a,t}$ векторов $\overline{t}$.

Приведем асимптотическую при $t \to \infty$ оценку вероятности

$$\mathcal{P}(\mathfrak{M}_t) = \sum_{\overline{t} \in \mathfrak{M}t} C_t^{\overline{t}} \overline{p}^{\overline{t}} = 1 - \sum_{\overline{t} \in \overline{\mathfrak{M}t}} C_t^{\overline{t}} \overline{p}^{\overline{t}}. \tag{1.61}$$

Из-за того, что число слагаемых суммы (1.61) растет неэкспоненциально с ростом t ($N(\mathfrak{M}_t)$, $N(\mathfrak{M}_{a,t}) \leqslant N(\mathfrak{A}_{a,t}) = e^{a \ln t + O(t^{-1})}$), как легко видеть, имеет место асимптотическая при $t \to \infty$ оценка с учетом соотношения (1.60)

$$\mathcal{P}(\mathfrak{M}_t) \approx \begin{cases} e^{-t\,[h(\overline{\Theta}^{(0)}, \overline{p}) - h(\overline{\Theta}^{(0)})]}, & \text{если } \overline{p} \notin \mathfrak{M}; \\ 1 - e^{-t\,[h(\overline{\Theta}^{(0)}, \overline{p}) - h(\overline{\Theta}^{(0)})]}, & \text{если } \overline{p} \in \mathfrak{M}, \end{cases} \tag{1.62}$$

где

$$h(\overline{\Theta}^{(0)}, \overline{p}) - h(\overline{\Theta}^{(0)}) = \max_{\overline{\Theta} \in \mathfrak{M}} [h(\overline{\Theta}, \overline{p}) - h(\overline{\Theta})] = \min_{\overline{\Theta} \in \overline{\mathfrak{M}}} [h(\overline{\Theta}, \overline{p}) - h(\overline{\Theta})] \tag{1.63}$$

и включение и невключение вектора $\overline{p}$ в $\mathfrak{M}$ в (1.62) рассматривается с некоторой его ε-окрестностью.

Зададим множество $\mathfrak{M}$ значений векторов $\overline{\Theta}$ линейными относительно координат условиями вида

$$\sum_{\alpha=1}^{a} \Theta_\alpha d_\alpha \leqslant d. \tag{1.64}$$

Тогда экстремум (1.63) может быть легко найден [21, 22] как экстремум функции

$$F(\Theta_1, \ldots, \Theta_a) = h(\overline{\Theta}, \overline{p}) - h(\overline{\Theta}) = -\sum_{\alpha=1}^{a} \Theta_\alpha \ln p_\alpha + \sum_{\alpha=1}^{a} \Theta_\alpha \ln \Theta_\alpha \tag{1.65}$$

a переменных $\Theta_1, \ldots, \Theta_a$, связанных двумя линейными условиями

$$\sum_{\alpha=1} \Theta_\alpha d_\alpha - d = 0 \quad \text{и} \quad \sum_{\alpha=1}^{a} \Theta_\alpha - 1 = 0, \tag{1.66}$$

или безусловный экстремум функции

$$Q(\Theta_1, \ldots, \Theta_a) = F(\Theta_1, \ldots, \Theta_a) - \lambda\left(\sum_{\alpha=1}^{a} \Theta_\alpha d_\alpha - d\right) + \mu\left(\sum_{\alpha=1}^{a} \Theta_\alpha - 1\right)$$

с неопределенными множителями Лагранжа. Далее для отыскания экстремума Q приравниваем частные производные $\partial Q / \partial \Theta_\alpha$ нулю

$$\frac{\partial Q}{\partial \Theta_\alpha} = -\ln p_\alpha + \ln \Theta_\alpha + 1 - \lambda d_\alpha + \mu = 0 \quad (\alpha = \overline{1,a}).$$

Отсюда

$$\Theta_\alpha = p_\alpha e^{-(1+\mu)}\, e^{\lambda d_\alpha}.$$

Для отыскания λ и μ получим из условий (1.66) соотношения

$$e^{-(1+\mu)}\sum_{\alpha=1}^{a} p_\alpha e^{\lambda d_\alpha} = 1 \quad \text{и} \quad e^{-(1+\mu)}\sum_{\alpha=1}^{a} d_\alpha p_\alpha e^{\lambda d_\alpha} = d. \tag{1.67}$$

Деля второе соотношение (1.67) на первое, получим уравнение для определения λ

$$\sum_{\alpha=1}^{a} d_\alpha p_\alpha e^{\lambda d_\alpha} = d \sum_{\alpha=1}^{a} p_\alpha e^{\lambda d_\alpha}. \tag{1.68}$$

Пусть $\min\limits_{\alpha} d_\alpha \leqslant d \leqslant \max\limits_{\alpha} d_\alpha$, тогда уравнение (1.68) имеет единственный вещественный корень λ_0. Через него из первого соотношения (1.67) выражается μ

$$e^{-(1+\mu)} = \left(\sum_{\alpha=1}^{a} p_\alpha e^{\lambda_0 d_\alpha}\right)^{-1}.$$

Откуда

$$\Theta_\alpha = \Theta_\alpha^{(0)} = p_\alpha e^{\lambda_0 d_\alpha}\left(\sum_{\alpha=1}^{a} p_\alpha e^{\lambda_0 d_\alpha}\right)^{-1}.$$

Подставив найденное значение $\Theta_\alpha^{(0)}$ в (1.65), получим его экстремальное значение, которое оказывается минимумом

$$h(\bar{\Theta}^{(0)}, \bar{p}) - h(\bar{\Theta}^{(0)}) = \sum_{\alpha=1}^{a} \Theta_\alpha^{(0)} \ln \frac{\Theta_\alpha^{(0)}}{p_\alpha} = \lambda_0 d - \ln\left(\sum_{\alpha=1}^{a} p_\alpha e^{\lambda_0 d_\alpha}\right),$$

где λ_0 — вещественный корень уравнения (1.68).

Для дальнейшего изложения удобно следующее обозначение найденного минимума, зависящего от векторов $\bar{p} = (p_1, \ldots, p_a)$, $\bar{d} = (d_1, \ldots, d_a)$ и константы d. Положим

$$k_{\bar{p},\bar{d}}(\varepsilon) = h(\bar{\Theta}^{(0)}, \bar{p}) - h(\bar{\Theta}^{(0)}) = \lambda_0[\gamma'(0) + \varepsilon] - \gamma(\lambda_0), \tag{1.69}$$

где

$$\gamma(\lambda) = \ln \sum_{\alpha=1}^{a} p_\alpha e^{\lambda d_\alpha}, \tag{1.70}$$

$$\varepsilon = d - \gamma'(0) > 0$$

и λ_0 является корнем уравнения

$$d = \gamma'(0) + \varepsilon = \gamma'(\lambda).$$

Соответствующее множество $\mathfrak{M}_t$ ($\mathfrak{M}$) значений $\bar{t}(\bar{\Theta})$, задаваемое линейными условиями (1.64), будем обозначать $\mathfrak{M}_t(\varepsilon)$ ($\mathfrak{M}(\varepsilon)$). Тогда асимптотическая оценка (1.62) для конкретного множества $\mathfrak{M}_t = \mathfrak{M}_t(\varepsilon)$ имеет вид:

$$\mathcal{P}(\mathfrak{M}_t(\varepsilon)) \approx \begin{cases} e^{-tk_{\bar{p}\bar{d}}(\varepsilon)}, & \text{если } \bar{p} \notin \mathfrak{M}(\varepsilon); \\ 1 - e^{-tk_{\bar{p}\bar{d}}(\varepsilon)}, & \text{если } \bar{p} \in \mathfrak{M}(\varepsilon). \end{cases} \tag{1.71}$$

Если $\varepsilon < 0$, то условия в соотношении (1.71) меняются местами.

Учитывая важность функции $k_{\overline{pd}}(\varepsilon)$ для дальнейшего изложения, произведем ее оценки снизу (приводящие к наиболее интересным оценкам снизу вероятности $\mathcal{P}(\mathfrak{M}_t(\varepsilon))$.

Используя формулу Тейлора, имеем

$$\gamma(\lambda) = \gamma(0) + \lambda\gamma'(0) + \frac{\lambda^2}{2}\gamma''(0) + \frac{\lambda^3}{6}\gamma'''(\lambda_1);$$

$$\gamma'(\lambda) = \gamma'(0) + \lambda\gamma''(0) + \frac{\lambda^2}{2}\gamma'''(\lambda_2);$$

$$\lambda\gamma'(\lambda) = \lambda\gamma'(0) + \lambda^2\gamma''(0) + \frac{\lambda^3}{2}\gamma'''(\lambda_2),$$

где $0 \leqslant \lambda_1,\ \lambda_2 \leqslant |\lambda|$.

Отсюда, учитывая, что $\gamma(0) \equiv 0$ (см. (1.70)), имеем

$$\lambda\gamma'(\lambda) - \gamma(\lambda) = \frac{\lambda^2}{2}\gamma''(0) + \frac{\lambda^3}{6}[3\gamma'''(\lambda_2) - \gamma'''(\lambda_1)]. \tag{1.72}$$

Легко показать, что

$$\gamma'(\lambda) = \sum_{\alpha=1}^{a} d_\alpha p_\alpha(\lambda);\ \gamma''(\lambda) = \sum_{\alpha=1}^{a} (d_\alpha - \gamma'(\lambda))^2 p_\alpha(\lambda) \geqslant 0;$$

$$\gamma'''(\lambda) = \sum_{\alpha=1}^{a} (d_\alpha - \gamma'(\lambda))^3 p_\alpha(\lambda),$$

где

$$p_\alpha(\lambda) = p_\alpha e^{\lambda d_\alpha}/\gamma(\lambda) \quad \left(\sum_{\alpha=1}^{a} p_\alpha(\lambda) = 1\right),$$

причем с ростом λ функция $\gamma'(\lambda)$ монотонно не убывает, так как $\gamma''(\lambda) \geqslant 0$

Пусть $|d_\alpha| \leqslant L/2$, тогда $|\gamma'(\lambda)| \leqslant L/2$ и

$$|\gamma'''(\lambda)| \leqslant |L/2 + L/2|^3 = L^3. \tag{1.73}$$

Из (1.72) и (1.73) имеем оценку

$$\frac{\lambda^2}{2}\gamma''(0) - \frac{2}{3}|\lambda^3| L^3 \leqslant \lambda\gamma'(\lambda) - \gamma(\lambda) \leqslant \frac{\lambda^2}{2}\gamma''(0) + \frac{2}{3}|\lambda^3| L^3. \tag{1.74}$$

Пусть

$$\alpha\frac{\lambda^2}{2}\gamma''(0) \leqslant \frac{\lambda^2}{2}\gamma''(0) - \frac{2}{3}|\lambda^3| L^3 \quad (0 \leqslant \alpha \leqslant 1). \tag{1.75}$$

Тогда из (1.74) и (1.75) имеем

$$\alpha\frac{\lambda^2}{2}\gamma''(0) \leqslant \lambda\gamma'(\lambda) - \gamma(\lambda) \leqslant (2-\alpha)\frac{\lambda^2}{2}\gamma''(0) \tag{1.76}$$

и для выполнения (1.75) необходимо и достаточно, чтобы

$$|\lambda| \leqslant \frac{1-\alpha}{L^3}\frac{3}{4}\gamma''(0). \tag{1.77}$$

Оценим теперь величину модуля корня $|\lambda_0(\varepsilon)|$ уравнения

$$\gamma'(\lambda) = \gamma'(0) + \lambda\gamma''(0) + \frac{\lambda^2}{2}\gamma'''(\lambda_2) = \gamma'(0) + \varepsilon.$$

Учитывая условие (1.77), получим

$$|\lambda|\gamma''(0)\left[1-\frac{3}{8}(1-\alpha)\right]\leqslant|\varepsilon|\leqslant|\lambda|\gamma''(0)\left[1+\frac{3}{8}(1-\alpha)\right],$$

откуда

$$\frac{|\varepsilon|}{\gamma''(0)\left[1+\frac{3}{8}(1-\alpha)\right]}\leqslant|\lambda_0(\varepsilon)|\leqslant\frac{|\varepsilon|}{\gamma''(0)\left[1-\frac{3}{8}(1-\alpha)\right]}. \tag{1.78}$$

Для совместности оценок (1.77) и (1.78) необходимо выполнение неравенства

$$\frac{\varepsilon}{\gamma''(0)\left[1+\frac{3}{8}(1-\alpha)\right]}\leqslant|\lambda_0|\leqslant\frac{1-\alpha}{L^3}\,\frac{3}{4}\gamma''(0). \tag{1.79}$$

Итак, если $|d_\alpha|<L/2$ и

$$\frac{\varepsilon}{[\gamma''(0)]^2}\leqslant\frac{3}{4L^3}(1-\alpha)\left[1+\frac{3}{8}(1-\alpha)\right], \tag{1.80}$$

то из (1.76) и (1.79) имеем

$$k_{\overline{p},\overline{d}}(\varepsilon)=\lambda_0\gamma'(\lambda_0)-\gamma(\lambda_0)\geqslant\frac{\alpha}{\left(1+\frac{3}{8}(1-\alpha)\right)^2}\cdot\frac{\varepsilon^2}{2\gamma''(0)}, \tag{1.81}$$

где $\lambda_0=\lambda_0(\varepsilon)$ находится из уравнения.

Ясно, что оценка (1.81) ухудшается с ростом ε, так как для выполнения (1.80) приходится брать малые α. Например, полагая

$$\frac{\alpha}{\left[1+\frac{3}{8}(1-\alpha)\right]^2}=\frac{2}{3}, \tag{1.82}$$

получим из (1.80)

$$\frac{\varepsilon}{(\gamma''(0))^2}\leqslant\frac{3}{4L^3}(1-\alpha)\sqrt{\frac{3}{2}\alpha}. \tag{1.83}$$

Из (1.82) имеем

$$\frac{2}{3}<\frac{7}{9}<\alpha=\frac{2}{3}\left[\frac{\sqrt{19}}{2}-1\right]^2<\frac{8}{10}.$$

Находим из (1.81), (1.82) и (1.83), что при

$$\frac{|\varepsilon|}{(\gamma''(0))^2}\leqslant\frac{1}{7L^3},$$

$$k_{\vec{p},\vec{d}}(\varepsilon)\geqslant\frac{\varepsilon^2}{3\gamma''(0)},$$

где $|\alpha_\alpha|\leqslant L/2$.

Заметим, что при $\varepsilon\to0$ из (1.78) имеем $\lambda_0(\varepsilon)\to0$, поэтому из (1.74) имеем

$$k_{\overline{p},\overline{d}}(\varepsilon)\underset{\varepsilon\to0}{\to}\frac{\varepsilon^2}{2\gamma''(0)}. \tag{1.84}$$

Из соотношения (1.84) следует, что уравнение $k_{\overline{p},\overline{d}}(\varepsilon)=\delta>0$ имеет два корня $\varepsilon'(\delta)<0<\varepsilon(\delta)$, которые при $\delta\to0$ стремятся к нулю.

1.6.3. *Пуассоновское приближение.* В предыдущем пункте рассматривалось асимптотическое поведение полиномиального распределения при росте $t \to \infty$, когда параметры распределения $\bar{p} = (p_1, \ldots, p_a)$ — произвольные фиксированные величины, не зависящие от t.

Рассмотрим теперь случай, когда

$$p_\alpha = c_\alpha / t \quad (\alpha = \overline{2,a})$$

зависят от t, так что при $t \to \infty$ $p_\alpha \to 0$ $(\alpha = \overline{2,a})$ и

$$p_1 = 1 - \sum_{\alpha=2}^{a} p_\alpha = 1 - \sum_{\alpha=2}^{a} c_\alpha / t = 1 - c/t \to 1.$$

В этом случае вместо симметричной записи производящей функции

$$g(\bar{u}) = \sum_{\bar{t}} C_t^{\bar{t}} \bar{p}^{\bar{t}} \bar{u}^{\bar{t}} = \left(\sum_{\alpha=1}^{a} p_\alpha u_\alpha \right)^t$$

полиномиального распределения удобна эквивалентная ей несимметричная запись

$$g(\bar{u}') = g(1, u_2, \ldots, u_a) = \sum_{\bar{t}} C_t^{\bar{t}} \bar{p}^{\bar{t}} \prod_{\alpha=2}^{a} u_\alpha^{t_\alpha} =$$

$$= \left(p_1 + \sum_{\alpha=2}^{a} p_\alpha u_\alpha \right)^t = \left(1 + \frac{1}{t} \sum_{\alpha=2}^{a} c_\alpha (u_\alpha - 1) \right)^t.$$

Переходя к пределу при $t \to \infty$, имеем

$$g(\bar{u}') \to \prod_{\alpha=2}^{a} e^{c_\alpha (u_\alpha - 1)}.$$

Отсюда следует [23], что в рассматриваемом асимптотическом случае распределение полиномиально распределенного вектора $\bar{t} = (t_1, t_2, \ldots, t_a)$ полностью определяется независимо распределенными по Пуассону $P_{c_\alpha}(t_\alpha) = \frac{c_\alpha^{t_\alpha}}{t_\alpha!} e^{-c_\alpha}$ с параметрами c_α компонентами t_α $(\alpha = \overline{2,a})$.

Введем в рассмотрение множество $\mathfrak{M}_t'(\varepsilon)$ векторов $\bar{t}$, определяемое линейными условиями

$$\sum_{\alpha=2}^{a} t_\alpha < c = \varepsilon t .$$

или эквивалентным им условием $t_1 \geqslant t - c$.

Ввиду того, что распределение Пуассона устойчиво при композиции, случайная величина $t' = \sum_{\alpha=2}^{a} t_\alpha$ имеет распределение Пуассона с параметром $\Lambda = \sum_{\alpha=2}^{a} c_\alpha$. Поэтому

$$\mathcal{P}(\mathfrak{M}_t'(\varepsilon)) = \sum_{\bar{t} \in \mathfrak{M}'_t(\varepsilon)} C_t^{\bar{t}} \bar{p}^{\bar{t}} \to \sum_{r=0}^{c-1} \frac{\Lambda}{r!} e^{-\Lambda} = 1 - \sum_{r=c}^{\infty} \frac{\Lambda}{r!} e^{-\Lambda}. \qquad (1.85)$$

Произведем оценку вероятности (1.85). Имеем, используя формулу Стирлинга, в предположении $c > e\Lambda$

$$\mathcal{P}(\mathfrak{M}'_t(\varepsilon)) \to 1 - \sum_{r=c}^{\infty} \frac{\Lambda^r}{r!} e^{-\Lambda} \geqslant 1 - e^{-\Lambda} \sum_{r=c}^{\infty} \frac{\Lambda^r}{\left(\frac{r}{e}\right)^r} \geqslant 1 -$$

$$- e^{-\Lambda} \left(\frac{\Lambda e}{c}\right)^c \frac{1}{1 - \frac{\Lambda e}{c}} \geqslant 1 - e^{-\Lambda - c\ln c + c\ln \Lambda e}.$$

Если $c = \varepsilon t$, где $\varepsilon > 0$, то, так как $\Lambda = \text{const}$, при $t \to \infty$, имеем

$$\mathcal{P}(\mathfrak{M}'_t(\varepsilon)) \to 1 - e^{-\varepsilon t \ln t + O(t)},$$

или, вводя обозначение $K(t) \sim e^{kt\ln t}$ для $K(t) = e^{kt\ln t + O(t)}$ имеем

$$\mathcal{P}(\mathfrak{M}'_t(\varepsilon)) \sim 1 - e^{-\varepsilon t \ln t}. \tag{1.86}$$

Рассмотрим далее известное [23] Пуассоновское приближение распределения Паскаля.

В самом деле случайная величина $t_2 = t - t_1$ согласно (1.52) имеет производящую функцию

$$\tilde{g}_{t_1}(u) = \left(\frac{1-q}{1-qu}\right)^{t_1} = [1 + q(u-1) + O(q^2)]^{t_1}.$$

Пусть $q = l/t$, где l не зависит от t_1. Тогда при $t_1 \to \infty$

$$\tilde{g}_{t_1}(u) \to e^{l(u-1)}.$$

Отсюда следует, что величина t_2, или, что то же, величина $t - t_1$, имеет распределение Пуассона с параметром $l = qt_1$, при малых $q = l/t_1$ и $t_1 \to \infty$.

Известно [19], что биномиальное распределение (1.36) при $p = k/t$, где k не зависит от t, при $t \to \infty$ имеет распределение Пуассона с параметром k.

Рассмотрим асимптотическое поведение распределения Мизеса (1.31), когда $s \to \infty$, $a \to \infty$ и $\lim\limits_{\substack{s\to\infty \\ a\to\infty}} \frac{s}{a} = \Lambda = \text{const}$. Используя формулу Стирлинга, получим [7] их соотношения (1.31)

$$P_{a,t}(\bar{k}) \to C_a^{\bar{k}} \prod_{r=0}^{t} \left(\frac{\Lambda^r}{r!}\right)^{k_r} = P_a(\bar{k}).$$

Отсюда следует, что в рассматриваемом асимптотическом случае распределение Мизеса стремится к полиномиальному распределению с параметрами

$$p_r = \frac{\Lambda^r}{r!} e^{-\Lambda} = P_\Lambda(r) \quad (r = 0,1,2...),$$

имеющими распределение Пуассона.

Далее, так как частные распределения компонент $\bar{k} = (k_0, \ldots, k_r, \ldots, k_t)$ в этом случае

$$\mathcal{P}(k_r) = \sum_{\bar{k}.k_r = \text{const}} P_a(\bar{k}) = C_a^{k_r} p_r^{k_r} (1 - p_r)^{a - k_r}$$

имеют биномиальное распределение, то при больших a имеет место аппроксимация

$$\mathcal{P}(k_r) \to P_a(k_r) \to \frac{\Omega_{a,r}^{k_r}}{k_r!} e^{-\Omega_{a,r}}, \tag{1.87}$$

где $\Omega_{a,r} = a\frac{\Lambda^r}{r!} e^{-\Lambda}$, причем $\Lambda = \lim\limits_{\substack{s\to\infty \\ a\to\infty}} \frac{s}{a}$.

1.6.4. *Вальдовское приближение.* Рассмотрим асимптотическое поведение распределений Бернулли и Паскаля при произвольном, но фиксированном параметре $p\,(q = 1 - p)$ и растущих t и t_1, соответственно.

Предварительно заметим, что имеет место тождество

$$Q_{t_1}(t) \equiv \frac{t_1}{t} P_t(t_1), \tag{1.88}$$

следующее из выражений (1.36) и (1.37).

Используя нормальную аппроксимацию биномиального распределения, из (1.88) получим [19] при $t \to \infty$

$$Q_{t_1}(t) \to \frac{t_1}{t} \frac{1}{\sqrt{2\pi t p q}} e^{-\frac{1}{2}\left(\frac{t_1 - tp}{\sqrt{tpq}}\right)^2}. \tag{1.89}$$

Далее, величины среднего и дисперсии распределения Паскаля имеют вид

$$\mathrm{E}t = g'_{t_1}(1) = \frac{t_1}{p} \quad \text{и} \quad \mathrm{D}t = g''_{t_1}(1) - (g'_{t_1}(1))^2 + g'_{t_1}(1) = t_1 \frac{q}{p^2},$$

соответственно. Теперь преобразуем правую часть соотношения (1.89)

$$Q_{t_1}(t) \to \sqrt{\frac{c}{2\pi}}\, y^{-3/2} e^{-\frac{c}{2}\left(\sqrt{y} - \frac{1}{\sqrt{y}}\right)^2} \frac{1}{\mathrm{E}t} = w_c(y) \frac{1}{\mathrm{E}t},$$

где

$$y = \frac{t}{\mathrm{E}t} = \frac{tp}{t_1} \quad \text{и} \quad c = \frac{(\mathrm{E}t)^2}{\mathrm{D}t} = t_1 \frac{1}{q}. \tag{1.90}$$

Плотность вероятности

$$w_c(y) = \sqrt{\frac{c}{2\pi}}\, y^{-3/2} e^{-\frac{c}{2}\left(\sqrt{y} - \frac{1}{\sqrt{y}}\right)^2}$$

называется распределением Вальда [34], ее функция распределения обозначается

$$W_c(y) = \int_0^y w_c(x)\, dx.$$

§ 1.7. Двойные расположения

1.7.1. *Множества $\mathscr{E}_s^m$.* Рассмотрим два множества предметов $A = (A_1, \ldots, A_\alpha, \ldots, A_a)$ и $B = (B_1, \ldots, B_\beta, \ldots, B_b)$ и их τ-проекции ρ_τ и $\rho^*_{\tau^*}$ на одно и то же множество мест $\sigma = (1, 2, \ldots, t, \ldots, s)$, соответственно.

Используя представление (1.1′), имеем

$$\rho_\tau = (\tau_1, \ldots, \tau_\alpha, \ldots, \tau_a)\left(\bigcup_{\alpha=1}^{a} \tau_\alpha = \tau\right)$$

и

$$\rho_{\tau^*}^* = (\tau_1^*, \ldots, \tau_\beta^*, \ldots, \tau_b^*)\,(\bigcup_{\beta=1}^{b} \tau_\beta^* = \tau^*).$$

В частности, расположения $\rho \in R$ и $\rho^* \in R$ имеют представления (1.2)

$$\rho = (\tau, \ldots, \tau_2, \ldots, \tau_a) \text{ и } \rho^* = (\tau_1^*, \ldots, \tau_\beta^*, \ldots, \tau_b^*)\,(\bigcup_{\alpha=1}^{a} \tau_\alpha = \bigcup_{\beta=1}^{b} \tau_\beta = \sigma),$$

соответственно.

Пару расположений $\rho, \rho^* \in (R, R^*)$ можно рассматривать как одно двойное расположение $r = (\rho, \rho^*)$ «составного» множества предметов $A \times B = \{A_\alpha A_\beta\}$ на местах $\sigma = (1, \ldots, s)$ с соответствующим представлением

$$r = \{\tau_{\alpha,\beta}\}\,(\bigcup_{\alpha\beta} \tau_{\alpha\beta} = \sigma),$$

где

$$\tau_{\alpha\beta} = \tau_\alpha \cap \tau_\beta;\ \bigcup_\beta \tau_{\alpha\beta} = \tau_\alpha;\ \bigcup_\alpha \tau_{\alpha\beta} = \tau_\beta^*;\ \bigcup_\alpha \tau_\alpha = \bigcup_\beta \tau_\beta^* = \sigma. \quad (1.91)$$

Соотношения (1.91) указывают на эквивалентность задания пары расположений ρ, ρ^* заданию двойного расположения r, и наоборот.

Введем числа $N(\tau_{\alpha\beta}) = m_{\alpha\beta}$; $N(\tau_\alpha) = m_\alpha$; $N(\tau_\beta^*) = m_\beta^*$, которые в соответствии с соотношениями (1.91) удовлетворяют условиям:

$$\sum_\beta m_{\alpha\beta} = m_\alpha;\ \sum_\alpha m_{\alpha\beta} = m_\beta^*;\ \sum_\alpha m_\alpha = \sum_\beta m_\beta = s. \quad (1.92)$$

Матрица $\mathbf{m} = \| m_{\alpha\beta} \| = (\overline{m_\alpha})$ с неотрицательными целочисленными элементами $m_{\alpha\beta}$ ($\alpha = \overline{1,a}$, $\beta = \overline{1,b}$) имеет важное значение для последующего изложения. Ее мы будем называть расстоянием между расположениями ρ и ρ^* и обозначать $\mathbf{m} = \mathbf{m}(\rho, \rho^*) = \| m_{\alpha\beta} \|$. Используя матричную запись, соотношения (1.92) можно переписать в следующей компактной форме

$$\overline{\mathbf{m}}\,\overline{e}_b = \overline{m};\ \overline{e}_a \mathbf{m} = \overline{m}^*;\ \overline{e}_a \mathbf{m} \overline{e}_b = \overline{e}_a \overline{m} = \overline{m}^* \overline{e}_b = s,$$

где $\overline{e}_b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\Big\}b$; $\overline{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_a \end{pmatrix}$ — вектор-столбцы; $\overline{e}_a = (\underbrace{1, \ldots, 1}_{a})$ и $\overline{m}^* = (m_1^* \ldots m_b^*)$—вектор-строки; штрихом обозначается операция транспонирования.

Наряду с множествами расположений $\mathscr{E}_\sigma^{\overline{m}} \subset R$ и $\mathscr{E}_\sigma^{\overline{m}^*} \subset R^*$ рассмотрим множество $\mathscr{E}_\rho^{\mathbf{m}}$ расположений $\rho^* \in R^*$ с расстоянием $\mathbf{m}(\rho, \rho^*)$ до $\rho = (\tau_1, \ldots \tau_2, \ldots, \tau_a) \in R$, в точности равным $\mathbf{m}$. Множество $\mathscr{E}_\rho^n$ имеет следующее представление

$$\mathscr{E}_\rho^{\mathbf{m}} = \bigtimes_{\alpha=1}^{a} \mathscr{E}_{\tau_\alpha}^{\overline{m}_\alpha} \quad (1.93)$$

в виде прямого произведения введенных в п. 1.3.1 множеств. Из соотношения (1.93) следует, что

$$N(\mathscr{E}_\rho^{\mathbf{m}}) = \prod_{\alpha=1}^{a} C_{m_\alpha}^{\overline{m}_\alpha}.$$

Изучим пересечения множеств $\mathscr{E}^{\mathbf{m}}_{\rho}$ при различных ρ и $\mathbf{m}$.

1.7.2 *Основная лемма о пересечениях.* Изучим пересечения множеств $\mathscr{E}^{\mathbf{m}_1}_{\rho_1}$ и $\mathscr{E}^{m_2}_{\rho_2}$ расположений $\rho^* \in R^*$, где $\mathbf{m}_1 = \mathbf{m}(\rho_1, \rho) = \| m^{(1)}_{\alpha\beta} \|$; $\mathbf{m}_2 = \mathbf{m}(\rho_2, \rho^*) = \| m^{(2)}_{\alpha'\beta} \|$, при условии, что расстояние между ρ_1 и $\rho_2 \in R$ в точности равно $\mathbf{m} = \mathbf{m}(\rho_1, \rho_2) = \| m_{\alpha\alpha'} \|$. При этом обязательно условие $\sum\limits_{\alpha} m^{(1)}_{\alpha\beta} = \sum\limits_{\alpha'} m^{(2)}_{\alpha'\beta} = m^*_{\beta}$.

Для этого введем в рассмотрение трехиндексные неотрицательные целые числа $m_{\alpha\alpha'\beta} \geqslant 0$ ($\alpha, \alpha' = \overline{1,a}$, $\beta = \overline{1,b}$), удовлетворяющие условиям

$$\sum_{\alpha=1}^{a} m_{\alpha\alpha'\beta} = m^{(2)}_{\alpha'\beta}, \quad \sum_{\alpha'=1}^{a} m_{\alpha\alpha'\beta} = m^{(1)}_{\alpha\beta}, \quad \sum_{\beta=1}^{b} m_{\alpha\alpha'\beta} = m_{\alpha\alpha'}. \tag{1.94}$$

Удобна также следующая векторная запись введенных чисел $\overline{m}_{\alpha\alpha'} = \{m_{\alpha\alpha'\beta}\}$ и матриц $\mathbf{m}_1 = (\overline{m}^{(1)}_{\alpha})$; $\mathbf{m}_2 = (\overline{m}^{(2)}_{\alpha'})$, с помощью которой соотношения (1.94) запишутся так

$$\sum_{\alpha=1}^{a} \overline{m}_{\alpha\alpha'} = \overline{m}^{(2)}_{\alpha'}, \quad \sum_{\alpha'=1}^{a} \overline{m}_{\alpha\alpha'} = \overline{m}^{(1)}_{\alpha}, \quad \overline{m}_{\alpha\alpha'} \cdot \overline{e}_b = m_{\alpha\alpha'}.$$

При этом обязательно условие $\overline{e}_a \mathbf{m}_1 = \overline{e}_a \mathbf{m}_2 = \overline{m}^*$.

Основная лемма о пересечениях 1.2. [10]

$$\mathscr{E}^{\mathbf{m}_1}_{\rho_1} \cap \mathscr{E}^{\mathbf{m}_2}_{\rho_2} = \begin{cases} \bigcup\limits_{(1.94)} \times\limits_{\alpha\alpha'} \mathscr{E}^{\overline{m}_{\alpha\alpha'}}_{m_{\alpha\alpha'}}, & \text{при } \overline{e}_a \mathbf{m}_1 = \overline{e}_a \mathbf{m}_2 = \overline{m}^* \\ \varnothing, & \text{при } \overline{e}_a \mathbf{m}_1 \neq \overline{e}_a \mathbf{m}_2, \end{cases} \tag{1.95}$$

где $\mathbf{m} = \mathbf{m}(\rho_1, \rho_2) = \| m_{\alpha\alpha} \|$; $\mathbf{m}_1 = \mathbf{m}(\rho_1, \rho^*) = \| m^{(1)}_{\alpha\beta} \|$; $\mathbf{m}_2 = \mathbf{m}(\rho_2, \rho^*) = \| m^{(2)}_{\alpha'\beta} \|$, и суммирование ведется по всем решениям $m_{\alpha\alpha'}$ системы (1.94).

Доказательство. Рассмотрим представления $\rho_1 = \{\tau_\alpha\}$, $\rho_2 = \{\tau'_{\alpha'}\}$ и $(\rho_1, \rho_2) = \{\tau_{\alpha\alpha'}\}$, где $\tau_{\alpha\alpha'} = \tau_\alpha \cap \tau'_{\alpha'}$. Согласно (1.93) имеем следующие представления

$$\mathscr{E}^{\mathbf{m}_1}_{\rho_1} = \times_{\alpha=1}^{a} \mathscr{E}^{\overline{m}^{(1)}_{\alpha}}_{\tau_\alpha} \quad \text{и} \quad \mathscr{E}^{\mathbf{m}_2}_{\rho_2} = \times_{\alpha'=1}^{a} \mathscr{E}^{\overline{m}^{(2)}_{\alpha'}}_{\tau'_{\alpha'}}, \tag{1.96}$$

Далее, используя соотношение (1.22), получим следующие представления

$$\mathscr{E}^{\overline{m}^{(1)}_{\alpha}}_{\tau_\alpha} = \bigcup_{\sum\limits_{\alpha'} \overline{m}^{(1)}_{\alpha\alpha'} = \overline{m}^{(1)}_{\alpha}} \times_{\alpha'=1}^{a} \mathscr{E}^{\overline{m}^{(1)}_{\alpha\alpha'}}_{\tau_{\alpha\alpha'}} \quad \text{и} \quad \mathscr{E}^{\overline{m}^{(2)}_{\alpha'}}_{\tau'_{\alpha'}} = \bigcup_{\sum\limits_{\alpha} \overline{m}^{(2)}_{\alpha\alpha'} = \overline{m}^{(2)}_{\alpha'}} \times_{\alpha=1}^{a} \mathscr{E}^{\overline{m}^{(2)}_{\alpha\alpha'}}_{\tau_{\alpha\alpha'}}, \tag{1.97}$$

где суммирование ведется по векторам $\overline{m}^{(1)}_{\alpha\alpha'} = (m^{(1)}_{\alpha\alpha'\beta})$ и $\overline{m}^{(2)}_{\alpha\alpha'} = (m^{(2)}_{\alpha\alpha'\beta})$ с трехиндексными компонентами.

Подставив выражения (1.97) в (1.96), учитывая перестановочность операций $\bigcup$ и $\times$, имеем окончательное представление

$$\mathscr{E}_{\rho_1}^{\mathbf{m}_1} = \bigcup_{\sum_{\alpha'} \bar{m}_{\alpha\alpha'}^{(1)} = m_\alpha^{(1)}} \times \mathscr{E}_{\tau_{\alpha\alpha'}}^{\bar{m}_{\alpha\alpha'}^{(1)}} \quad \text{и} \quad \mathscr{E}_{\rho_2}^{\mathbf{m}_2} = \bigcup_{\sum_{\alpha} \bar{m}_{\alpha\alpha'}^{(2)} = m_{\alpha'}^{(2)}} \times \mathscr{E}_{\tau_{\alpha\alpha'}}^{\bar{m}_{\alpha\alpha'}^{(2)}}. \tag{1.98}$$

Возьмем пересечение $\mathscr{E}_{\rho_1}^{\mathbf{m}_1} \cap \mathscr{E}_{\rho_2}^{\mathbf{m}_2}$ множеств $\mathscr{E}_{\rho_1}^{\mathbf{m}_1}$ и $\mathscr{E}_{\rho_2}^{\mathbf{m}_2}$, представленных в форме (1.98). Будем иметь

$$\mathscr{E}_{\rho_1}^{\mathbf{m}_1} \cap \mathscr{E}_{\rho_2}^{\mathbf{m}_2} = \bigcup_{\substack{\sum_{\alpha'} \bar{m}_{\alpha\alpha'}^{(1)} = \bar{m}_\alpha^{(1)} \\ \sum_{\alpha} \bar{m}_{\alpha\alpha'}^{(2)} = \bar{m}_{\alpha'}^{(2)}}} \times \mathscr{E}_{\tau_{\alpha\alpha'}}^{\bar{m}_{\alpha\alpha'}^{(1)}} \cap \mathscr{E}_{\tau_{\alpha\alpha'}}^{\bar{m}_{\alpha\alpha'}^{(2)}}. \tag{1.99}$$

Но согласно соотношению (1.13) имеем

$$\mathscr{E}_{\tau_{\alpha\alpha'}}^{m_{\alpha\alpha'}^{(1)}} \cap \mathscr{E}_{\alpha\alpha'}^{m_{\alpha\alpha'}^{(2)}} = \begin{cases} \mathscr{E}_{\tau_{\alpha\alpha'}}^{\bar{m}_{\alpha\alpha'}}, & \text{при } \bar{m}_{\alpha\alpha'}^{(1)} = \bar{m}_{\alpha\alpha'}^{(2)} = \bar{m}_{\alpha\alpha'}; \\ \varnothing, & \text{при } \bar{m}_{\alpha\alpha'}^{(1)} \neq \bar{m}_{\alpha\alpha'}^{(2)}. \end{cases}$$

Поэтому из соотношения (1.99) получим утверждение леммы. При этом условия суммирования переходят в условия (1.94), для наличия непустого пересечения необходимо выполнение условий $\bar{e}_a \mathbf{m}_1 = \bar{e}_a \mathbf{m}_2 = \bar{m}^*$.

Изучим специальный случай отсутствия непустого пересечения множеств $\mathscr{E}_{\rho_1}^{\mathbf{m}_1}$ и $\mathscr{E}_{\rho_2}^{\mathbf{m}_2}$. Пусть $R = R^*$, когда $a = b$. Определим на матричных расстояниях $\mathbf{m} = \mathbf{m}(\rho_1, \rho_2)$ числовую функцию

$$d[\mathbf{m}(\rho_1, \rho_2)] = d(\rho_1, \rho_2) = \sum_{\alpha \neq \alpha'} m_{\alpha\alpha'} = s - \sum_{\alpha=1}^{a} m_{\alpha\alpha}, \tag{1.100}$$

которую будем называть расстоянием Хэмминга[1]. Ясно, что одно и то же хеммингoвское расстояние имеют все пары (ρ_1, ρ_2) с матричными расстояниями $\mathbf{m}$ с одним и тем же следом $S(\mathbf{m}) = \sum_{\alpha=1}^{a} m_{\alpha\alpha}$. В соответствии с этим определим множество $\rho^* \in R^*$

$$\mathscr{E}_\rho^d = \bigcup_{d(\rho, \rho^*) = d} \mathscr{E}_\rho^{\mathbf{m}}.$$

Используя результаты основной леммы о пересечениях в терминах хэммингoвских расстояний, можно сформулировать условия, достаточные для того, чтобы множества $\mathscr{E}_{\rho_1}^{\mathbf{m}_1}$ и $\mathscr{E}_{\rho_2}^{\mathbf{m}_2}$ не пересекались.

Следствие 1.2.1. Пусть $d = d(\rho_1, \rho_2)$, $d_1 = d(\rho_1, \rho^*)$ и $d_2 = d(\rho_2, \rho^*)$, тогда если $d > d_1 + d_2$, то

$$\mathscr{E}_{\rho_1}^{d_1} \cap \mathscr{E}_{\rho_2}^{d_2} = \varnothing,$$

откуда

$$D_d(d_1, d_2) = N(\mathscr{E}_{\rho_1}^{d_1} \cap \mathscr{E}_{\rho_2}^{d_2}) = 0.$$

[1] Введенное расстояние (1,100) впервые рассматривалось им в [24] для бинарного случая $a=b=2$.

Доказательство. В самом деле, из определения чисел d и $m_{\alpha\alpha\beta}$ имеем:

$$\left.\begin{aligned} d &= \sum_{\alpha\neq\alpha'} m_{\alpha\alpha'} = \sum_{\substack{\alpha\neq\alpha'\\ \beta}} m^{(1)}_{\alpha\alpha'\beta} = \sum_{\alpha\neq\alpha'} m^{(1)}_{\alpha\alpha'\alpha} + \sum_{\substack{\alpha\neq\alpha'\\ \beta\neq\alpha}} m^{(1)}_{\alpha\alpha'\beta} = \\ &= \sum_{\substack{\alpha\neq\alpha'\\ \beta'}} m^{(2)}_{\alpha\alpha'\beta'} = \sum_{\alpha\neq\alpha'} m^{(2)}_{\alpha\alpha'\alpha'} + \sum_{\substack{\alpha\neq\alpha'\\ \beta'\neq\alpha'}} m^{(2)}_{\alpha\alpha'\beta'};\\ d_1 &= \sum_{\alpha\neq\beta} m^{(1)}_{\alpha\beta} = \sum_{\substack{\alpha\neq\beta\\ \alpha'}} m^{(1)}_{\alpha\alpha'\beta} = \sum_{\alpha\neq\beta} m^{(1)}_{\alpha\alpha\beta} + \sum_{\substack{\alpha\neq\alpha'\\ \beta\neq\alpha}} m^{(1)}_{\alpha\alpha'\beta};\\ d_2 &= \sum_{\alpha'\neq\beta'} m^{(2)}_{\alpha'\beta'} = \sum_{\substack{\alpha'\neq\beta'\\ \alpha}} m^{(2)}_{\alpha'\alpha'\beta'} = \sum_{\alpha'\neq\beta'} m^{(2)}_{\alpha'\alpha'\beta'} + \sum_{\substack{\alpha\neq\alpha'\\ \beta'\neq\alpha'}} m^{(2)}_{\alpha\alpha'\beta'}. \end{aligned}\right\} \quad (1.101)$$

Пусть $d > d_1 + d_2$. Но из (1.101) имеем

$$\sum_{\alpha\neq\alpha'} m^{(1)}_{\alpha\alpha'\alpha} + \sum_{\alpha\neq\alpha'} m^{(2)}_{\alpha\alpha'\alpha'} = 2d - \left(\sum_{\substack{\alpha\neq\alpha'\\ \beta\neq\alpha}} m^{(1)}_{\alpha\alpha'\beta} + \right.$$

$$\left. + \sum_{\substack{\alpha\neq\alpha'\\ \beta'\neq\alpha'}} m^{(2)}_{\alpha\alpha'\beta'} \right) \geqslant 2d - (d_1 + d_2) > d.$$

Откуда

$$\sum_{\alpha\neq\alpha'} (m^{(1)}_{\alpha\alpha'\alpha} + m^{(2)}_{\alpha\alpha'\alpha'}) > \sum_{\alpha\neq\alpha'} m_{\alpha\alpha'}.$$

Поэтому из-за неотрицательности чисел $m_{\alpha\alpha'\beta}$ найдется хотя бы одна пара значений α, α' $(\alpha \neq \alpha')$, при которой

$$m^{(1)}_{\alpha\alpha'\alpha} + m^{(2)}_{\alpha\alpha'\alpha'} > m_{\alpha\alpha'}. \quad (1.102)$$

Покажем, что условие (1.102) достаточно для того, чтобы пересечение соответствующих множеств $\mathscr{E}^{\overline{m}^{(1)}_{\alpha\alpha'}}_{\tau_{\alpha\alpha'}} \cap \mathscr{E}^{\overline{m}^{(2)}_{\alpha\alpha'}}_{\tau_{\alpha\alpha'}}$ было пусто. Рассмотрим τ-проекции $\rho^{(1)}_{\tau_{\alpha\alpha'}} \subset \mathscr{E}^{\overline{m}^{(1)}_{\alpha\alpha'}}_{\tau_{\alpha\alpha'}}$ и $\rho^{(2)}_{\tau_{\alpha\alpha'}} \subset \mathscr{E}^{\overline{m}^{(2)}_{\alpha\alpha'}}_{\tau_{\alpha\alpha'}}$ и покажем, что они не могут совпадать при соблюдении условия (1.102), т. е. пересечение рассматриваемых множеств пусто. В самом деле, символ α встречается в $\rho^{(1)}_{\tau_{\alpha\alpha'}}$ $m^{(1)}_{\alpha\alpha'\alpha}$ раз. Этот же символ встречается $\rho^{(2)}_{\tau_{\alpha\alpha'}}$ не более $m_{\alpha\alpha'} - m^{(2)}_{\alpha\alpha'\alpha'}$ раз. Но из условия (1.102) следует, что $m^{(1)}_{\alpha\alpha'\alpha} > m_{\alpha\alpha'} - m^{(2)}_{\alpha\alpha'\alpha'}$, т. е. $\overline{m}^{(1)}_{\alpha\alpha'} \neq \overline{m}^{(2)}_{\alpha\alpha'}$, что приводит к $\mathscr{E}^{\overline{m}^{(1)}_{\alpha\alpha'}}_{\tau_{\alpha\alpha}} \cap \mathscr{E}^{\overline{m}^{(2)}_{\alpha\alpha'}}_{\tau_{\alpha\alpha'}} = \varnothing$. Отсюда следует $\mathscr{E}^{\overline{m}_1}_{\rho_1} \cap \mathscr{E}^{\overline{m}_2}_{\rho_2} = \varnothing$ и утверждение следствия.

Из теоретико-множественных соотношений основной леммы следуют соответствующие числовые соотношения. Имеет место

Следствие 1.2.2

$$D_{\mathfrak{m}}(\mathfrak{m}_1, \mathfrak{m}_2) = N(\mathscr{E}^{\mathfrak{m}_1}_{\rho_1} \cap \mathscr{E}^{\mathfrak{m}_2}_{\rho_2}) = \begin{cases} \sum\limits_{(1.94)} \prod\limits_{\alpha\alpha'} C^{\overline{m}_{\alpha\alpha'}}_{m_{\alpha\alpha'}}, & \text{при } \overline{e}_a\mathfrak{m}_1 = \overline{e}_a\mathfrak{m}_2 = \overline{m}^*\\ 0, & \text{при } \overline{e}_a\mathfrak{m}_1 \neq \overline{e}_a\mathfrak{m}_2. \end{cases} \quad (1.103)$$

Доказательство. Для получения соотношения (1.103) достаточно взять функцию N от обеих частей соотношения (1.95),

1.7.3. *Производящие и некоторые неасимптотические оценки.* Из соотношения (1.103) можно получить компактное выражение для производящей функции (второй производящей) с весами чисел $D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2)$. В самом деле, используя обобщенную формулу бинома (1.19), получим

$$g(\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2) = \left[\sum_{\alpha\alpha'} u_{\alpha\alpha'} \sum_{\beta} u^{(1)}_{\alpha\beta} u^{(2)}_{\alpha'\beta}\right]^s =$$

$$= \sum_{\mathbf{m}} C^{\mathbf{m}}_s \mathbf{u}^{\mathbf{m}} \prod_{\alpha\alpha'} \left(\sum_{\beta} u^{(1)}_{\alpha\beta} u^{(2)}_{\alpha'\beta}\right)^{m_{\alpha\alpha'}} =$$

$$= \sum_{\mathbf{m}} C^{\mathbf{m}}_s u^{\mathbf{m}} \prod_{\alpha\alpha'} \sum_{\sum_{\beta} m_{\alpha\alpha'\beta} = m_{\alpha\alpha'}} C^{\overline{m}_{\alpha\alpha'}}_{m_{\alpha\alpha'}} \prod_{\beta} (u^{(1)}_{\alpha\beta} u^{(2)}_{\alpha'\beta})^{m_{\alpha\alpha\beta'}} =$$

$$= \sum_{\mathbf{m}} C^{\mathbf{m}}_s \mathbf{u}^{\mathbf{m}} \sum_{\mathbf{m}_1, \mathbf{m}_2} \sum_{(1.94)\alpha\alpha'} \prod C^{\overline{m}_{\alpha\alpha'}}_{m_{\alpha\alpha'}} \cdot \mathbf{u}_1^{\mathbf{m}_1} \mathbf{u}_2^{\mathbf{m}_2} = \sum_{\mathbf{m}_1, \mathbf{m}_2} \left(\sum_{\mathbf{m}} C^{\mathbf{m}}_s D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2)\, \mathbf{u}^{\mathbf{m}}\right) \mathbf{u}_1^{\mathbf{m}_1} \mathbf{u}_2^{\mathbf{m}_2},$$

где $\mathbf{u} = \| u_{\alpha\alpha'} \|$, $\mathbf{u}_1 = \| u^{(1)}_{\alpha\beta} \|$; $\mathbf{u}_2 = \| u^{(2)}_{\alpha\beta} \|$ — формальные матричные аргументы.

Введем обозначения:

$$g_1(\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2) = \sum_{\alpha\alpha'} \mathbf{u}_{\alpha\alpha'} \sum_{\beta} \mathbf{u}^{(1)}_{\alpha\beta} \mathbf{u}^{(2)}_{\alpha'\beta};$$

$$g_{\mathbf{m}_1\mathbf{m}_2}(\mathbf{u}) = \sum_{\mathbf{m}} C^{\mathbf{m}}_s \mathbf{u}^{\mathbf{m}} D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2), \tag{1.104}$$

тогда

$$g(\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2) = [g_1(\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2)]^s = \sum_{m_1 m_2} g_{\mathbf{m}_1\mathbf{m}_2}(\mathbf{u})\, \mathbf{u}_1^{\mathbf{m}_1} \mathbf{u}_2^{\mathbf{m}_2}. \tag{1.105}$$

Пусть $\mathbf{u}_1$ и $\mathbf{u}_2 > 0$, т. е. элементы матриц положительны. Тогда из (1.105) имеем не асимптотическую оценку

$$g_{\mathbf{m}_1\mathbf{m}_2}(\mathbf{u}) \leqslant [g_1(\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2)]^s / \mathbf{u}_1^{\mathbf{m}_1} \mathbf{u}_2^{\mathbf{m}_2} =$$
$$= e^{s[h(\mu_1, u_1) + h(\mu_2, u_2) + \ln g_1(u, u_1, u_2)]}, \tag{1.106}$$

где

$$\mu_1 = \frac{1}{s} \mathbf{m}_1 = \| \mu_{\alpha,1} \mu^{\beta}_{\alpha,1} \|, \quad \mu_2 = \frac{1}{s} \mathbf{m}_2 = \| \mu_{\alpha,2} \mu^{\beta}_{\alpha,2} \|$$

и

$$h(\mu, \mathbf{u}) = -\sum_{\alpha\beta} \mu_{\alpha} \mu^{\beta}_{\alpha} \ln u_{\alpha\beta}.$$

Положим

$$\mathbf{u} = \| u_{\alpha} u_{\alpha'} \|, \quad \mathbf{u}_1 = \mathbf{u}_2 = \| u^{(1)}_{\alpha\beta} \| = \| u^{\beta}_{\alpha} / \sqrt{\textstyle\sum_{\gamma} u_{\gamma} u^{\beta}_{\gamma}} \|,$$

где

$$\sum_{\alpha} u_{\alpha} = \sum_{\beta} u^{\beta}_{\alpha} = 1,$$

тогда

$$g_1(\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2) = \sum_{\alpha\alpha'} u_\alpha u_{\alpha'} \sum_\beta \frac{u_\alpha^\beta u_{\alpha'}^\beta}{\sum_\gamma u_\gamma u_\gamma^\beta} = \sum_\beta \sum_\alpha u_\alpha u_\alpha^\beta \equiv 1$$

и оценка (1.106) упростится

$$g_{\mathbf{m}_1\mathbf{m}_2}(\mathbf{u}) \leqslant e^{s[h(\mu_1,u_1)+h(\mu_2,u_1)]}. \qquad (1.107)$$

Далее, используя выражение (1.104), получим из оценки (1.107) асимптотическую оценку при $s \to \infty$

$$D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2) \leqslant g_{\mathbf{m}_1\mathbf{m}_2}(\mathbf{u})/C_s^{\mathbf{m}}\mathbf{u}^{\mathbf{m}} \leqslant e^{s[h(\mu_1,u_1)+h(\mu_2,u_1)-h(\mu,u)-h(\mu)]+O(1us)}.$$

Заметим, что вторая производящая (1.105) имеет второе представление

$$g(\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2) = \left(\sum_{\alpha\alpha'} u_{\alpha\alpha'} \sum_\beta u_{\alpha\beta}^{(1)} u_{\alpha'\beta}^{(2)}\right)^s = \sum_{\mathbf{m}} C_s^{\mathbf{m}}\mathbf{u}^{\mathbf{m}} g_{\mathbf{m}}(\mathbf{u}_1, \mathbf{u}_2),$$

где

$$g_{\mathbf{m}}(\mathbf{u}_1, \mathbf{u}_2) = \prod_{\alpha\alpha'} \left(\sum_\beta u_{\alpha\beta}^{(1)} u_{\alpha'\beta}^{(2)}\right)^{m_{\alpha\alpha'}} = \sum_{\mathbf{m}_1\mathbf{m}_2} D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2)\, \mathbf{u}_1^{\mathbf{m}_1}\mathbf{u}_2^{\mathbf{m}_2} \qquad (1.108)$$

производящая функция чисел $D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2)$.

1.7.4. *Бинарный случай.* Рассмотрим частный случай двойных расположений, когда $a = b = 2$. В этом случае векторы $\overline{m} = (m, s-m)$ и $\overline{m}^* = (m^*, s - m^*)$ при фиксированном s определяются своими первыми компонентами.

Матрицы

$$\mathbf{m} = \begin{pmatrix} m_{11} & m - m_{11} \\ m^* - m_{11} & s - m - m^* + m_{11} \end{pmatrix},$$

при фиксированных векторах $\overline{m}$ и $\overline{m}^*$ определяются заданием единственного элемента m_{11}. Однако система (1.94) не имеет единственного решения и общие соотношения не упрощаются.

Существенное упрощение здесь можно получить при переходе к множествам $\mathscr{C}_\rho^d$, когда решение системы (1.94) единственно (бинарный симметричный случай; см. **§ 8.4**).

В самом деле, рассмотрим производящую (1.108) для случая $a = b = 2$. Имеем

$$g_{\mathbf{m}}(\mathbf{u}_1, \mathbf{u}_2) = (u_{11}^{(1)}u_{11}^{(2)} + u_{12}^{(1)}u_{12}^{(2)})^{m_{11}} (u_{11}^{(1)}u_{21}^{(2)} + u_{12}^{(1)}u_{22}^{(2)})^{m_{12}} \times$$

$$\times (u_{21}^{(1)}u_{11}^{(2)} + u_{22}^{(1)}u_{12}^{(2)})^{m_{21}} (u_{21}^{(1)}u_{21}^{(2)} + u_{22}^{(1)}u_{22}^{(2)})^{m_{22}} = \sum_{\mathbf{m}_1\mathbf{m}_2} D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2)\, \mathbf{u}_1^{\mathbf{m}_1}\mathbf{u}_2^{\mathbf{m}_2}. \qquad (1.109)$$

Легко видеть, что для получения производящей чисел $D_d(d_1, d_2)$, где $d = m_{12} + m_{21}$, $d_1 = m_{12}^{(1)} + m_{21}^{(1)}$ и $d_2 = m_{12}^{(2)} + m_{21}^{(2)}$, достаточно положить в (1.109) $\mathbf{u}_1^{(0)} = \begin{pmatrix} 1 & u \\ u & 1 \end{pmatrix}$ и $\mathbf{u}_2^{(0)} = \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}$,

тогда

$$g_{\mathbf{m}}(\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)}) = g_d(u, v) = (uv+1)^{s-d}(u+v)^d = \sum_{d_1 d_2} D_d(d_1, d_2)\, u^{d_1} v^{d_2}. \qquad (1.110)$$

Преобразуем выражение (1.110), используя бином Ньютона

$$g_d(u, v) = (uv + 1)^{s-d}(u + v)^d = \sum_{k,\, l} C_d^k C_{s-d}^l u^{k+l} v^{l+d-k}.$$

Положим

$$\left.\begin{aligned} k + l &= d_1 \\ d - (k - l) &= d_2 \end{aligned}\right\}, \tag{1.111}$$

откуда при условии $d + d_1 + d_2 = 0 \pmod 2$ и $d \leqslant d_1 + d_2$ (см. следст. 1.2.1) у системы (1.111) существует единственное целочисленное решение

$$k = (d_1 - d_2 + d)/2, \quad l = (d_1 + d_2 - d)/2.$$

В случае $d + d_1 + d_2 \equiv 1 \pmod 2$ или $d > d_1 + d_2$ система (1.111) не имеет целочисленных решений.

Итак

$$g_d(u, v) = \sum_{d_1\, d_2} C_d^{\frac{d_1 - d_2 + d}{2}} C_{s-d}^{\frac{d_1 + d_2 - d}{2}} u^{d_1} v^{d_2} = \sum_{d_1\, d_2} D_d(d_1, d_2)\, u^{d_1} v^{d_2},$$

откуда имеем окончательно

$$D_d(d_1, d_2) = \begin{cases} C_d^{\frac{d_1 - d_2 + d}{2}} C_{s-d}^{\frac{d_1 + d_2 - d}{2}}, & \text{при } d + d_1 + d_2 = 0 \bmod 2 \\ 0, & \text{при } d + d_1 + d_2 \equiv 1 \bmod 2. \end{cases} \tag{1.112}$$

Этот результат впервые был получен в [7]. Прямой вывод соотношения (1.112) без использования производящих содержится в [6]. Этот результат послужил основой для дальнейших обобщений, приведших к основной лемме о пересечениях. Ряд интересных асимптотических выражений, связанных с (1.112), см. в **§ 8.4.**

Глава 2

СТОХАСТИЧЕСКИЕ ФУНКЦИИ

§ 2.1. Вводные замечания

2.1.1. *Стохастические функции.* Рассмотрим детерминированную функцию $F(X_k) = Y_l$, соотносящую один объект произвольной природы $X_k \in X = (X_1, \ldots, X_k)$ другому объекту произвольной природы[1] $Y_l \in Y = (Y_1, \ldots, Y_L)$.

Пусть объекты X имеют случайную природу, то есть на них определены вероятности

$$\mathcal{P}(X_k)\,(k = \overline{1, K} \sum_{k=1}^{K} \mathcal{P}(X_k) = 1). \tag{2.1}$$

Тогда объекты Y также будут иметь случайную природу и называться функцией случайных объектов X. Здесь случайность Y является следствием случайности X.

Пусть объекты X детерминированы и заданы условные вероятности объектов Y, при фиксированных объектах X

$$\mathcal{P}(Y_l/X_k)\,(k = \overline{1, K}, l = \overline{1, L}, \sum_{l=1}^{L} \mathcal{P}(Y_l/X_k) = 1). \tag{2.2}$$

Тогда объекты Y будут иметь случайную природу и называться случайной функцией объектов X. Здесь случайность Y является следствием стохастической зависимости $\mathcal{P}(Y_l/X_k)$ объектов Y от детерминированных объектов X (в отличие от детерминированной зависимости $F(X_k) = Y_l$).

Понятие стохастической функции Y является естественным обобщением двух введенных выше понятий, когда объект X имеет стохастическую природу, определяемую (2.1), и объект Y находится в стохастической зависимости от объекта X, определяемой (2.2). Здесь случайность объекта Y является следствием случайности объекта X и стохастической зависимости объекта Y от объекта X. Точное определение понятия дискретной стохастической функции, а также соображения о неестественности сведения его к существующему общему понятию стохастического процесса приведены в п. 2.7.2.

2.1.2. *Исторические замечания.* Впервые схема стохастических функций встречается в 1912 г. в работе А. А. Маркова «Об испытаниях, связанных в цепь ненаблюдаемыми событиями» [8]. Долгие годы после опубликования этой работы схема стохастических функций не привлекала внимания ни отечественных, ни зарубежных ученых. Лишь в 1939 г. в работе В. И. Романовского «О цепных корреляциях» [9] эта

---

[1] Индексы k и l могут пробегать произвольное множество значений, в том числе и бесконечное, однако мы будем ограничиваться конечными множествами.

схема А. А. Маркова была рассмотрена вновь. Впоследствии, в 1949 г., в монографии «Дискретные цепи Маркова» [25] В. И. Романовский посвятил этой схеме всю гл. V, где отмечает (стр. 265), что до работы [9] он считал эту схему новой [1]. Автор некоторое время также считал, что схема стохастической функции нова, пока не узнал из [25] об аналогичном заблуждении В. А. Романовского.

Отметим, что А. А. Марков пришел к схеме стохастических функций, углубляя схему испытаний, связанных в цепь (см. п. 2.3.1), а В. И. Романовский пришел к ней, обобщая понятие корреляции математической статистики на случай цепной связи выборочных значений (см. гл. 3) — «цепные корреляции». Ни тот, ни другой не приводили каких-либо приложений рассмотренной ими вероятностной модели. Это, по-видимому, явилось одной из причин того, что схема стохастических функций периодически забывалась.

В гл. 4 будет подробно показано, что эта вероятностная модель описывает выход канала связи с шумами, к которому подключен источник сообщений. Эта общая схема теории связи привлекла за последние годы внимание многих ученых радиоинженеров и математиков, особенно благодаря фундаментальным исследованиям В. А. Котельникова [1] и К. Шеннона [2]. Однако развитие их идей наталкивается на ряд принципиальных аналитических трудностей, связанных с неадэкватностью аппарата современного математического анализа с существенно дискретной спецификой рассмотрений.

2.1.3. *Дискретные рассмотрения.* Традиционный математический переход от дискретных рассмотрений к непрерывным диктуется в основном физической проблематикой и в ряде случаев упрощает решение сложных в дискретной постановке задач. Иногда в непрерывной постановке такие задачи имеют стандартное решение методами математического анализа. Однако решение многих современных математических задач, связанных в основном с развитием новой техники, выходят за рамки существующих непрерывных математических методов. Это объясняется не только ограниченностью числа задач, допускающих простые «формульные» решения непрерывными методами. По-видимому, дискретные рассмотрения характерны для всей теоретической проблематики современной техники. Они восходят к известной теореме В. А. Котельникова о дискретном представлении непрерывных сигналов с ограниченным спектром.

Дискретизация по амплитуде, вплоть до предельно «грубой» бинарной, оправдывается возможностями использования универсальных машин дискретного счета. Но и чисто аналитические дискретные рассмотрения начинают занимать все большее место в математическом аппарате области знаний, которую теперь принято называть кибернетикой. Конечность и дискретность рассмотрений считаются характерными особенностями этой новой области, абстрагирующейся от физических вещественных и энергетических особенностей сложных технических и природных объектов [27]. При этом внимание концентрируется на целенаправленном функционировании таких сложных объектов, состоящих из конечного, но весьма большого числа элементов. Сами элементы при этом рассматриваются как «черные ящики», без попыток проникновения во внутренний механизм их реакций на внешние воздействия.

---

[1] По-видимому, речь идет о его более ранней монографии 1938 г. «Математическая статистика» [26], где схеме стохастических функций посвящены стр. 303—307 и 480 без ссылок на [25].

Наряду с указанными особенностями кибернетики одной из важнейших ее черт является допущение элемента случайности как во внешних воздействиях, так и в реакциях на них черного ящика (в теории игр доказывается, что в некоторых ситуациях случайные реакции предпочтительнее детерминированных). Описанная выше схема стохастических функций является математической моделью черного ящика. Рассматриваемый далее дискретный вариант этой модели учитывает в основном специфику частного случая черного ящика — дискретного канала с шумами, с источником сообщений на входе и искаженными шумами сообщениями на выходе. Однако полученные общие соотношения могут быть использованы для исследования и других кибернетических вариантов черных ящиков.

2.1.4. *Содержание главы.* Для главы характерен «событийный», а не «величинный» стиль изложения вероятностного материала, так как изложение в терминах случайных величин предпочтительнее для физических приложений. Рассмотрение лишь дискретных и конечных образований освобождает от большого числа условностей и принципиальных трудностей, связанных с непрерывным подходом. При этом не исключается возможность асимптотических рассмотрений, приводящих к упрощениям сложных конечных выражений. Более того, именно для асимптотического случая начинают проявляться основные закономерности, выявление которых составляет цель второй части книги. Это не противоречит общей тенденции ограничиться дискретными и конечными построениями, если асимптотику рассматривать как случай очень большого, но все же конечного. Оценка порядков пренебрегаемых членов всюду способствует такому рассмотрению.

Перейдем к описанию содержания главы 2. В § 2.2 приводится точное определение дискретного стохастического аргумента. Символика гл. 1 и матричные обозначения позволяют дать компактную запись вводимых понятий.

В § 2.3 и 2.4 рассмотрены частные случаи дискретного стохастического аргумента, при которых упрощаются общие соотношения для распределений или производящих.

В § 2.5 указаны известные предельные распределения для дискретного стохастического аргумента при возрастании числа s моментов времени, на которых определены его значения и частоты событий.

Иллюстрации использования общих соотношений в простейшем бинарном случае посвящен § 2.6. Помимо упрощения общих соотношений, в этом случае удается получить ряд новых соотношений, вывод которых для общего случая затруднителен.

В § 2.7 вводится точное определение дискретной стохастической зависимости и дискретной стохастической функции. Рассмотрение стохастической зависимости в отличие от стохастического аргумента ограничивается простейшим случаем независимых переходов в дискретные моменты времени. Основным результатом является установление аналитических связей между производящими стохастического аргумента и функции.

В § 2.8 на основании основного соотношения между производящими устанавливаются связи между моментами распределений частот стохастического аргумента и функции.

Наконец, в § 2.9 общие соотношения гл. 2 иллюстрируются в простейшем бинарном случае.

§ 2.2. Дискретный стохастический аргумент

2:2.1. *Общее определение дискретного стохастического аргумента (СА).* Рассмотрим упорядоченное множество мест $\sigma = (1, 2, \ldots, t, \ldots, s)$, элементы которого будем называть моментами времени. Обобщая рассмотрения гл. 1, соотнесем каждому моменту t неупорядоченное множество предметов[1]

$$A^t = (A_1^{(t)}, \ldots, A_\alpha^{(t)}, \ldots, A_a^{(t)}) (t = \overline{1, s}).$$

Рассмотрим множество R обобщенных расположений $x \in R$, определяемых обобщенным соотношением

$$x = \{A_{\alpha_1}^{(1)}, \ldots, A_{\alpha_t}^{(t)}, \ldots, A_{\alpha_s}^{(s)}\} = \bigtimes_{t=1}^{s} \rho_{\{t\}}^{\alpha_t}, \tag{2.3}$$

где элементарная τ-проекция $\rho_{\{t\}}^{\alpha_t}$ отображает в момент t элементы лишь множества $A^{(t)}$.

Ясно, что второе представление расположений гл. 2 в нашем общем случае не имеет смысла.

Определим также множества R_τ обобщенных τ-проекций

$$x_\tau = \{A_{\alpha_{i_1}}^{(i_1)}, \ldots, A_{\alpha_{i_k}}^{(i_k)}, \ldots, A_{\alpha_{i_t}}^{(i_t)}\} = \bigcup_{i_k \in \tau} \rho_{\{i_k\}}^{\alpha_{i_k}}, \tag{2.4}$$

где $\tau = (i_1, \ldots, i_k, \ldots, i_t) \subseteq \sigma$, в частности при $\tau = \sigma$ $x_\sigma = x$.

Совокупность обобщенных τ-проекций x_τ стохастически определена[2], если заданы вероятности

$$\mathcal{P}(x_\tau) = p(x_\tau) \tag{2.5}$$

для всех $x_\tau \in R_\tau$ и $\tau \subseteq \sigma$.

Набор $p(x_\tau)$ вероятностей (2.5) не произволен, а должен удовлетворять условиям согласованности, следующим из общих свойств вероятностей,

$$\left.\begin{aligned} &1.\ \sum_{\alpha_{i_k}=1}^{a} p((\alpha_{i_1}, \ldots, \alpha_{i_{k-1}} \alpha_{i_k} \alpha_{i_{k+1}}, \ldots, \alpha_{i_t})) = \\ &\quad = p((\alpha_{i_1}, \ldots, \alpha_{i_{k-1}} \alpha_{i_{k+1}}, \ldots, \alpha_{i_t})) \\ &\text{и нормировки} \\ &2.\ \sum_{x \in R} p(x) = 1. \end{aligned}\right\} \tag{2.6}$$

Из условия 1 следует, что для задания набора $p(x_\tau)$ вероятностей (2.5) достаточно задать лишь набор $p(x)$ вероятностей обобщенных расположений $(x \in R)$ в виде a^s произвольных неотрицательных величин, дающих в сумме единицу. Тогда суммированием по всем α_i с $i \in \sigma - \tau$ получим

$$p(x_\tau) = \sum_{\sigma - \tau} p(x). \tag{2.7}$$

---

[1] В гл. 1 все множества $A^{(t)}$ были тождественны между собой. В обобщенной схеме можно считать, что и число a_t элементов $A^{(t)}$ зависит от t. Тогда можно формально положить $a = \max_{1 \leqslant t \leqslant s} a_t$ и при дальнейших стохастических рассмотрениях некоторым не возникающим предметам $A_\alpha^{(t)}$ приписывать нулевую вероятность.

[2] Объект стохастически определен, если известна вероятность его появления.

Набор вероятностей $p(x_\tau)$, удовлетворяющий соотношениям (2.6), будем называть стохастическим аргументом (*СА*) и обозначать $\mathcal{P}(A)$. Обобщенные расположения x будем называть значениями СА, обобщенные τ-проекции x_τ — τ-значениями СА, а предметы $A_\alpha^{(t)}$ — элементарными значениями СА. Так определенный СА является частным дискретным, конечным во времени случаем общего понятия вероятностного процесса. Для случая $a = 2$ приведенное определение СА содержится в [28].

Если все множества $A^{(t)}$ тождественны $A^{(t)} = A\,(t = \overline{1, s})$, то будем говорить, что имеет место однородный СА. В этом случае τ-значения СА x соответствуют τ-проекциям ρ_τ гл. 1 и имеет смысл второе их представление в виде

$$(x_\tau = \tau_1, \ldots, \tau_\alpha, \ldots, \tau_a)\left(\bigcup_{\alpha=1}^{a} \tau_\alpha = \tau, \tau_\alpha \cap \tau_{\alpha'} = \varnothing\right),$$

где τ_α означает множество моментов τ, в которых появляется элементарное значение СА A_α.

Числа $N(\tau_\alpha) = t_\alpha\,(\alpha = \overline{1, a})$ будем называть частотами СА. Составленный из них вектор частот $\bar{t} = (t_1, \ldots, t_\alpha, \ldots, t_a)$ таков, что он имеет целочисленные неотрицательные компоненты, дающие в сумме величину $t = N(\tau)$.

Наряду с распределением $p(x_\tau)$ τ-значений СА в дальнейшем будет изучаться распределение $P_t(\bar{t})$ частот СА, которое определяется в виде

$$P_t(\bar{t}) = \sum_{x_\tau \in \mathscr{E}_\tau^{\bar{t}}} p(x_\tau), \tag{2.8}$$

где множество $\mathscr{E}_\tau^{\bar{t}}$ имеет смысл, указанный в гл. 1.

2.2.2. *Производящие СА.* Для дальнейшего изложения удобно использование производящих СА. Чтобы определить их, будем интерпретировать элементарные значения СА $A_\alpha^{(t)}\,(\alpha = \overline{1, a})$ α-мерными вектор-столбцами

$$\bar{e}_\alpha = \overset{\alpha}{\left\{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right\}} a\,(\alpha = \overline{1, a}).$$

Тогда представлению (2.3) значения x СА соответствует матричное представление

$$\mathbf{x} = \{\bar{e}_{\alpha_1}, \ldots, \bar{e}_{\alpha_t}, \ldots, \bar{e}_{\alpha_s}\}$$

в виде $(a \times s)$-матрицы, состоящей из s вектор-столбцов.

Аналогично τ-значение x_τ СА имеет соответствующее (2.4) представление

$$\mathbf{x}_\tau = \{\bar{e}_{\alpha_{i_1}}, \ldots, \bar{e}_{\alpha_{i_k}}, \ldots, \bar{e}_{\alpha_{i_t}}\}$$

в виде $(a \times t)$-подматрицы матрицы x.

Введем формальный матричный аргумент $\mathbf{u} = (\bar{u}_1, \ldots, \bar{u}_t, \ldots, \bar{u}_s)$ в виде $(a \times s)$-матрицы, состоящий из s a-мерных вектор-столбцов

$$\bar{u}_t = \begin{pmatrix} u_1^{(t)} \\ \vdots \\ u_\alpha^{(t)} \\ \vdots \\ u_a^{(t)} \end{pmatrix}.$$

Тогда, в соответствии с определением п. 1.5.1, производящая СА имеет вид

$$g(\mathbf{u}) = \sum_{\mathbf{x} \in R} p(\mathbf{x})\, \mathbf{u}^{\mathbf{x}}. \tag{2.9}$$

Такого рода производящие без матричной символики для $a = 2$ впервые рассмотрены в [28]. Далее обозначим через $\mathbf{u}_\tau$ $(a \times t)$-матрицу, получающуюся из матрицы $\mathbf{u}$ заменой в последней столбцов, стоящих на местах τ a-мерными единичными вектор-столбцами $\bar{e} = \left(\begin{smallmatrix} 1 \\ \vdots \\ 1 \end{smallmatrix}\right) \Big\} a$. Тогда, подставив в (2.9) $\mathbf{u}_\tau$ вместо $\mathbf{u}$, в соответствии с соотношением (2.7) получим

$$g(\mathbf{u}_\tau) = \sum_{\mathbf{x}_\tau \in R_\tau} p(\mathbf{x}_\tau)\, \mathbf{u}_\tau^{\mathbf{x}_\tau}.$$

Рассмотрим случай тождественных $A^{(t)} = A\,(t = \overline{1, s})$, тогда a-мерный вектор-столбец частот $\bar{t}$ имеет следующее выражение

$$\bar{t} = \mathbf{x}_\tau \bar{e} = \sum_{k=1}^{t} \bar{e}_{\alpha_{i_k}}, \tag{2.10}$$

где $\bar{e}$ — t-мерный единичный вектор-столбец.

Тогда из соотношений (2.8) и (2.10) следует, что производящая частот $\bar{t}$ выражается через производящую τ-значений СА соотношением

$$g_\tau(\bar{u}) = g(\mathbf{u}_\tau)\big|_{\bar{u}_{\alpha_{i_k}} = \bar{u}} = \sum_{\bar{t}} P_t(\bar{t})\, \bar{u}^{\bar{t}}, \tag{2.11}$$

в котором все вектор-столбцы матрицы $\mathbf{u}_\tau$ заменены одним и тем же вектор-столбцом $\bar{u}$.

2. 2. 3. *Группировка. Неравенство Буля.* Важными операциями, которые можно проводить над СА, в общем случае являются группировка значений и временная группировка. Опишем первую операцию.

В самом деле, разобьем множество индексов $[1, 2, \ldots, \alpha, \ldots, a]$ на $a' \leqslant a$ непересекающихся подмножеств $I_i\,(i = \overline{1, a'})$

$$\bigcup_{i=1}^{a'} I_i = \{1, \ldots, \alpha, \ldots, a\},\ I_i \cap I_{i'} = \varnothing \text{ (при } i \neq i').$$

Будем называть группированными элементарными значениями

$$\widetilde{A}_{I_i}^{(t)} = \bigcup_{\alpha \in I_i} A_\alpha^{(t)}\,(i = \overline{1, a'}).$$

Соответствующие группированные τ-значения имеют вид

$$\widetilde{\mathbf{x}}_\tau = \{\widetilde{A}^{(1)}_{l_i}, \ldots, \widetilde{A}^{(k)}_{l_{i_k}}, \ldots, \widetilde{A}^{(t)}_{l_{i_t}}\}.$$

Тогда определенный из них СА будем называть группированием СА. Соответствующий набор вероятностей имеет вид

$$p(\widetilde{\mathbf{x}}_\tau) = \sum_{\alpha \in I_{i_k}(k=\overline{1,t})} p(\mathbf{x}_\tau).$$

Определим теперь временную группировку t-го порядка. В отличие от рассмотренной группировки она приводит не к укрупнению, а в некотором смысле к размельчению элементарных значений СА. В самом деле, пусть множество σ моментов времени состоит из $s = t \cdot s'$ моментов; тогда σ можно представить в виде $\sigma = \bigcup_{i=1}^{s'} \tau_i$, где $\tau_i = ((i-1)t+1, \ldots, it)$ $(i = \overline{1, s'})$.

Примем множества τ_i за новые группированные моменты времени (места); тогда множество моментов времени имеет вид

$$\sigma' = (\tau_1, \ldots, \tau_i, \ldots, \tau_{s'}).$$

За группированные элементарные значения A_{τ_i} примем τ-значения $\mathbf{x}_{\tau_i}$. Они составляют совокупность $A_i = \{x_{\tau_i}\} = R_{\tau_i}$, состоящую из a^t элементов.

Так как $\mathbf{x} = \bigcup_{t=1}^{s} p^{\alpha_{(t)}}_{(t)} = \bigcup_{i=1}^{s'} x_{\tau_i}$, то значения СА, построенные из группированных элементарных значений, оказываются одновременно построенными из негруппированных элементарных значений и поэтому имеют ту же вероятность $p(x)$.

Без дополнительных специальных предположений о структуре вероятностей (2.6) могут быть получены лишь некоторые общие неравенства [28].

Важнейшее из них так называемое неравенство Буля будет существенно использовано в дальнейшем изложении и поэтому ниже приводится его простой вывод. В самом деле, из общего теоретико-множественного соотношения (1.4) имеем для соответствующих вероятностей [1]

$$\min_{i=\overline{1,M}} \mathcal{P}(\mathscr{E}_i) \geqslant \mathcal{P}(\mathscr{E}) = \mathcal{P}\Big(\bigcap_{i=1}^{M} \mathscr{E}_i\Big) \geqslant 1 - \sum_{i=1}^{M} [1 - \mathcal{P}(\mathscr{E}_i)] =$$

$$= \sum_{i=1}^{M} \mathcal{P}(\mathscr{E}_i) - (M-1). \qquad (2.12)$$

Соотношения (2.12) называются неравенствами Буля. Используем их для оценки вероятностей τ-значений $p(\mathbf{x}_\tau)$ по вероятностям элементарных значений

$$p(A^{(t)}_\alpha) = p^{(t)}_\alpha = \sum_{\alpha_1, \ldots, \alpha_{t-1}\, \alpha_{t+1} \ldots \alpha_s} p(\mathbf{x})\,(\alpha = \overline{1, a}).$$

[1] Если рассматриваемые множества $\mathscr{E}_i$ конечны, то соотношение (2.12) сохраняется при замене символа $\mathcal{P}$ на символ N.

Имеем

$$p_{\alpha_{i_k}}^{(t)} \geqslant p(\mathbf{x}_\tau) = p((A_{\alpha_{i_1}}, \ldots, A_{\alpha_{i_k}}, \ldots; A_{\alpha_{i_k}})) \geqslant \sum_{k=1}^{t} p_{\alpha_{i_k}}^{(t)} - (t-1).$$

Наиболее существенные нижние оценки Буля эффективны лишь при близости соответствующих вероятностей $\mathcal{P}(\mathscr{E}_i)$ к единице.

Для возможности дальнейших положительных исследований стохастического аргумента необходимы специальные предположения о структуре вероятностей (2.5).

§ 2.3. Марковский аргумент

2.3.1. *Общий случай.* Будем называть СА марковским аргументом k-го порядка [25], если условная вероятность появления элементарного значения $A_{\alpha_t}^{(t)}$ в t-й момент при наступлении некоторого τ-значения $\mathbf{x}_\tau$, где $\tau = (1, 2, \ldots, t-1)$, имеет вид

$$\mathcal{P}(A_{\alpha_t}^{(t)}/\mathbf{x}_\tau) = \begin{cases} \mathcal{P}(A_{\alpha_1}^{(1)}), & \text{если } t = 1 \\ \mathcal{P}(A_{\alpha_t}^{(t)}/\mathbf{x}_{\tau'}), & \text{если } 2 \leqslant t \leqslant s, \end{cases}$$

причем

$$\tau' = \begin{cases} (1, 2, \ldots, t-1), & \text{если } 2 \leqslant t \leqslant k \\ (t-k, t-k+1, \ldots, t+1), & \text{если } k+1 \leqslant t \leqslant s. \end{cases}$$

Другими словами, вероятность наступления $A_{\alpha_t}^{(t)}$ зависит лишь от значений, имевших место в k предшествующих моментах времени, и не зависит от значений, имевших место в более ранние моменты. При $k = 1$ марковский аргумент называется простым марковским аргументом. Легко показать [29], что временная группировка $k+1$-го порядка марковского аргумента k-го порядка приводит к простому группированному марковскому аргументу. Поэтому в дальнейшем мы будем изучать лишь последний и называть его коротко марковским аргументом. Структура марковского аргумента не оправдывает рассмотрения вероятностей произвольных τ-значений. Поэтому будем рассматривать лишь вероятности значений $\mathbf{x}$. Исходя из определения марковского аргумента, они имеют вид

$$p(\mathbf{x}) = p_{\alpha_1}(1)\, \mathcal{P}_{\alpha_1}^{\alpha_2}(2)\, \mathcal{P}_{\alpha_2}^{\alpha_3}(3) \ldots \mathcal{P}_{\alpha_{t-1}}^{\alpha_t}(t) \ldots \mathcal{P}_{\alpha_{s-}}^{\alpha_s}(s),$$

где $p_{\alpha_1}(1)$ — компоненты вектор-строки $\overline{p}_1 = (p_1(1), \ldots, p_\alpha(1), \ldots, p_a(1))$ абсолютных вероятностей $\mathcal{P}(A_{\alpha_1}^{(1)}) = p_{\alpha_1}(1)$ в момент 1, а $\mathcal{P}_{\alpha_{t-1}}^{\alpha_t}(t) = \mathcal{P}(A_{\alpha t}^{(t)}/A_{\alpha_{t-1}}^{(t-1)})$ элементы матрицы $\mathcal{P}_t = \| \mathcal{P}_{\alpha_{t-1}}^{\alpha_t}(t) \|$ условных вероятностей.

Производящая (2.9) здесь имеет вид

$$g(\mathbf{u}) = \sum_{\mathbf{x}} p(\mathbf{x})\, \mathbf{u}^{\mathbf{x}} = \sum_{\mathbf{x}} p_{\alpha_1}(1)\, u_{\alpha_1}^{(1)}\, \mathcal{P}_{\alpha_1}^{\alpha_2}(2)\, u_{\alpha_2}^{(2)} \ldots \mathcal{P}_{\alpha_{s-1}}^{\alpha_s}(s)\, u_{\alpha_s}^{(s)}. \qquad (2.13)$$

Введем обозначение для диагональной матрицы

$$D(\overline{w}) = \begin{pmatrix} w_1 & & 0 \\ & w_2 & \\ & & \ddots \\ 0 & & w_a \end{pmatrix}.$$

Тогда покомпонентное произведение двух вектор-строк $\bar{z}$ и $\bar{w}$ имеет вид

$$\bar{z}\cdot\bar{w} = \bar{z}D(\bar{w}) = \bar{w}D(\bar{z}) = (z_1w_1, \ldots, z_1w_1)$$

и соотношение (2.13), как легко видеть, имеет следующую матричную запись

$$g(\mathbf{u}) = \bar{p}_1 D(\bar{u}_1)\prod_{t=2}^{s} \mathcal{P}_t D(\bar{u}_t)\,\bar{e}, \tag{2.14}$$

где $\bar{e}$ — единичный a-мерный вектор-столбец.

Подчеркнем, что во всем дальнейшем изложении вероятностные и частотные векторы $\bar{p}$, $\bar{q}$, $\bar{m}$, $\bar{m}^*$ ... рассматриваются как вектор-строки, а формальные векторы $\bar{u}$, $\bar{v}$, $\bar{e}$ ... рассматриваются как вектор-столбцы. $\mathbf{p}'$ означают транспонированную матрицу $\mathbf{p}$. Штрихи у векторов означают переход вектор-столбцов в вектор-строки, и наоборот.

2.3.2. *Однородный случай.* Рассмотрим частный случай марковского однородного аргумента, когда матрица условных вероятностей $\mathcal{P}_t = \mathcal{P} = \|\mathcal{P}_a^{a'}\|$ не зависит от момента времени. В этом случае вероятность $p(\mathbf{x})$ и производящая $g(\mathbf{u})$ имеют вид

$$p(\mathbf{x}) = p_{a_1}(1)\,\mathcal{P}_{a_1}^{a_2}\mathcal{P}_{a_2}^{a_3}\ldots\mathcal{P}_{a_{t-1}}^{a_t}\ldots\mathcal{P}_{a_{s-1}}^{a_s}$$

и

$$g(\mathbf{u}) = \bar{p}_1 D(\bar{u}_1)\prod_{t=2}^{s} \mathcal{P}D(\bar{u}_t)\,\bar{e}$$

соответственно.

Удобно в качестве вектора абсолютных вероятностей $\bar{p}_1$ взять собственный вектор матрицы $\mathcal{P}$ с собственным значением 1, т. е. вектор, удовлетворяющий соотношению $\bar{p}\mathcal{P} = \bar{p}$. Известно [25], что такой вектор при весьма общих ограничениях всегда существует и каков бы ни был вектор $\bar{p}_1$, вектор абсолютных вероятностей элементарных значений в t-й момент

$$\bar{p}_t = \bar{p}_1\mathcal{P}^{t-1} \xrightarrow[t\to\infty]{} \bar{p},$$

где $\bar{p}$ совпадает со строкой матрицы $\bar{e}\bar{p} = \lim\limits_{t\to\infty}\mathcal{P}^t$, состоящей из постоянных столбцов.

В рассматриваемом случае вектор $\bar{p}_t = \bar{p}\mathcal{P}^{t-1} = \bar{p}$ не зависит от момента времени и марковский однородный аргумент называется устойчивым. Производящая для него имеет вид

$$g(\mathbf{u}) = \bar{p}D(\bar{u}_1)\prod_{t=2}^{s}\mathcal{P}D(\bar{u}_t)\,\bar{e} = \bar{p}\prod_{t=1}^{s}\mathcal{P}D(\bar{u}_t)\,\bar{e}. \tag{2.15}$$

Рассмотрим для однородного устойчивого марковского аргумента случай тождественных $A^{(t)} = A$. Тогда имеет смысл рассмотрение вероятностей $P_s(\bar{m})$ частот $\bar{m}$. В этом случае производящая частот, согласно (2.11) и (2.15) имеет

$$g(\bar{u}) = \sum_{\bar{m}} P_s(\bar{m}) = \bar{p}\cdot[\mathcal{P}D(\bar{u})]^s\cdot\bar{e}. \tag{2.16}$$

Выражение, соответствующее соотношению (2.16) для характеристической функции, приводится в [25].

2.3.3. *Слабо зависимый случай.* Рассмотрим однородный устойчивый марковский аргумент с тождественными $A^{(t)} = A$, определяемый единственной стохастической матрицей $\mathcal{P}$. На основе ее собственного вектора $\bar{p}$ абсолютных вероятностей может быть построена матрица $\bar{\bar{e}}\bar{\bar{p}}$ с тождественными строками, совпадающими с $\bar{p}$ (с постоянными столбцами). Матрица $\bar{\bar{e}}\bar{\bar{p}}$ имеет тот же собственный вектор $\bar{p}$ с собственным значением 1, что и матрица $\mathcal{P}$. Эта матрица играет особую роль. В самом деле, легко показать, что $\bar{\bar{e}}\bar{\bar{p}}\mathcal{P} = \mathcal{P}\bar{\bar{e}}\bar{\bar{p}} = \bar{\bar{e}}\bar{\bar{p}}$, для любого целого k

$$(\bar{\bar{e}}\bar{\bar{p}})^k = \bar{\bar{e}}\bar{\bar{p}} \quad \text{и} \quad \bar{\bar{e}}\bar{\bar{p}}\,(\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}}) = O,$$

где матрица O вся состоит из нулевых элементов. Поэтому

$$\mathcal{P}^k = [\bar{\bar{e}}\bar{\bar{p}} + (\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}})]^k = (\bar{\bar{e}}\bar{\bar{p}})^k + (\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}})^k = \bar{\bar{e}}\bar{\bar{p}} + (\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}})^k.$$

В невырожденных случаях, когда строки матрицы $\mathcal{P}$ не состоят из нулей и единицы,

$$(\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}})^k \xrightarrow[k\to\infty]{} O, \quad \text{откуда} \quad \mathcal{P}^k \xrightarrow[k\to\infty]{} \bar{\bar{e}}\bar{\bar{p}}.$$

Таким образом, условные вероятности элементарных значений, удаленных от фиксированных элементарных значений на k моментов времени, при $k \to \infty$ стремятся к своим абсолютным вероятностям, т. е. не зависят от фиксированных элементарных значений.

Более того, по близости (поэлементной) матриц $\bar{\bar{e}}\bar{\bar{p}}$ и $\mathcal{P}$ можно судить о степени зависимости элементарных значений в соседние моменты времени, так как при $\bar{\bar{e}}\bar{\bar{p}} = \mathcal{P}$ они не зависят друг от друга, и мы имеем в этом случае бернуллиевский стохастический аргумент с абсолютными вероятностями событий $\bar{p}$.

В связи с изложенным естественно следующее преобразование производящей (2.15)

$$g(\mathbf{u}) = \bar{p}\prod_{t=1}^{s}\mathcal{P}D(\bar{u}_t)\,\bar{e} = \bar{p}\prod_{t=1}^{t}[\bar{\bar{e}}\bar{\bar{p}} + (\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}})]\,D(\bar{u}_t)\,\bar{e} =$$

$$= \bar{p}\left[\prod_{t=1}^{s}\bar{\bar{e}}\bar{\bar{p}}D(\bar{u}_t) + (\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}})\,D(\bar{u}_1)\prod_{t=2}^{s}\bar{\bar{e}}\bar{\bar{p}}D(\bar{u}_t) + \right.$$

$$\left. + \sum_{t=2}^{s}\prod_{r=1}^{t-1}\bar{\bar{e}}\bar{\bar{p}}D(\bar{u}_r)\,(\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}})\,D(\bar{u}_t)\prod_{r=t+1}^{s}\bar{\bar{e}}\bar{\bar{p}}D(\bar{u}_r) + R\right]\bar{e}, \qquad (2.17)$$

где в случаях, когда верхний предел произведения превосходит нижний, соответствующая матрица полагается равной $E = D(\bar{e})$. Зметим, что матрица R состоит из суммы произведений матриц, среди сомножителей которых матрица $\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}}$ встречается не менее двух раз. Предположим теперь, что элементы матрицы $\mathcal{P} - \bar{\bar{e}}\bar{\bar{p}}$ по абсолютной величине не превосходят некоторого малого числа θ. Тогда легко показать, что элементы матрицы R по абсолютной величине не будут превосходить θ^2, если рассматривать компоненты формального аргумента $\mathbf{u}$, не превосходящие по абсолютной величине единицы. Этот факт мы будем записывать так: $R = O(\theta^2)\,\bar{\bar{e}}\bar{\bar{e}}'$. Далее имеем

$$\bar{p}D(\bar{u}) = \bar{p}\cdot\bar{u} \quad \text{и} \quad \bar{\bar{e}}\bar{\bar{p}}D(\bar{u}) = \bar{\bar{e}}\bar{\bar{p}}\cdot\bar{u},$$

поэтому

$$\overline{\overline{ep}D(\overline{u})\overline{e}} = \overline{\overline{epu}} \quad \text{и} \quad \prod_{t=1}^{s} \overline{\overline{ep}D(\overline{u}_t)\overline{e}} = \overline{e}\prod_{t=1}^{s} \overline{p}\,\overline{u}_t,$$

откуда

$$\overline{p}\prod_{t=1}^{s} \overline{\overline{ep}D(\overline{u}_t)\overline{e}} = \prod_{t=1}^{s} \overline{\overline{pu}_t}.$$

Кроме того, $\overline{p}(\mathcal{P} - \overline{\overline{ep}}) = \overline{0} = (\overbrace{0, \ldots, 0}^{a})$ и $\overline{\overline{pee'e}} = a$.

Поэтому из (2.17) имеем

$$g(\mathbf{u}) = \prod_{t=1}^{s} \overline{\overline{pu}_t} + \overline{p}\sum_{t=2}^{s}\prod_{r=1}^{t-1} \overline{\overline{ep}D(\overline{u}_r)}\,(\mathcal{P} - \overline{\overline{ep}})\,D(\overline{u}_t)\prod_{r=t+1}^{s} \overline{\overline{ep}D(\overline{u}_r)\overline{e}} +$$

$$+ O(\theta^2)\,\overline{\overline{pee'e}} = \prod_{t=1}^{s} \overline{\overline{pu}_t} + \overline{p}\sum_{t=2}^{s}\prod_{r=1}^{t-2} \overline{\overline{ep}D(u_r)}\,\overline{\overline{ep}D(\overline{u}_{t-1})}\,(\mathcal{P} - \overline{\overline{ep}}) \times$$

$$\times D(\overline{u}_t)\overline{e}\prod_{r=t+1}^{s} \overline{\overline{pu}_r} + aO(\theta^2) = \prod_{t=1}^{s} \overline{\overline{pu}_t} + \overline{\overline{pe}}\sum_{t=2}^{s}\prod_{r=1}^{t-2} (\overline{\overline{pu}_r})\,\overline{pD}(\overline{u}_{t-1})\,(\mathcal{P} - \overline{\overline{ep}}) \times$$

$$\times D(\overline{u}_t)\overline{e}\prod_{r=t+1}^{s} \overline{\overline{pu}_r} + O(\theta^2) = \prod_{t=1}^{s} \overline{\overline{pu}_t} + \prod_{r=1}^{s} \overline{\overline{pu}_r}\sum_{t=2}^{s} \frac{\overline{p}\,D(\overline{u}_{t-1})\,(\mathcal{P} - \overline{\overline{ep}})\,D(\overline{u}_t)\,\overline{e}}{\overline{\overline{pu}}_{t-1}\qquad \overline{\overline{pu}}_t} +$$

$$+ O(\theta^2) = \overline{\overline{pu}}_1\prod_{t=2}^{s}\left[\overline{\overline{pu}_t} + \frac{\overline{u}'_{t-1}D(\overline{p})\,(\mathcal{P} - \overline{\overline{ep}})\,\overline{u}_t}{\overline{\overline{pu}}_{t-1}}\right] + O(\theta^2).$$

Итак, имеем окончательно

$$g(\mathbf{u}) = \overline{\overline{pu}}_1\prod_{t=2}^{s}\left[\overline{\overline{pu}_t} + \frac{\overline{u}'_{t-1}D(\overline{p})\,(\mathcal{P} - \overline{\overline{ep}})\,\overline{u}_t}{\overline{\overline{pu}}_{t-1}}\right] + O(\theta^2).$$

Полагая $\overline{u}_t \equiv \overline{u}$, из (2.11) получим производящую частот

$$g(\overline{u}) = \overline{\overline{pu}}\left[\overline{\overline{pu}} + \frac{\overline{u}'D(\overline{p})\,(\mathcal{P} - \overline{\overline{ep}})\,\overline{u}}{\overline{\overline{pu}}}\right]^{s-1} + O(\theta^2) =$$

$$= (\overline{\overline{pu}})^s + (s-1)(\overline{\overline{pu}})^{s-1}\overline{u}'D(\overline{p})\,(\mathcal{P} - \overline{\overline{ep}})\,\overline{u} + O(\theta^2) =$$

$$= \left[(\overline{\overline{pu}}) + \frac{s-1}{s}\frac{\overline{u}'D(\overline{p})\,(\mathcal{P} - \overline{\overline{ep}})\,\overline{u}}{(\overline{\overline{pu}})}\right]^s + O(\theta^2).$$

Имеем окончательно

$$g(\overline{u}) = \left[\overline{\overline{pu}} + \frac{s-1}{s}\frac{\overline{u}'D(\overline{p})\,(\mathcal{P} - \overline{\overline{ep}})\,\overline{u}}{\overline{\overline{pu}}}\right]^s + O(\theta^2). \qquad (2.17')$$

До сих пор мы ограничивались рассмотрениями производящих функций распределений стохастического аргумента, так как явные выражения для самих распределений трудно найти. Однако в дальнейшем изложении свойства распределений могут быть изучены с помощью производящих.

Перейдем теперь к простейшему случаю стохастического аргумента.

2.3.4. *Независимый случай.* В независимом неоднородном случае (схема Пуассона) матрицы переходов $\mathcal{P}_t$ выражаются в матрицы с постоянными столбцами $\mathcal{P}_t = \overline{\overline{e}}\,\overline{p}_t$, где вектор $\overline{p}_t = (p_1^{(t)}, \ldots, p_a^{(t)}, \ldots, p_a^{(t)})$ состоит из абсолютных вероятностей, различных для разных моментов времени t. В этом случае производящая получится из общей производящей (2.14) заменой $\mathcal{P}_t = \overline{\overline{e}}\,\overline{p}_t$, откуда

$$g(\mathbf{u}) = \prod_{t=1}^{s} \overline{p}_t\,\overline{u}_t \tag{2.18}$$

и производящая частот имеет вид

$$g(\overline{u}) = \prod_{t=1}^{s} \overline{p}_t\,\overline{u}.$$

Наконец, для однородного случая (схема Бернулли), когда $\overline{p}_t = \overline{p}$, имеем производящую значений

$$g(\mathbf{u}) = \prod_{t=1}^{s} \overline{p}\overline{u}_t$$

и производящую частот при $\overline{u}_t = \overline{u}$

$$g(\overline{u}) = \prod_{t=1}^{s} \overline{p}\overline{u} = (\overline{p}\overline{u})^s = \left(\sum_{a=1}^{a} p_a u_a\right)^s. \tag{2.18'}$$

Отсюда распределение частот для бернуллиевского СА является полиномиальным

$$P_s(\overline{m}) = C_s^{\overline{m}}\,\overline{p}^{\overline{m}}. \tag{2.19}$$

Дифференцирование производящей (2.18') дает

$$\frac{\partial g(\overline{u})}{\partial u_a}\bigg|_{\overline{u}=\overline{e}} = \mathrm{E}m_a = sp_a, \quad \frac{\partial^2 g(\overline{u})}{\partial u_a\,\partial u_{a'}}\bigg|_{\overline{u}=\overline{e}} = \mathrm{E}m_a m_{a'} =$$

$$= s(s-1)\,p_a p_{a'} \quad (\alpha \neq \alpha');$$

$$\frac{\partial^2 g(\overline{u})}{\partial u_a^2}\bigg|_{\overline{u}=\overline{e}} = \mathrm{E}m_a(m_a - 1) = s(s-1)\,p_a^2.$$

Откуда имеем окончательно

$$\mathrm{E}m_a = sp_a,$$

$$\mathrm{E}(m_a - \mathrm{E}m_a)(m_{a'} - \mathrm{E}m_{a'}) = \begin{cases} -sp_a p_{a'}, & \text{при } \alpha \neq \alpha'; \\ sp_a(1-p_a), & \text{при } \alpha = \alpha'. \end{cases} \tag{2.20}$$

§ 2.4. Симметричный аргумент

Для марковского аргумента зависимость между элементарными значениями носит локальный характер. Рассмотрим схему стохастического аргумента, элементарные значения которого зависимы между собой во всех моментах времени $\sigma = (1, 2, \ldots, s)$.

Именно, будем называть СА симметричным аргументом, если все его значения $x \in \mathcal{E}_s^{\overline{m}}$ имеют одну и ту же вероятность $p(x) = p(\overline{m})$, зависящую лишь от частот $\overline{m}$.

Если ввести в рассмотрение распределение частот $P_s(\overline{m})$, то легко видеть, что

$$p(x) = P_s(\overline{m}) / C_s^{\overline{m}}.$$

Случай симметричного аргумента впервые рассматривался Финетти [28] под названием симметричных событий. Частным случаем рассмотренной схемы являются урновые схемы без возвращения (см. пп. 1.4.1 и 2.6.3).

Ясно, что для симметричного аргумента распределение частот имеет вид

$$P_s(\overline{m}) = C_s^{\overline{m}} p(\overline{m}).$$

Поэтому, назначая соответствующие вероятности

$$p(\overline{m}) = \frac{1}{C_s^{\overline{m}}} \sum_{x \in \mathcal{E}_s^m} \tilde{p}(x),$$

где $\tilde{p}(x)$ — вероятности значений произвольного СА, можно получить симметричный аргумент, имеющий распределение частот, совпадающее с распределением частот произвольного СА [28].

Далее, можно показать, что функция $p(\overline{m})$ векторного аргумента $\overline{m}$ всегда представима в виде

$$p(\overline{m}) = \sum_{k=1}^{N_{a,s}} a_k \overline{p}_k^{\overline{m}}, \tag{2.21}$$

где $\overline{p}_k = (p_{\alpha k})$ — некоторые специальным образом подобранные a-мерные вероятностные вектора $(\overline{p}_k \overline{e} = 1)$, а коэффициенты $a_k (k = 1, \ldots, N_{a,s} = C_{a+s-1}^{a-1})$ однозначно определяются по $p(\overline{m})$ и $\overline{p}_k$.

В самом деле, рассмотрим первые простые числа $p_1, p_2, \ldots, p_\alpha, \ldots, p_a$ и положим

$$p_{\alpha k} = \left(p_\alpha \Big/ \sqrt[k]{\sum_{\alpha=1}^{a} p_\alpha^k} \right)^k; \tag{2.22}$$

ясно, что $\sum_{\alpha=1}^{a} p_{\alpha k} = 1$. Подставим (2.22) в (2.21). Будем иметь

$$p(\overline{m}) = \sum_{k=1}^{N_{a,s}} a_k \prod_{\alpha=1}^{a} \left(p_\alpha \Big/ \sqrt[k]{\sum_{\alpha=1}^{a} p_\alpha^k} \right)^{k m_\alpha} = \sum_{k=1}^{N_{a,s}} \frac{a_k}{\left(\sum_{\alpha=1}^{a} p_\alpha^k \right)^s} (\overline{p}^{\overline{m}})^k.$$

Но из теории чисел известно, что $\overline{p}^{\overline{m}_1} \neq \overline{p}^{\overline{m}_2}$, если $\overline{m}_1 \neq \overline{m}_2$. Поэтому все N_{as} значений $\overline{p}^{\overline{m}}$ различны. Но тогда, рассматривая $\dfrac{a_k}{\left(\sum_{\alpha=1}^{a} p_\alpha^k \right)^s} = b_k$ за $N_{a,s}$ неизвестных, получим для их определения $N_{a,s}$ линейных уравнений с детерминантом Вандермонда $W(\overline{p}^{\overline{m}})$ в качестве детерминанта системы, отличным

от нуля из-за того, что $\overline{p}^{\overline{m}_1} \neq \overline{p}^{\overline{m}_2}$. Поэтому a_k однозначно определяются, и окончательно имеем для произвольного распределения частот

$$P_s(\overline{m}) = \sum_{k=1}^{N_{a,s}} a_k C_s^{\overline{m}} \overline{p}_k^{\overline{m}}.$$

§ 2.5. Предельные распределения стохастического аргумента

2.5.1. *Предельное распределение значений СА.* До недавнего времени рассмотрения СА велись на уровне частот и их предельных распределений в марковском случае.

Шенноном [2] было открыто замечательное свойство широкого класса СА на уровне значений. Именно, оказалось, что в асимптотическом случае значения аргумента разбиваются на два класса: небольшой класс равновероятных высоковероятных значений и большой класс равновероятных маловероятных значений.

Покажем, что этот факт имеет место для независимого аргумента. Для этого рассмотрим разбиение совокупности значений аргумента на $N_{a,s} = C_{a+s-1}^{a-1}$ непересекающихся множеств с постоянными частотами

$$R = \bigcup_{\overline{m}} \mathscr{E}_\sigma^{\overline{m}}$$

и выделим совокупность

$$\mathscr{E}_\sigma^{\mathfrak{M}_1(\varepsilon)} = \bigcup_{\overline{m} \in \mathfrak{M}_1(\varepsilon)} \mathscr{E}_\sigma^{\overline{m}}$$

всех $\mathscr{E}_\sigma^{\overline{m}}$, частота $\overline{m}$ которых мало отклоняется от своего среднего $|\overline{\mu} - \overline{p}| \leqslant \varepsilon$, где разность векторов понимается в покомпонентном смысле. Тогда из соотношений (1.62) и (1.63) имеем при $s \to \infty$

$$\mathscr{P}\left(\mathscr{E}_\sigma^{\mathfrak{M}_1(\varepsilon)}\right) \approx 1 - e^{-s[h(\overline{\mu}^{(0)}, \overline{p}) - h(\overline{\mu}^{(0)})]},$$

где

$$h(\overline{\mu}^{(0)}, \overline{p}) - h(\overline{\mu}^{(0)}) = \max_{\mu \in \mathfrak{M}_1(\varepsilon)} (h(\overline{\mu}, \overline{p}) - h(\mu)).$$

Очевидно и следующее соотношение

$$N\left(\mathscr{E}_\sigma^{\mathfrak{M}_1(\varepsilon)}\right) = \sum_{\overline{m} \in \mathfrak{M}_1(\varepsilon)} C_s^{\overline{m}} \approx e^{sh(\overline{\mu}^{(1)})}, \tag{2.23}$$

где

$$h(\overline{\mu}^{(1)}) = \max_{\overline{\mu} \in \mathfrak{M}_1(\varepsilon)} h(\overline{\mu}).$$

Ясно, что при $\varepsilon \to 0$ $\overline{\mu}^{(1)} \to \overline{p}$,

$$N\left(\mathscr{E}_\sigma^{\mathfrak{M}_1(\varepsilon)}\right) \to e^{sh(\overline{p})} \tag{2.24}$$

$$\mathscr{P}\left(\mathscr{E}_\sigma^{\mathfrak{M}_1(\varepsilon)}\right) \approx 1 - e^{-s\varepsilon^2 \sum_{\alpha=1}^{a} \frac{1}{p_\alpha} + sO(\varepsilon^3)}. \tag{2.25}$$

Таким образом, из (2.24) и (2.25) заключаем, что все значения аргумента $x \in \mathscr{E}_\sigma^{\mathfrak{M}_1(\varepsilon)}$ составляют высоковероятную часть множества всех значений R и при этом составляют лишь малую долю R, так как $N(R) = a^s = e^{s \ln a} \gg \gg N\left(\mathscr{E}_\sigma^{\mathfrak{M}_1(\varepsilon)}\right) \approx e^{sh(p)}$.

Будучи элементами множеств $\mathscr{E}_\sigma^{\overline{m}}$, близких к множеству $\mathscr{E}_\sigma^{[\overline{sp}]}$, все они при $s \to \infty$ делаются равновероятными и имеют вероятность

$$p(x) \approx e^{-sh(\overline{p})}. \tag{2.26}$$

В теории информации [2] показывается, что аналогичным свойством обладает широкий класс аргументов, включающий рассмотренные марковский и равномернорасположенный аргументы.

2.5.2. *Предельные распределения частот СА.* Наиболее изучены условия нормального приближения совместного распределения частот в марковском случае при $s \to \infty$ [25]. Отсюда следует нормальность распределения для независимого неоднородного и однородного аргумента. В последнем случае частоты $\overline{m} = (m_1, \ldots, m_\alpha, \ldots, m_a)$ имеют совместное a-мерное нормальное распределение [19] со средним $\mathrm{E}\overline{m} = (sp_1, \ldots, sp_\alpha, \ldots, sp_a)$ и матрицей вторых центральных моментов

$$\mathrm{E}(m_\alpha - sp_\alpha)(m_{\alpha'} - sp_{\alpha'}) = \begin{cases} sp_\alpha(1 - p_\alpha), & \text{при } \alpha = \alpha', \\ -sp_\alpha p_{\alpha'}, & \text{при } \alpha \neq \alpha'. \end{cases}$$

Если в неоднородном случае

$$p_\alpha = \begin{cases} 1 - \lambda/s \left(\lambda = \sum_{\alpha=2}^{a} \lambda_\alpha\right), & \text{при } \alpha = 1, \\ \lambda_\alpha / s, & \text{при } \alpha \neq 1, \end{cases}$$

то так же как и в п. 1.6.3, можно показать, что

$$P_s(\overline{m}) \xrightarrow[s \to \infty]{} \prod_{\alpha=2}^{a} \frac{\lambda_\alpha^{m_\alpha}}{m_\alpha!} e^{-\lambda_\alpha}. \tag{2.27}$$

§ 2.6. Бинарный случай стохастического аргумента

2.6.1. *Вводные замечания.* Если в момент t совокупность возможных элементарных значений $A^{(t)}$ состоит из двух взаимноисключающих значений $A^{(t)} = (A_t, \overline{A}_t) = (A_1^{(t)}, A_2^{(t)})$, то мы будем говорить о бинарном СА. Этот простейший по структуре множества $A^{(t)}$ случай важен тем, что для него общие соотношения $a > 2$ существенно упрощаются, а также могут быть получены новые соотношения, получение которых для $a > 2$ затруднительно. Следует подчеркнуть, что соответствующей группировкой событий $A^{(t)} = \{A_1^{(t)}, \ldots, A_\alpha^{(t)}, \ldots, A_a^{(t)}\}$, например $A^{(t)} = \{A_\alpha^{(t)}, \overline{A}_\alpha^{(t)}\}$, где $\overline{A}_\alpha^{(t)} = \{A_1^{(t)}, \ldots, A_{\alpha-1}^{(t)} A_{\alpha+1}^{(t)}, \ldots, A_a^{(t)}\}$, можно вместо СА $\mathscr{P}(A)$ рассмотреть бинарный СА $\mathscr{P}(\widetilde{A})$, где $\widetilde{A} = \{A_\alpha^{(t)} \overline{A}_\alpha^{(t)}\}$ и $\mathscr{P}(\widetilde{A})$ может быть получено из $\mathscr{P}(A)$.

В ряде случаев (см. п. 3.3.1) потери от такого рода «угрубления» оказываются незначительными.

2.6.2. *Распределения частот в марковском случае.* Начнем с рассмотрения марковского СА. Рассмотрим лишь однородный устойчивый случай.

Имеем здесь матрицу переходов

$$\mathcal{P} = \begin{pmatrix} \mathcal{P}_1^1 \mathcal{P}_1^2 \\ \mathcal{P}_2^1 \mathcal{P}_2^2 \end{pmatrix}$$

и вектор абсолютных вероятностей $\bar{p} = (p, q)$ $(q = 1 - p)$, причем $\bar{p}\mathcal{P} = \bar{p}$. Всегда имеем

$$\mathcal{P}_1^1 + \mathcal{P}_1^2 = \mathcal{P}_2^1 + \mathcal{P}_2^2 = p + q = 1.$$

Введем обозначение А. А. Маркова [25]:

$$\mathcal{P}_1^1 - \mathcal{P}_2^1 = \mathcal{P}_2^2 - \mathcal{P}_1^2 = \delta.$$

Тогда из п. 2.3.3. имеем представление $\mathcal{P}$ [23]

$$\mathcal{P} = \bar{\bar{e}}\bar{p} + \delta Q, \tag{2.28}$$

где матрицы

$$\bar{\bar{e}}\bar{p} = \begin{pmatrix} pq \\ pq \end{pmatrix}; \quad Q = \begin{pmatrix} q & -q \\ -p & p \end{pmatrix}$$

таковы, что $(\bar{\bar{e}}\bar{p})^k = \bar{\bar{e}}\bar{p}$, $Q^k = Q$, $\bar{\bar{e}}\bar{p}Q = O$ и $\mathcal{P}^k = \bar{\bar{e}}\bar{p} + \delta^k Q$. При $|\delta| < 1$ и $k \to \infty \mathcal{P}^k \to$ $\to \bar{\bar{e}}\bar{p}$.

В рассматриваемом случае удается получить явное выражение для распределения $P_s(m)$ частот $\bar{m} = (m, s - m)$, которое здесь вырождается в одномерное распределение числа m элементарных значений A, имевших место в s моментах времени $\sigma = (1, \ldots, t, \ldots, s)$.

Пусть $0 < \mathcal{P}_1^1, \mathcal{P}_2^2 < 1$. Известно (см. напр. [19]) явное выражение для $P_s(m)$. Используем развитый комбинаторный аппарат для вывода этого распределения.

Теорема 2.1. Вероятность $P_s(m)$ частоты m события A в s моментах времени для устойчивого, однородного бинарного аргумента имеет вид

$$P_s(m) = \sum_{l=0}^{\min(m, s-m)} \mathcal{P}_s(m, l), \tag{2.29}$$

где

$$\mathcal{P}_s(m, l) = \left[\frac{p}{p_{11}} C_{m-1}^{l} C_{s-m-1}^{l-1} + \left(\frac{p}{p_{21}} + \frac{q}{p_{12}} \right) C_{m-1}^{l-1} C_{s-m-1}^{l-1} + \right.$$
$$\left. + \frac{q}{p_{22}} C_{m-1}^{l-1} C_{s-m-1}^{l} \right] \left(\frac{p_{12} p_{21}}{p_{11} p_{22}} \right)^l p_{11}^m p_{22}^{s-m}, \tag{2.30}$$

причем положено $C_{-1}^0 = C_{-1}^{-1} \equiv 1$ и $\mathcal{P} = \begin{pmatrix} \mathcal{P}_1^1 \mathcal{P}_1^2 \\ \mathcal{P}_2^1 \mathcal{P}_2^2 \end{pmatrix} = \begin{pmatrix} p_{11} p_{12} \\ p_{21} p_{22} \end{pmatrix}$.

Доказательство. Рассмотрим значение $x \in R$, состоящее в появлении $A = (A, \bar{A})$ в $\sigma = (1, \ldots, s)$ моментах, и выделим моменты τ и τ' идущих подряд элементарных значений A и $\bar{A}$ соответственно. Тогда каждое значение однозначно представимо в одной из следующих четырех форм:

$$x = \begin{cases} (\tau_1, \tau_1', \ldots, \tau_\lambda, \tau_\lambda', \ldots, \tau_l, \tau_l', \tau_{l+1}) & \text{(I)}; \\ (\tau_1, \tau_1', \ldots, \tau_\lambda, \tau_\lambda', \ldots, \tau_l, \tau_l') & \text{(II)}; \\ (\tau_1', \tau_1, \ldots, \tau_\lambda', \tau_\lambda, \ldots, \tau_l', \tau_l) & \text{(III)}; \\ (\tau_1', \tau_1, \ldots, \tau_\lambda', \tau_\lambda, \ldots, \tau_l', \tau_l, \tau_{l+1}') & \text{(IV)}. \end{cases} \tag{2.31}$$

Введем числа $N(\tau_\lambda) = r_\lambda$ и $N(\tau'_\lambda) = r'_\lambda$. Тогда при фиксированных l, m и s числа $\{r_\lambda\}$ и $\{r'_\lambda\}$ должны удовлетворять следующим соотношениям соответственно для каждого из четырех вариантов (2.31):

$$\begin{aligned}
&\sum_{x=1}^{l+1} r_\lambda = m, \sum_{\lambda=1}^{l} r'_\lambda = s-m && \text{(I)};\\
&\sum_{\lambda=1}^{l} r_\lambda = m, \sum_{\lambda=1}^{l} r'_\lambda = s-m && \text{(II), (III)};\\
&\sum_{\lambda=1}^{l} r_\lambda = m, \sum_{x=1}^{l+1} r'_\lambda = s-m && \text{(IV).}
\end{aligned} \tag{2.32}$$

Введем в рассмотрение числа $\overline{k} = \{k_i\}$ и $\overline{k'} = \{k'_j\}$, указывающие на количества среди $\{r_\lambda\}$ и $\{r'_\lambda\}$ чисел, в точности равных i и j соответственно.

Тогда соотношения (2.32) для чисел $\{k_i\}$ и $\{k_j\}$ перепишутся в виде:

$$\begin{aligned}
&\sum_{i=1}^{\infty} k_i = l+1, \sum_{i=1}^{\infty} ik_i = m; \sum_{j=1}^{\infty} k'_j = l \ldots, \sum_{j=1}^{\infty} jk'_j = s-m && \text{(I)};\\
&\sum_{i=1}^{\infty} k_i = l; \sum_{i=1}^{\infty} ik_i = m; \sum_{j=1}^{\infty} k'_j = l, \sum_{j=1}^{\infty} jk'_j = s-m && \text{(II), (III);}\\
&\sum_{i=1}^{\infty} k_i = l; \sum_{i=1}^{\infty} ik_i = m; \sum_{j=1}^{\infty} k'_j = l+1, \sum_{j=1}^{\infty} jk'_j = s-m && \text{(IV).}
\end{aligned} \tag{2.33}$$

Вычислим теперь вероятность $p(x)$ в каждом из четырех вариантов (2.31), используя числа $\{k_\lambda\}$ и $\{k'_\lambda\}$ и соотношения (2.33). Имеем для варианта (I)

$$p(x) = pp_{11}^{\sum_{i=1}^{\infty}(i-1)k_i - 1} \cdot p_{21}^{\sum_{i=1}^{\infty} k_i - 1} p_{22}^{\sum_{j=1}^{\infty}(j-1)k'_j} p_{12}^{\sum_{j=1}^{\infty} k'_j} = \frac{p}{p_{11}} \left(\frac{p_{21}}{p_{11}} \frac{p_{12}}{p_{22}}\right)^l p_{11}^m p_{22}^{s-m}.$$

Аналогично рассчитывая $p(x)$ для остальных вариантов, будем иметь:

$$p(x) = \begin{cases}
\dfrac{p}{p_{11}} \left(\dfrac{p_{21}p_{12}}{p_{11}p_{22}}\right)^l p_{11}^m p_{22}^{s-m} & \text{(I)};\\
\dfrac{p}{p_{21}} \left(\dfrac{p_{21}}{p_{11}} \dfrac{p_{12}}{p_{22}}\right)^l p_{11}^m p_{22}^{s-m} & \text{(II)};\\
\dfrac{q}{p_{12}} \left(\dfrac{p_{21}}{p_{11}} \dfrac{p_{12}}{p_{22}}\right)^l p_{11}^m p_{22}^{s-m} & \text{(III)};\\
\dfrac{q}{p_{22}} \left(\dfrac{p_{21}}{p_{11}} \dfrac{p_{12}}{p_{22}}\right)^l p_{11}^m p_{22}^{s-m} & \text{(IV).}
\end{cases} \tag{2.34}$$

Вероятности (2.34) будут иметь все значения x, для которых соответствующие числа $\overline{k}$ и $\overline{k'}$ удовлетворяют системам (2.32). Число таких значений соответственно

для каждого варианта (2.31) имеет вид:

$$\begin{aligned}
&\sum C_{l+1}^{\overline{k}} \cdot \sum C_{l}^{\overline{k}'} = C_{m-1}^{l} \cdot C_{s-m-1}^{l-1} && \text{(I);}\\
&\sum C_{l}^{\overline{k}} \cdot \sum C_{l}^{\overline{k}'} = C_{m-1}^{l-1} \cdot C_{s-m-1}^{l-1} && \text{(II), (III);}\\
&\sum C_{l}^{\overline{k}} \cdot \sum C_{l+1}^{\overline{k}'} = C_{m-1}^{l-1} \cdot C_{s-m-1}^{l} && \text{(IV),}
\end{aligned} \tag{2.35}$$

где суммирование ведется по $\overline{k}$ и $\overline{k}'$, удовлетворяющим соответствующим системам (2.33), и используются формулы суммирования (1.50).

Беря сумму вероятностей (2.34), умноженных на соответствующие коэффициенты (2.35), и суммируя полученное выражение по l, получим утверждение теоремы.

Для дальнейшего изложения необходимо следующее соотношение:

$$\sum_{s=2}^{\infty}\sum_{m=1}^{s-1} C_{m-1}^{a} C_{s-m-1}^{b} u^m v^s = \left(\frac{uv}{1-uv}\right)^{a+1}\left(\frac{v}{1-v}\right)^{b+1}. \tag{2.36}$$

Для доказательства соотношения (2.36) рассмотрим разложение

$$(1-uv)^{-(a+1)}(1+v)^{-(b+1)} = \sum_{\alpha=0}^{\infty} C_{a+\alpha}^{a} u^{\alpha} v^{\alpha} \sum_{\beta=0}^{\infty} C_{b+\beta}^{b} v^{\beta} = \sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty} C_{a+\alpha}^{a} C_{b+\beta}^{b} u^{\alpha} v^{\alpha+\beta}.$$

Проведем замену переменных суммирования, положив

$$\begin{aligned}
a+\alpha &= m-1,\\
b+\beta &= s-m-1,
\end{aligned}$$

откуда

$$\alpha = m-1-a,$$

$$\alpha+\beta = s-2-a-b;$$

тогда

$$(1-uv)^{-(a+1)}(1-v)^{-(b+1)} = \sum_{s=2}^{\infty}\sum_{m=1}^{s-1} C_{m-1}^{a} C_{s-m-1}^{b} u^{m-1-a} v^{s-2-a-b} =$$

$$= (uv)^{-(a+1)} v^{-(b+1)} \sum_{s=2}^{\infty}\sum_{m=1}^{s-1} C_{m-1}^{a} C_{s-m-1}^{b} u^m v^s.$$

Поделив обе части полученного соотношения на $(uv)^{-(a+1)}v^{-(b+1)}$, получим соотношение (2.36).

Имеет место

Следствие 2.1.1. Вторая производящая

$$g(u, v) = \sum_{s=0}^{\infty}\sum_{m=0}^{s} P_s(m)\, u^m v^s = \frac{1+[(p-p_{11})u+q-p_{22}]v}{1-(p_{11}u+p_{22})v-(p_{21}-p_{11})uv^2}.$$

Доказательство. Имеем $g(u, v) = \sum_{s=0}^{\infty} g_s(u) v^s$, где $g_s(u) = \sum_{m=0}^{s} P_s(m) u^m$. Но

$P_0(0) = 1$, $P_1(0) = q$ и $P_1(1) = p$, откуда $g_0(u) = P_0(0) u^0 = 1 \cdot 1 = 1$ и $g_1(u) = P_1(0) + P_1(1) u = q + pu$.

Поэтому

$$g(u, v) = 1 + (pu + q)v + \sum_{s=2}^{\infty} \sum_{m=0} P_s(m) u^m v^s = 1 + (pu + q)v +$$

$$+ \sum_{s=2}^{\infty} \left[\frac{q}{p_{22}} (p_{22}v)^s + \frac{p}{p_{11}} (p_{11}uv)^s \right] + \sum_{s=2}^{\infty} \sum_{m=1}^{s-1} P_s(m) u^m v^s = 1 + (pu + q)v +$$

$$+ \frac{q}{p_{22}} \frac{p_{22}^2 v^2}{1 - p_{22}v} + \frac{p}{p_{11}} \frac{p_{11}^2 u^2 v^2}{1 - p_{11}uv} + \sum_{s=2}^{\infty} \sum_{m=1}^{s-1} P_s(m) u^m v^s.$$

Преобразуем первые слагаемые

$$S_1 = 1 + (pu + q)v + \frac{q}{p_{22}} \frac{p_{22}^2 v^2}{1 - p_{22}v} + \frac{p}{p_{11}} \frac{p_{11}^2 u^2 v^2}{1 - p_{11}uv} =$$

$$= 1 + (pu + q)v + \frac{q}{p_{22}} \left(\frac{p_{22}v}{1 - p_{22}v} - p_{22}v \right) + \frac{p}{p_{11}} \left(\frac{p_{11}uv}{1 - p_{11}uv} - p_{11}uv \right) =$$

$$= 1 + (pu + q)v - (pu + q)v + p \frac{uv}{1 - p_{11}uv} + q \frac{v}{1 - p_{22}v}.$$

Итак

$$S_1 = 1 + p \frac{uv}{1 - p_{11}uv} + q \frac{v}{1 - p_{22}v}. \tag{2.37}$$

Преобразуем двойную сумму, используя представление (2.29) для $P_s(m)$

$$S_2 = \sum_{s=2}^{\infty} \sum_{m=1}^{s-1} P_s(m) u^m v^s = \sum_{s=2}^{\infty} \sum_{m=1}^{s-1} \sum_{l=0}^{\infty} \mathscr{P}_s(m, l) u^m v^s = \sum_{l=1}^{\infty} \sum_{s=2}^{\infty} \sum_{m=1}^{s-1} \mathscr{P}_s(m, l) u^m v^s.$$

Используя выражение (2.30) для $\mathscr{P}_s(m, l)$ и формулу суммирования (2.36), будем иметь

$$S_2 = \sum_{l=1}^{\infty} \left[\frac{p}{p_{11}} \left(\frac{p_{11}uv}{1 - p_{11}uv} \right)^{l+1} \left(\frac{p_{22}v}{1 - p_{22}v} \right)^l + \left(\frac{p}{p_{21}} + \frac{q}{p_{12}} \right) \left(\frac{p_{11}uv}{1 - p_{11}uv} \right)^l \left(\frac{p_{22}v}{1 - p_{11}v} \right)^l + \right.$$

$$\left. + \frac{q}{p_{22}} \left(\frac{p_{11}uv}{1 - p_{11}uv} \right)^l \left(\frac{p_{22}v}{1 - p_{22}v} \right)^{l+1} \right] \left(\frac{p_{21}p_{12}}{p_{11}p_{22}} \right)^l = \left(p \frac{uv}{1 - p_{11}uv} + \frac{p}{p_{21}} + \frac{q}{p_{12}} + q \frac{v}{1 - p_{22}v} \right) \times$$

$$\times \sum_{l=1}^{\infty} \left[\frac{p_{21}p_{12}uv^2}{(1 - p_{11}uv)(1 - p_{22}v)} \right]^l = \left(p \frac{uv}{1 - p_{11}uv} + \frac{p}{p_{21}} + \frac{q}{p_{12}} + q \frac{v}{1 - p_{22}v} \right) \times$$

$$\times \frac{p_{21}p_{12}uv^2}{(1 - p_{11}uv)(1 - p_{22}v)} \cdot \frac{1}{1 - \dfrac{p_{21}p_{12}uv^2}{(1 - p_{11}uv)(1 - p_{22}v)}} =$$

$$= \left(p \frac{uv}{1 - p_{11}uv} + \frac{p}{p_{21}} + \frac{q}{p_{12}} + q \frac{v}{1 - p_{22}v} \right) \frac{p_{21}p_{12}uv^2}{(1 - p_{11}uv)(1 - p_{22}v) + p_{21}p_{12}uv^2}.$$

Но $(1 - p_{11}uv)(1 - p_{22}v) - p_{21}p_{12}uv^2 = 1 - (p_{11}u + p_{22})v + (p_{11}p_{22} - p_{21}p_{12})uv^2 = 1 - (p_{11}u + p_{22})v - (p_{21} - p_{11})uv^2$, поэтому окончательно

$$S_2 = \left(p \frac{uv}{1 - p_{11}uv} + \frac{p}{p_{21}} + \frac{q}{p_{12}} + q \frac{v}{1 - p_{12}v} \right) \frac{p_{21}p_{12}uv^2}{1 - (p_{11}u + p_{12})v - (p_{21} - p_{11})uv^2}. \tag{2.38}$$

Далее, используя соотношение (2.37) и (2.38), имеем

$$g(u, v) = S_1 + S_2 = 1 + p\frac{uv}{1 - p_{11}uv} + q\frac{v}{1 - p_{22}v} +$$

$$+ \frac{\left(p\frac{uv}{1 - p_{11}uv} + q\frac{v}{1 - p_{22}v}\right) p_{21}p_{12}uv^2 + (pp_{12} + qp_{21})\, uv^2}{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{12})\, uv^2} =$$

$$= \frac{1 - (p_{11} + p_{22})\, v - (p_{21} - p_{12})\, uv^2 + \left(p\frac{uv}{1 - p_{11}uv} + q\frac{v}{1 - p_{22}v}\right)(1 - p_{11}uv)(1 - p_{22}v)}{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2} +$$

$$+ \frac{(pp_{12} + qp_{21})\, uv^2}{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2} =$$

$$= \frac{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2 + puv(1 - p_{22}v) + qv(1 - p_{11}uv) + (pp_{12} + pq_{21})\, uv^2}{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2} =$$

$$= \frac{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2 + (pu + q)\, v - (pp_{22} + qp_{11})\, uv^2 + (pp_{22} + qp_{21})\, uv^2}{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2} =$$

$$= \frac{1 + [(p - p_{11})\, u + q - p_{22}]\, v + [p(p_{12} - p_{22}) + q(p_{21} - p_{11}) - (p_{21} - p_{11})]\, uv^2}{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2} =$$

$$\frac{1 + [(p - p_{11})\, u + q - p_{22}]\, v + (p_{21} - p_{11})(p + q - 1)\, uv^2}{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2} = \frac{1 + [(p - p_{11})\, u + q - p_{22}]\, v}{1 - (p_{11}u + p_{22})\, v - (p_{21} - p_{11})\, uv^2},$$

что и требовалось доказать.

Заметим, что согласно (2.28)

$$\begin{pmatrix} p_{11}p_{12} \\ p_{21}p_{22} \end{pmatrix} = \begin{pmatrix} p + \delta q & q - \delta q \\ p - \delta p & q + \delta p \end{pmatrix}.$$

Используя эти соотношения, запишем $g(u, v)$ в форме

$$g(u, v) = \frac{1 - \delta(qu + p)\, v}{1 - [(p + \delta q)\, u + q + \delta p]\, v + \delta uv^2} = \sum_{s=1}^{\infty} g_s(u)\, v^s, \tag{2.39}$$

впервые полученной А. А. Марковым [25] с помощью конечно-разностных уравнений.

Используя разложение на простейшие дроби методом неопределенных коэффициентов, получим соотношение

$$\frac{1 - Av}{1 - Bv + Cv^2} = \left(\frac{1}{2} - \frac{2A - B}{4\sqrt{B^2/4 - C}}\right)\Big/[1 - (B/2 + \sqrt{B^2/4 - C})\, v] +$$

$$+ \left(\frac{1}{2} + \frac{2A - B}{4\sqrt{B^2/4 - C}}\right)\Big/[1 - (B/2 - \sqrt{B^2/4 - C})\, v]. \tag{2.40}$$

Тогда, положив $A = \delta(qu + p)$, $B = pu + q + \delta(qu + p)$ и $C = \delta u$ и разлагая правую часть (2.40) в ряд по v, получим из (2.39) выражение

$$g_s(u) = \left\{\frac{1}{2} - \frac{2\delta u - [pu + q + \delta(qu + p)]}{2\sqrt{[pu + q + \delta(qu + p)]^2 - 4\delta u}}\right\} \times \left\{\frac{1}{2}[pu + q + \delta(qu + p) + \right.$$

$$\left. + \sqrt{[pu + q + \delta(qu + p)]^2 - 4\delta u}\right\}^s +$$

$$+ \left\{\frac{1}{2} + \frac{2\delta u - [pu + q + \delta(qu + p)]}{2\sqrt{[pu + q + \delta(qu + p)]^2 - 4\delta u}}\right\} \times \left\{\frac{1}{2}[pu + q + \delta(qu + p) - \right.$$

$$\left. - \sqrt{[pu + q + \delta(qu + p)]^2 - 4\delta u}\right\}^s.$$

Легко проверить, что при $u=1$ $g_s(1)=1$ и при $\delta=0$ $g_s(u)=(pu+q)^s$. При фиксированной матрице $\mathscr{P}$ и $s\to\infty$ распределение $P_s(\overline{m})$ аппроксимируется нормальным распределением со средним sp и дисперсией $spq\dfrac{1+\delta}{1-\delta}$ [25].

Рассмотрим случай слабо зависимого СА. Имеем

$$g(u,v)=\frac{1-\delta(qu+p)v}{1-(pu+q)v-\delta(qu+p-uv)v}=\frac{1-\delta(qu+p)v}{1-(pu+q)v}\times$$
$$\times\frac{1}{1-\delta\dfrac{qu+p-uv}{1-(pu+q)v}v}=\frac{1-\delta(qu+p)v}{1-(pu+q)v}\times$$
$$\times\left[1+\delta\frac{qu+p-uv}{1-(pu+q)v}v+O(\delta^2)\right].$$

После несложных преобразований получим

$$g(u,v)=\frac{1-(pu+q)v+\delta pq(u-1)^2v^2}{[1-(pu+q)v]^2}+O(\delta^2)(u-1)\equiv$$
$$\equiv\frac{1}{1-(pv+q)v}+\delta pq(u-1)^2\frac{v^2}{[1-(pv+q)v]^2}+(u-1)O(\delta^2). \quad (2.41)$$

Из (2.41) следует

$$g(u,v)=\sum_{s=0}^{\infty}g_s(u)v^s=\sum_{s=0}^{\infty}(pu+q)^sv^s+$$
$$+\delta pq(u-1)^2v^2\sum_{k=0}^{\infty}C_{2-1+k}^{2-1}(pu+q)^kv^k+(u-1)O(\delta^2)=$$
$$=1+\sum_{s=1}^{\infty}[(pu+q)^s+\delta(s-1)pq(u-1)^2(pu+q)^{s-2}+(u-1)O(\delta^2)]v^s.$$

Откуда

$$g_s(u)=(pu+q)^s+\delta(s-1)pq(u-1)^2(pu+q)^{s-2}+(u-1)O(\delta^2)=$$
$$=\left[pu+q+\delta\frac{s-1}{s}pq\frac{(u-1)^2}{pu+q}\right]^s+(u-1)O(\delta^2), \quad (2.42)$$

что, как легко показать, является частным бинарным случаем соотношения (2.17′).

2.6.3. *Распределение частот в симметричном случае (урновые схемы).* Рассмотрим бинарный вариант (шары двух цветов) урновой схемы без возвращения и выборки с добавлением одного шара того же цвета, что и вынутый. Обе эти схемы являются частным случаем известной схемы Пойа [23] добавления или вынимания k шаров. Эта общая схема также приводит к модели симметричного СА.

Итак, пусть урна содержит s шаров, пронумерованных целыми числами от 1 до s, из которых m белых и $s-m$ черных. Номера шаров будем интерпретировать местами $\sigma=(1, 2, \ldots, t, \ldots, s)$, а их цвет предметами $A=(A, \bar{A})$, где A соответствует белому цвету, а $\bar{A}$ — черному. В бинарном случае вероятность появления t_1 белых шаров ($t-t_1$ черных) среди выбранных t

шаров без добавления имеет вид

$$\underline{P}_{t,\,s,\,m}(t_1) = \frac{C_t^{t_1} \cdot C_{s-t}^{m-t_1}}{C_s^m} = \frac{C_m^{t_1} \cdot C_{s-m}^{t-t_1}}{C_s^t}. \tag{2.43}$$

Аналогичные вероятности для выборки с возвращением имеют вид

$$P_{t,\,s,\,m}(t_1) = C_t^{t_1} \left(\frac{m}{s}\right)^{t_1} \left(\frac{s-m}{s}\right)^{t-t_1}$$

и для выборки с добавлением

$$\overline{P}_{t,\,s,\,m}(t_1) = \frac{C_{m+t_1-1}^{t_1} \cdot C_{s-m+t-t_1-1}^{t-t_1}}{C_{s+m-1}^{t}} = \frac{C_{m+t_1-1}^{m-1} \cdot C_{s-m+t-t_1-1}^{s-m-1}}{C_{s+m-1}^{s-1}}. \tag{2.44}$$

Заметим, что соответствующие вторые производящие с весами имеют вид

$$\underline{g}_{s,\,m}(u, v) = (1+uv)^m (1+v)^{s-m} = \sum_t C_s^t \underline{g}_{t,\,s,\,m}(u) v^t \tag{2.45}$$

и

$$\overline{g}_{s,\,m}(u, v) = (1-uv)^{-m} (1-v)^{-(s-m)} = \sum_t C_{s-1+m}^{s-1} \overline{g}_{t,\,s,\,m}(u) v^t, \tag{2.46}$$

где $\underline{g}_{t,\,s,\,m}(u) = \sum_{t_1} \underline{P}_{t,\,s,\,m}(t_1)\, u^{t_1}$ и $\overline{g}_{t,\,s,\,m}(u) = \sum_{t_1} \overline{P}_{t,\,s,\,m}(t_1)\, u^{t_1}$,

соответствующие первые производящие.

Беря первую и вторую частные производные по u от выражений (2.45) и (2.46), полагая $u=1$ и сравнивая коэффициенты при одинаковых степенях v, получим для соответствующих средних

$$\underline{\mathrm{E}}m = \overline{\mathrm{E}}m = \mathrm{E}m = t\frac{m}{s}$$

и для соответствующих вторых факториальных моментов

$$\underline{\mathrm{E}}m(m-1) = t(t-1)\frac{m(m-1)}{s(s-1)} \quad \text{и} \quad \overline{\mathrm{E}}m(m-1) = t(t+1)\frac{m(m+1)}{s(s+1)},$$

откуда дисперсии

$$\underline{\mathrm{D}}m = t\frac{m}{s}\frac{s-m}{s}\cdot\frac{s-t}{s-1} \quad \text{и} \quad \overline{\mathrm{D}}m = s\frac{m}{s}\frac{s-m}{s}\cdot\frac{s+t}{s+1},$$

что отличается от дисперсии

$$\mathrm{D}m = t\frac{m}{s}\frac{s-m}{s}$$

для выборки с возвращением, соответствующей независимому СА.

Из (2.43) и (2.44) ясно, что две рассматриваемые урновые схемы приближаются к схеме с возвращением, при $s \to \infty$, т. е. при больших s они моделируют слабо зависимый симметричный СА.

Заметим, что рассмотренные бинарные симметричные СА совпадают с марковским слабозависимым СА по параметрам $p(\mathrm{E}m)$ и $\mathrm{D}m$, если положить

$$p = \frac{m}{s} \text{ и } 2\delta \approx \begin{cases} -\dfrac{t-1}{s-1}, & \text{для выборки без возвращения при } \delta < 0, \\ \dfrac{t-1}{s+1}, & \text{для выборки с добавлением при } \delta > 0. \end{cases}$$

Таким образом, в слабо зависимом случае можно с достаточной точностью моделировать распределение частот в марковском случае симметричными урновыми схемами.

2.6.4. *Соотношение между вторыми двойственными производящими.* Рассмотрим случай соотношения (1.24) и связанных с ним множеств

$$\mathscr{E}_{\tau}^{(t_1,\, t-t_1)} \text{ и } \mathscr{E}_{(1,\ldots,\, t-1)}^{(t_1-1,\, t-t_1)} \times \mathscr{E}_{(t)}^{(1,\,0)},$$

когда $\tau = (1, 2, \ldots, t)$. Введем множество «хвостов» расположений

$$\mathscr{E}_{t}^{(t_1)} = \bigcup_{r_1=t_1+1}^{t} \mathscr{E}_{(1,\ldots,\, r)}^{(r_1,\, t-r_1)} \text{ и } \mathscr{E}_{t_1}^{(t)} = \bigcup_{r=t+1}^{\infty} \mathscr{E}_{(1,\ldots,\, r-1)}^{(t_1-1,\, r-t_1)} \times \mathscr{E}_{(r)}^{(1,\,0)}.$$

Тогда

$$\mathscr{E}_{t}^{(t_1)} = R_{\tau} - \mathscr{E}_{t_1}^{(t)}. \tag{2.47}$$

Соотношение (2.47) означает, что множество τ-значений, у которых в первых t моментов имеет место более t_1 элементарных значений A, совпадает с множеством τ-значений, у которых t_1 элементарных значений A имеет место не более, чем в t первых моментах [23].

Введем числа элементов

$$\begin{gathered} N(\mathscr{E}_{\tau}^{(r_1,\, t-r_1)}) = a_t(r_1),\ N(\mathscr{E}_{1,\ldots,\, r-1}^{(t_1-1,\, r-t_1)} \times \mathscr{E}_{(r)}^{(1,0)}) = b_{t_1}(r); \\ N(\mathscr{E}_{t}^{(t_1)}) = A_t(t_1) = \sum_{r_1=t_1+1}^{\infty} a_t(r_1),\ N(\mathscr{E}_{t_1}^{(t)}) = B_{t_1}(t) = \sum_{r=t+1}^{\infty} b_{t_1}(r) \end{gathered} \tag{2.48}$$

и соответствующие производящие функции

$$\begin{gathered} g_t(u) = \sum_{t_1=0}^{t} a_t(t_1),\ \widetilde{g}_{t_1}(v) = \sum_{t=0}^{\infty} b_{t_1}(t)\, v^t; \\ g(u, v) = \sum_{t=0}^{\infty} g_t(u)\, v^t,\ \widetilde{g}(u, v) = \sum_{t_1=0}^{\infty} \widetilde{g}_{t_1}(v)\, u^{t_1}; \\ G_t(u) = \sum_{t_1=0}^{\infty} A_t(t_1)\, u^{t_1},\ \widetilde{G}_{t_1}(v) = \sum_{t=0}^{\infty} B_{t_1}(t)\, v^t; \\ G(u, v) = \sum_{t=0}^{\infty} G_t(u)\, v^t,\ \widetilde{G}(u, v) = \sum_{t_1=0}^{\infty} G_{t_1}(v)\, u^{t_1}, \end{gathered} \tag{2.49}$$

где бесконечные ряды предполагаются сходящимися, а их члены полагаются тождественно равными нулю, если верхние индексы суммирования выходят за пределы определения соответствующих членов. Возьмем число элементов от обеих частей (2.47). Будем иметь с учетом (2.48)

$$A_t(t_1) = B_{t_1}(-1) - B_{t_1}(t). \tag{2.50}$$

Умножая обе части (2.50) на $u^{t_1}v^t$ и суммируя по всем t_1 и t с учетом (2.49), получим

$$G(1, v) + uG(u, v) = \frac{1}{1-v}\,\widetilde{G}(u, 1) - \widetilde{G}(u, v). \tag{2.51}$$

Если числа $a_t(t_1)$ и $b_{t_1}(t)$ имеют смысл вероятностей соответствующих τ-значений, то, как легко видеть, $A_t(-1) = B_{t_1}(-1) = 1$ и соотношение

(2.51) примет вид

$$\frac{1}{1-v} + vG(u, v) = \frac{1}{1-v} \cdot \frac{1}{1-u} - \widetilde{G}(u, v). \tag{2.52}$$

Перейдем теперь от соотношения между вторыми производящими хвостов (2.52) к соотношению между вторыми производящими самих вероятностей. Для этого воспользуемся легко проверяемыми соотношениями [23]

$$G_t(u) = \frac{1-g_t(u)}{1-u}, \quad \widetilde{G}_{t_1}(v) = \frac{1-\widetilde{g}_{t_1}(v)}{1-v},$$

откуда следуют соотношения

$$G(u, v) = \frac{\frac{1}{1-v} - g(u, v)}{1-u}, \quad \widetilde{G}(u, v) = \frac{\frac{1}{1-u} - \widetilde{g}(u, v)}{1-v}.$$

Подставляя полученные значения $G(u, v)$ и $\widetilde{G}(u, v)$ в соотношение (2.52), после несложных преобразований получим соотношение между вторыми двойственными производящими [7]

$$\widetilde{g}(u, v) = \frac{1-u(1-v)\,g(u, v)}{1-u}. \tag{2.53}$$

В качестве одного из возможных приложений соотношения (2.53) изучим с его помощью обобщенное распределение Паскаля для марковского бинарного СА. Для бернуллиевского СА распределение Паскаля было рассмотрено в п. 1.4.2.

Подставим в соотношение (2.53) выражение для второй производящей (2.39). После несложных преобразований получим

$$\widetilde{g}(u, v) = \frac{1-(q-\delta p)\,v}{1-(q+\delta p)\,v - [(p+\delta q) - \delta v]\,uv}.$$

Отсюда, разлагая $\widetilde{g}(u, v)$ в ряд по степеням u^{t_1}, получим производящую $\widetilde{g}_{t_1}(v)$ в качестве коэффициента при u^{t_1}

$$\widetilde{g}_{t_1}(v) = \left[\frac{(p+\delta q) - \delta v}{1-(q+\delta p)\,v}\,v\right]^{t_1}.$$

Если вместо случайного числа t первых моментов, в которых впервые встретится t_1 элементарных значений A, рассмотреть случайное число $t_2 = t - t_1$ элементарных значений $\bar{A}$ в этих моментах, то, как легко видеть, его производящая

$$\widetilde{g}^*_{t_1}(v) = \widetilde{g}_{t_1}(v)\,v^{-t_1} = \left[\frac{(p+\delta y) - \delta y}{1-(q+\delta p)\,v}\right]^{t_1} = \left[p_{22}\frac{p_{21}v}{1-p_{22}v} + p_{11}\right]^{t_1}.$$

Итак,

$$\widetilde{g}^*_{t_1}(v) = g_{t_1}(\widetilde{g}_1(v)), \tag{2.54}$$

где

$$g_{t_1}(u) = [p_{22}u + p_{11}]^{t_1} \text{ и } \widetilde{g}_1(v) = \frac{p_{21}v}{1-p_{22}v}.$$

Соотношение (2.54) эквивалентно следующей теореме, формулируемой в терминах случайных величин.

Теорема 2.2. [7]. Случайное число t_2 появлений $\bar{A}$ до появления A в в точности t_1 раз в случае простого устойчивого марковского СА, опреде-

ляемого матрицей переходов $\mathcal{P} = \begin{pmatrix} p_{11}p_{12} \\ p_{21}p_{12} \end{pmatrix}$, может быть представлено в виде суммы случайного числа ν^* независимых одинаково распределенных случайных величин μ_i^*

$$t_2 = \sum_{i=0}^{\nu^*} \mu_i^*. \tag{2.55}$$

При этом ν^* имеет биномиальное распределение $\mathcal{P}(\nu^* = n) = C_{t_1}^n p_{12}^n p_{11}^{t_1-n}$, а все μ_i^* имеют одно и то же геометрическое распределение $\mathcal{P}(\mu_i^* = r) = p_{22}^{r-1}p_{21}$, причем ν^* и μ_i^* независимы между собой.

Доказательство следует из соотношения (2.54) и известной теоремы о производящих для сложных распределений [23] (см. также п. 1.5.1).

Следствие 2.2.1. При фиксированном t_2 случайное число t_1 появлений A до появления $\bar{A}$ впервые в точности t_2 раз может быть представлено в виде

$$t_1 = \sum_{i=0}^{\nu} \mu_i, \tag{2.56}$$

где $\mathcal{P}(\nu = n) = C_{t_2}^n p_{21}^n p_{22}^{t_2-n}$ и $\mathcal{P}(\mu_i = r) = p_{11}^{r-1}p_{12}$, причем случайные величины ν и μ_i независимы между собой.

Доказательство. Соотношение (2.56) следует из соображений симметрии из соотношения (2.55).

Ряд предельных теорем могут быть получены из (2.55) и (2.56), при стремлении к бесконечности t_1 и t_2 соответственно. В частности, при $t_2 \to \infty$ следует известная предельная теорема о представимости t_1 в виде суммы случайного числа ν одинаково геометрически распределенных независимых случайных величин μ_i, независимых от ν, распределенных по закону Пуассона со средним $\lambda = tp_{21}$.

§ 2.7. Дискретная стохастическая функция

2.7.1. *Общее определение дискретной стохастической зависимости* (СЗ). В п. 2.1.1. было приведено общее описание стохастических аргумента, зависимости и функции. Дадим теперь точное определение дискретного варианта стохастической зависимости.

Пусть одним и тем же моментам времени $\sigma = (1, 2, \ldots, t, \ldots, s)$ соотнесены два набора множеств предметов

$$A^{(t)} = (A_1^{(t)}, \ldots, A_a^{(t)}, \ldots, A_a^{(t)}) \text{ и } B^{(t)} = (B_1^{(t)}, \ldots, B_\beta^{(t)}, \ldots, B_b^{(t)}) \; (t = \overline{1, s}),$$

где, вообще говоря, $a \neq b$.

Определим на множестве моментов времени σ обобщенные τ-проекции в смысле п. 2.2.1

$$y_\tau = \left\{B_{\beta_{i_1}}^{(i_1)}, \ldots, B_{\beta_{i_k}}^{(i_k)}, \ldots, B_{\beta_{i_t}}^{(i_t)}\right\} \in R_\tau^*,$$

где $\tau = (i_1, \ldots, i_k, \ldots, i_t) \subseteq \sigma$, в частности, при $\tau = \sigma$, $y_\tau = y$. Зафиксируем обобщенное расположение

$$x = \left\{A_{\alpha_1}^{(1)}, \ldots, A_{\alpha_t}^{(t)}, \ldots, A_{\alpha_s}^{(s)}\right\} \in R.$$

Пусть совокупность обобщенных τ-проекций y_τ стохастически определена при фиксированном x заданием условных вероятностей

$$\mathcal{P}(y_\tau/x) = p_x(y_\tau)$$

для всех $y_\tau \in R^*_\tau$ и $\tau \subseteq \sigma$. Так же как и набор вероятностей $p(x_\tau)$ в п. 2.2.1, набор условных вероятностей $p_x(y_\tau)$ должен удовлетворять условиям

$$\left.\begin{aligned} &1'.\ \sum_{\beta_{i_k}}^{b} p_x((\beta_{i_1}, \ldots, \beta_{i_{k-1}}, \beta_{i_k}, \beta_{i_{k+1}}, \ldots, \beta_{i_t})) = \\ &= p_x((\beta_{i_1}, \ldots, \beta_{i_{k+1}}, \beta_{i_{k+1}}, \ldots, \beta_{i_t})); \\ &2'.\ \sum_{y \in R^*} p_x(y) = 1. \end{aligned}\right\} \qquad (2.57)$$

Так же как и в указанном случае, для задания набора условных вероятностей $p_x(y_\tau)$ (2.57) достаточно задать лишь набор условных вероятностей $p_x(y)$ обобщенных расположений y в виде b^s произвольных неотрицательных величин, дающих в сумме единицу. Тогда суммированием по всем β_i с индексами $i \in \sigma - \tau$ получим

$$p_x(y_\tau) = \sum_{\sigma - \tau} p_x(y).$$

Определенный соотношениями (2.57) набор условных вероятностей $p_x(y_\tau)$ для всех $x \in R$ будем называть стохастической зависимостью (СЗ) и обозначать $\mathcal{P}(B/A)$. Обобщенные расположения x и предметы $A_\alpha(t)$ будем называть значениями и элементарными значениями аргумента соответственно. Обобщенные расположения y, τ-проекции y_τ и предметы $B^{(t)}_\beta$ будем называть значениями, τ-значениями и элементарными значениями случайной функции значения аргумента x[1].

Если все множества $B^{(t)}$ тождественны $B^{(t)} = B(t = \overline{1, s})$, то будем говорить, что имеет место однородная случайная функция. В этом случае y_τ допускает представление

$$y_\tau = (\tau^*_1, \ldots, \tau^*_\beta, \ldots, \tau^*_b)\left(\bigcup_{\beta=1}^{b} \tau^*_\beta = \tau,\ \tau^*_\beta \cap \tau^*_{\beta'} = \varnothing(\beta \neq \beta')\right),$$

где τ^*_β означает множество моментов τ, в которых появляются элементарные значения случайной функции B_β. Неотрицательные целые числа $N(\tau^*_\beta) = t^*_\beta\ (\beta = \overline{1, b})$ являются компонентами вектора частот $\bar{t}^* = (t^*_1, \ldots, t^*_\beta, \ldots, t^*_b)\left(\sum_{\beta=1}^{b} t^*_\beta = t\right)$.

Далее можно определить условные вероятности частот случайной функции

$$P_{tx}(\bar{t}) = \sum_{y_\tau \in \mathscr{E}^{\bar{t}}_\tau} p_x(y_\tau), \qquad (2.58)$$

где множество $\mathscr{E}^{\bar{t}}_\tau$ определено в гл. 1.

Как и в п. 2.2.2 можно перейти к матричному представлению значения y случайной функции

$$\mathrm{y} = \{\bar{e}_{\beta_1}, \ldots, \bar{e}_{\beta_t}, \ldots, \bar{e}_{\beta_s}\}$$

[1] Обоснование такой терминологии содержится в п. 2.1.1.

в виде $(b\times s)$-матрицы, состоящей из s вектор-столбцов $\bar{e}_\beta = {}^{\beta}\left\{\begin{pmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{pmatrix}\right\}b$,

ассоциированных элементарным значениям $B_\beta^{(t)}$ случайной функции.

При этом τ-значения y_τ случайной функции

$$\mathbf{y}_\tau = \{\bar{e}_{\beta_{t_1}}, \ldots, \bar{e}_{\beta_{t_k}}, \ldots, \bar{e}_{\beta_{t_t}}\}$$

соответствует $(b\times t)$-подматрице матрицы $\mathbf{y}$.

Введем формальный матричный аргумент

$$\mathbf{v} = (\bar{v}_1, \ldots, \bar{v}_t, \ldots, \bar{v}_s)$$

в виде $(b\times s)$-матрицы, состоящей из s b-мерных вектор-столбцов $\bar{v}_t = \begin{pmatrix} v_1^{(t)}\\ \vdots\\ v_\beta^{(t)}\\ \vdots\\ v_b^{(t)}\end{pmatrix}$.

Тогда производящая значений случайной функции имеет вид

$$g_x(\mathbf{v}) = \sum_{\mathbf{y}\in R^*} p_x(\mathbf{y})\,\mathbf{v}^{\mathbf{y}} \tag{2.59}$$

и производящая τ-значений случайной функции

$$g_x(\mathbf{v}_\tau) = \sum_{\mathbf{y}_\tau\in R_\tau^*} p_x(\mathbf{y}_\tau)\,\mathbf{v}_\tau^{\mathbf{y}_\tau}$$

получается из производящей (2.59) заменой столбцов матрицы $\mathbf{v}$, стоящих на местах τ, b-мерными единичными вектор-столбцами.

Рассмотрим случай тождественных $B^{(t)} = B\,(t = \overline{1, S})$. Тогда b-мерный вектор-столбец частот $\bar{t}^*$ имеет следующее выражение

$$\bar{t}^* = \mathbf{y}_\tau \bar{e} = \sum_{k=1}^{t} \bar{e}_{\alpha_{t_k}}, \tag{2.60}$$

где $\bar{e}$ — t-мерный единичный вектор-столбец.

Тогда производящая частот $\bar{t}$ случайной функции, учитывая соотношения (2.58) и (2.60), имеет вид

$$g_{t,x}(\bar{v}) = g_x(v_\tau)\Big|_{v_{\beta_{t_k}} = \bar{v}} = \sum_{\bar{t}} P_{tx}(\bar{t}^*)\,\bar{v}^{\bar{t}^*},$$

в котором все вектор-столбцы матрицы v_τ заменены на один и тот же вектор-столбец $\bar{v}$.

2.7.2. *Общее определение дискретной стохастической функции* (СФ). Теперь можно дать точное определение понятия дискретной стохастической

функции (СФ). В самом деле, на основании формулы полной вероятности по вероятностям $p(x)$ и $p_x(y)$ можно вычислить вероятности

$$\mathcal{P}(y) = p^*(y) = \sum_{x \in R} p(x)\, p_x(y). \tag{2.61}$$

А на основании последних вычислить вероятности $p(y_\tau)$, удовлетворяющие условиям, аналогичным условиям (2.6) п. 2.2.1. Так, определенную совокупность вероятностей $p(y_\tau)$ будем называть стохастической функцией (СФ) и обозначать $\mathcal{P}(B)$. Обобщенное расположение y назовем значением СФ, обобщенную τ-проекцию y назовем τ-значением СФ и предмет $B_\beta^{(\alpha)}$ назовем элементарным значением СФ.

Аналогично производящим СА определяется производящая СФ.

Производящая значений СФ

$$g^*(\mathbf{v}) = \sum_{\mathbf{y} \in R^*} p^*(\mathbf{y})\, \mathbf{v}^{\mathbf{y}} = \sum_{x \in R} p(x)\, g_x(\mathbf{v}). \tag{2.62}$$

Производящая τ-значений СФ

$$g^*(\mathbf{v}_\tau) = \sum_{\mathbf{y}_\tau \in R_\tau^*} p^*(\mathbf{y}_\tau)\, \mathbf{v}_\tau^{\mathbf{y}_\tau} = \sum_{x \in R} p(x)\, g_x(\mathbf{v}_\tau).$$

Производящая частот СФ

$$g^*(\bar{v}) = \sum_{\bar{t}^*} P_{\bar{t}}^*(\bar{t}^*)\, \bar{v}^{\bar{t}^*} = \sum_{x \in R} p(x)\, g_{tx}(\bar{v}).$$

Вероятностная природа СФ, определяемая вероятностной природой СА и СЗ, богаче каждого из двух определяющих ее факторов. Как уже отмечалось во введении, эта схема может служить вероятностной моделью черного ящика с заданным потенциальным запасом возможных случайных выходных реакций (СЗ) на каждое воздействие извне на его вход. При этом воздействия извне на вход черного ящика могут оказаться случайными (СА), в силу чего на выходе черного ящика возникают случайные реакции (СФ) на эти внешние воздействия.

Перейдем к обсуждению соображений в пользу введения понятия стохастической функции.

Современное понятие стохастического процесса обладает общностью, достаточной для включения только что рассмотренной схемы в виде частного случая. В самом деле, по вероятностям $p(x)$ и $p_x(y)$ вычисляются совместные вероятности

$$p(x, y) = p(x)\, p_x(y)$$

и затем пара значений (x, y) может рассматриваться как стохастический аргумент (процесс), как это делается в комментариях к работе А. А. Маркова [8] в [30].

Однако такого рода рассмотрения, симметричные относительно x и y, часто не отражают специфики ряда задач, например задач выбора между гипотезами (гл. 3), где ограничиваются заданием лишь СЗ в виде так называемой функции правдоподобия $p_x(y)$. Далее в рассмотренном выше общем случае черного ящика первоначально заданным является СЗ $\mathcal{P}(B/A)$ и в этом случае стохастический аргумент $\mathcal{P}(A)$ (внешнее воздействие) и стохастическая функция $\mathcal{P}(B)$ (реакция) существенно неравноправны. В гл. 4

подробно комментируется естественность понятия стохастической функции для общей теории связи, где стохастический аргумент $\mathcal{P}(A)$ (вход) и стохастическая функция $\mathcal{P}(B)$ (выход) существенно неравноправны. В самом деле, СА стохастически задан набором вероятностей $\mathcal{P}(A)$ вне зависимости от СФ. Напротив СФ становится стохастически определенной при фиксированной СЗ соотношением (2.61) лишь при стохастической определенности СА. Бо́льшая стохастическая определенность СА при фиксированной СЗ и задании конкретных значений СФ (см. обратные стохастические функции п. 2. 7. 6) указывает лишь на целесообразность выделения понятия СА и СФ.

2.7.3. *Специальные случаи* СЗ. Как и для СА, положительное изучение СЗ связано со специальной структурой условных вероятностей $p_x(y)$. Отправляясь от изученных в § 2. 3 и 2.4 марковского и симметричного случаев СА, здесь можно рассматривать большое число специальных случаев СЗ в зависимости от структуры вхождения x и y в условную вероятность $p_x(y)$.

Например, в предположении конечной по времени зависимости условных вероятностей $p_x(y)$ по x соответствующей группировкой (см. п. 2, 3, 1) можно ограничиться случаем

$$p_x(y) = P_{A^{(t)}_{\alpha(t)}}(y) \quad (t = \overline{1, s}), \tag{2.63}$$

когда фиксация одной из компонент $x = (A^{(1)}_{\alpha}, \ldots, A^{(t)}_{\alpha}, \ldots, A^{(s)}_{\alpha})$ достаточна для определения условной вероятности (2.63). Далее, в зависимости от предположений о марковости или симметричности СЗ будем иметь следующие выражения для (2.63)

$$p_x(y) = \mathcal{P}(B^{(t)}_{\beta_t} / A^{(t)}_{\alpha_t})\, \mathcal{P}(B^{(t)}_{\beta_{t+1}} / B^{(t)}_{\beta_t}, A^{(t)}_{\alpha_t}) \ldots \mathcal{P}(B^{(s)}_{\beta_s} / B^{(s-1)}_{\beta_{s-1}}, A^{(t)}_{\alpha_t})$$

и

$$p_x(y) = \mathcal{P}(y / A^{(t)}_{\alpha_t}) = \frac{\mathcal{P}(\bar{\mathcal{E}}^{m^*}_{\sigma} / A_{\alpha_t})}{C^m_s}, \quad \text{для } y \in \bar{\mathcal{E}}^{m^*}_{\sigma}$$

соответственно.

В дальнейшем изложении мы ограничимся рассмотрением простейшего случая СЗ, определяемой набором условных вероятностей

$$\mathcal{P}(B^{(t)}_{\beta(t)} / A^{(t)}_{\alpha(t)}) = p^{\beta}_{\alpha}(t).$$

Такого рода СЗ будем называть СЗ с независимыми переходами. В этом случае, если $x = \{A^{(t)}_{\alpha_t}\}$ и $y = \{B^{(t)}_{\beta_t}\}$, то

$$p_x(y) = \prod_{t=1}^{s} p^{\beta_t}_{\alpha_t}(t) = \prod_{t=1}^{s} \bar{p}^{\bar{e}_{\beta_t}}_{\alpha_t}(t). \tag{2.64}$$

Итак, здесь СЗ стохастически определена заданием набора матриц переходов

$$\mathbf{p}(t) = \| p^{\beta}_{\alpha}(t) \| = \{\bar{p}_{\alpha}(t)\} \quad (t = \overline{1, s}).$$

В дальнейшем изложении, говоря о СЗ, мы будем иметь в виду СЗ с независимыми переходами. В этом случае производящая СЗ имеет особо простой вид. В самом деле, из (2.59) и (2.64) имеем условную производящую

$$g_x(\mathbf{v}) = \sum_{\mathbf{y}} p_x(\mathbf{y})\, \mathbf{v}^{\mathbf{y}} = \sum_{\mathbf{y}} p_x(\mathbf{y}) \prod_{t=1}^{s} \bar{v}_t^{\bar{e}_{\beta_t}} = \sum_{\mathbf{y}} \prod_{t=1}^{s} (\bar{p}_{\alpha_t}(t).\bar{v}_t)^{\bar{e}_{\beta_t}}. \tag{2.65}$$

Заметим, что скалярное произведение $\bar{f}\bar{v}$ выражается через векторное покомпонентное произведение $(\bar{f}\cdot\bar{v})$ по формуле

$$\bar{f}\bar{v} \equiv \sum_{\gamma=1}^{c} (\bar{f}\cdot\bar{v})^{\bar{e}_\gamma} \equiv \sum_{\gamma=1}^{c} f_\gamma v_\gamma.$$

Используя это тождество, преобразуем (2.65)

$$\sum_{\mathbf{y}} \prod_{t=1}^{s} (\bar{p}_{a_t}(t)\cdot\bar{v}_t)^{\bar{e}_{\beta_t}} = \prod_{t=1}^{s} \left(\sum_{\beta_t=1}^{b} p_{a_t}^{\beta_t}(t)\, v_{\beta_t}^{(t)} \right) = \prod_{t=1}^{s} (\mathbf{p}(t)\,\bar{v}_t)^{\bar{e}_{a_t}}.$$

Итак, имеем окончательно соотношение

$$g_x(\mathbf{v}) = g_x(\bar{v}_1, \ldots, \bar{v}_s) = \prod_{t=1}^{s} (\mathbf{p}(t)\,\bar{v}_t)^{\bar{e}_{a_t}}, \tag{2.66}$$

важное для дальнейшего изложения.

Из (2.66), полагая $\bar{v}_t = \bar{v}$, получим условное распределение частот

$$g_x(\bar{v}) = \prod_{t=1}^{s} (\mathbf{p}(t)\,\bar{v})^{\bar{e}_{a_t}}. \tag{2.67}$$

2.7.4. *Основное соотношение между производящими значений СА и СФ.*

Теорема 2.3. Производящая СФ $g^*(\bar{v}_1, \ldots, \bar{v}_t, \ldots, \bar{v}_s)$ связана с производящей СА $g(\bar{u}_1, \ldots, \bar{u}_t, \ldots, \bar{u}_s)$ соотношением

$$g^*(\bar{v}_1, \ldots, \bar{v}_t, \ldots, \bar{v}_s) = g(\mathbf{p}(1)\bar{v}_1, \ldots, \mathbf{p}(t)\bar{v}_t, \ldots, \mathbf{p}(s)\bar{v}_s) \tag{2.68}$$

Доказательство. Имеем, согласно (2.62) и (2.66),

$$g^*(\bar{v}_1, \ldots, \bar{v}_t, \ldots, \bar{v}_s) = \sum_{\mathbf{y}} p(\mathbf{y}) \prod_{t=1}^{s} \bar{v}_t^{\bar{e}_{\beta_t}} = \sum_{x} p(x)\, g_x(\bar{v}_1, \ldots, \bar{v}_t, \ldots, \bar{v}_s) =$$

$$= \sum_{\mathbf{x}} p(\mathbf{x}) \prod_{t=1}^{s} (\mathbf{p}(t)\,\bar{v}_s)^{\bar{e}_{a_t}}.$$

Сравнив полученное выражение с выражением производящей СА

$$g(\mathbf{u}) = g(\bar{u}_1, \ldots, \bar{u}_t, \ldots, \bar{u}_s) = \sum_{\mathbf{x}} p(\mathbf{x}) \prod_{t=1}^{s} \bar{u}_t^{\bar{e}_{a_t}},$$

получим утверждение теоремы.

Из теоремы 2.3 могут быть выведены следствия.

Следствие 2.3.1. Если СА образован по схеме Пуассона, т. е. $p(x) = \prod_{t=1}^{s} p_{a_t}(t)$, то и СФ образована по той же схеме, т. е.

$$p^*(y) = \prod_{t=1}^{t} p^*_{\beta_t}(t),$$

где

$$p^*_{\beta_t}(t) = \sum_{a_t=1}^{a} p_{a_t}(t)\, p_{a_t}^{\beta_t}(t),$$

или в векторной записи

$$\overline{p_t^*} = \overline{p}_t \mathbf{p}(t), \tag{2.69}$$

Доказательство. В самом деле из (2.68) и (2.69) и (2.18) имеем

$$g^*(\overline{v}_1, \ldots, \overline{v}_t, \ldots, \overline{v}_s) = g\{\mathbf{p}(1)\overline{v}_1, \ldots, \mathbf{p}(t), \overline{v}_t, \ldots, \mathbf{p}(s)\overline{v}_s\} =$$

$$= \prod_{t=1}^{s} \overline{p}_t p(t) \overline{v}_t = \prod_{t=1}^{s} \overline{p_t^*} \overline{v}_t,$$

откуда следует утверждение следствия.

Следствие 2.3.2. Каков бы ни был СА, если СЗ определена матрицами переходов $\mathbf{p}(t) = \overline{e}\overline{p_t}$ с постоянными столбцами, то СФ образована по схеме Пуассона вида

$$p^*(y) = \prod_{t=1}^{s} \overline{p_t}^{\overline{e}\overline{\beta}_t}.$$

Доказательство. Из (2.68) имеем

$$g^*(\overline{v}_1, \ldots, \overline{v}_t, \ldots, \overline{v}_s) = g\{\mathbf{p}(1)\overline{v}_1, \ldots, \mathbf{p}(t)\overline{v}_t, \ldots, \mathbf{p}(s)\overline{v}_s\} =$$

$$= g(\overline{e}\,\overline{p}_1\overline{v}_1, \ldots, \overline{e}\,\overline{p}_t\overline{v}_t, \ldots, \overline{e}\,\overline{p}_s\overline{v}_s) = \prod_{t=1}^{s} (\overline{p}_t\overline{v}_t)\, g(\underbrace{\overline{e}, \ldots, \overline{e}}_{a}) =$$

$$= \prod_{t=1}^{s} (\overline{p}_t\overline{v}_t) \cdot 1,$$

что и требовалось доказать.

Заметим, что рассмотренный в следствии 2.3.2 случай является вырожденным случаем, когда СФ оказывается независимой от СА в вероятностном смысле.

Следствие 2.3.3. Если СА является марковским, определяемым производящей функцией

$$g(\overline{u}_1, \ldots, \overline{u}_t, \ldots, \overline{u}_s) = \overline{p}_1 D(\overline{u}_1) \prod_{t=1}^{s} \mathcal{P}_t D(\overline{u}_t)\overline{e}, \tag{2.70}$$

то СФ имеет производящую функцию

$$g^*(v_1, \ldots, v_t, \ldots, v_s) = \overline{p}_1 D(\mathbf{p}(1)v_1) \prod_{t=2}^{s} \mathcal{P}_t D(\mathbf{p}(t)\overline{v}_t) \tag{2.71}$$

и поэтому не является марковской.

Доказательство. Утверждение следствия очевидно из-за несовпадения структуры производящих (2.70) и (2.71), что и требовалось доказать.

На факт, указанный в следствии 2.3.3, впервые обратил внимание В. И. Романовский [25, стр. 270]. Более того, прямыми методами им было показано, что в этом случае СФ образует так называемую бесконечно усложняющуюся цепь, в которой вероятность появления события $B_{\beta(t)}^{(t)}$ зависит от всех событий, появившихся до него, и остающуюся при этом в однородном случае устойчивой (постоянная абсолютная вероятность $q_\beta = \mathcal{P}(B_\beta^{(t)})$), если устойчив СА.

Сопоставляя результаты следствий 2.3.1 и 2.3.3, заключаем, что если схема Пуассона инвариантна по отношению к рассматриваемому виду СЗ, то схема Маркова не обладает таким свойством. Можно показать, что слабозависимая схема инвариантна в указанном выше смысле.

Следствие 2.3.4. Если слабо зависимый СА определяется производящей

$$g(\mathbf{u}) = (\bar{p}\bar{u}_1)\prod_{t=2}^{s}\left[\bar{p}\bar{u}_t + \frac{\bar{u}'_{t-1}D(\bar{p})(P-\bar{e}\bar{p})\bar{u}_t}{\bar{p}\bar{u}_{t-1}}\right] + O(\theta^2) \qquad (2.72)$$

и однородные СЗ определяются матрицей переходов $\mathbf{p} = \| p_\alpha^\beta \|$ и СА — матрицей переходов $P = \| P_\alpha^{\alpha'} \|$, то СФ определяется производящей

$$g^*(\mathbf{v}) = g(\mathbf{p}\bar{v}_1, \ldots, \mathbf{p}\bar{v}_s) = (\bar{q}\bar{v}_1)\prod_{t=2}^{s}\left[\bar{q}\bar{v}_t + \frac{\bar{v}'_{t-1}D(\bar{q})(Q-\bar{e}\bar{q})\bar{v}_t}{\bar{q}\bar{v}_{t-1}}\right] + O(\theta^2),$$

где $\bar{q} = \bar{p}\mathbf{p}$; $Q = \| Q_\beta^{\beta'} \| = D(\overline{1/q})\,\mathbf{p}'D(\bar{p})P\mathbf{p}$ или в развернутом виде $Q_\beta^{\beta'} = \sum_{\alpha\alpha'} \frac{1}{q_\beta} p_\alpha^\beta p_\alpha P_\alpha^{\alpha'} p_{\alpha'}^{\beta'}$. Отсюда следует, что СФ снова слабо зависимая.

Доказательство. Используя (2.72) и (2.68), получим

$$g^*(\mathbf{v}) = g(\mathbf{p}\bar{v}_1, \ldots, \mathbf{p}\bar{v}_s) = (\bar{q}\bar{v}_1)\prod_{t=2}^{s}\left[\bar{q}\bar{v}_t + \frac{\bar{v}'_{t-1}D(\bar{q})\cdot D(\overline{1/q})\,\mathbf{p}'D(\bar{p})(P-\bar{e}\bar{p})\mathbf{p}\bar{v}_t}{\bar{q}\bar{v}_{t-1}}\right] +$$

$$+ O(\theta^2) = (\bar{q}\bar{v}_1)\prod_{t=2}^{s}\left[\bar{q}\bar{v}_t + \frac{\bar{v}'_{t-1}D(\bar{q})\cdot D(\overline{1/q})\,\mathbf{p}'D(\bar{p})(P-\bar{e}\bar{p})\mathbf{p}\bar{v}_t}{\bar{q}\bar{v}_{t-1}}\right] + O(\theta^2). \qquad (2.72')$$

Заметим, что $D(\overline{1/q})\,\mathbf{p}'D(\bar{p})$ является марковской $(b\times a)$-матрицей, так как

$$D(\overline{1/q})\,\mathbf{p}'D(\bar{p})\bar{e} = D(\overline{1/q})\,\bar{p}'\bar{p}' = D(\overline{1/q})(\bar{p}\mathbf{p})' = D(\overline{1/q})\,\bar{q}' = \bar{e}, \qquad (2.73)$$

откуда

$$D(\overline{1/q})\,\mathbf{p}'D(\bar{p})\bar{e}\bar{p}\mathbf{p} = \bar{e}\bar{p}\mathbf{p} = \bar{e}\bar{q}. \qquad (2.74)$$

Рассмотрим $(b\times b)$-матрицу

$$Q = D(\overline{1/q})\,\mathbf{p}'D(\bar{p})P\mathbf{p} = \| Q_\beta^{\beta'} \|. \qquad (2.75)$$

Из (2.75) следует, что ее элементы $Q_\beta^{\beta'}$ имеют представление

$$Q_\beta^{\beta'} = \sum_{\alpha\alpha'} \frac{1}{q_\beta} p_\alpha^\beta p_\alpha P_\alpha^{\alpha'} p_{\alpha'}^{\beta'} = \sum_{\alpha\alpha'} P(A_\alpha/B_\beta)\,P(A_{\alpha'}/A_\alpha)\,P(B_{\beta'}/A_{\alpha'}) = P(B_{\beta'}/B_\beta),$$

из которого следует, что матрица Q является марковской. Легко проверить, что $Q\bar{e} = \bar{e}$ и $\bar{q}Q = \bar{q}$, т. е. $\bar{q}$ является собственным вектором Q с собственным значением 1. В самом деле

$$\bar{q}Q = \bar{q}D(\overline{1/q})\,\mathbf{p}'D(\bar{p})P\mathbf{p} = \bar{e}'\mathbf{p}'D(\bar{p})P\mathbf{p} = \bar{e}'D(\bar{p})P\mathbf{p} = \bar{p}P\mathbf{p} = \bar{p}\mathbf{p} = \bar{q}.$$

Раскрывая скобки в соотношении (2.72′) и используя соотношения (2.73) и (2.74), получим утверждение следствия.

2.7.5. *Группировка значений* СЗ. Пусть заданы основные множества $A^{(t)} = \{A_\alpha^{(t)}\}, B^{(t)} = \{B_\beta^{(t)}\}$, стохастический аргумент $\mathcal{P}(A)$ и стохастическая зависимость, определяемая набором матриц вероятностей переходов

$$\mathbf{p}(t) = \| p_\alpha^\beta(t) \| \; (t = \overline{1,s},\ \alpha = \overline{1,a},\ \beta = \overline{1,b}).$$

В ряде случаев необходим переход от «мелких» событий $A^{(t)}$ и $B^{(t)}$ к более укрупненным событиям $\widetilde{A}^{(t)} = \{\widetilde{A}_{\alpha'}^{(t)}\}$ и $\widetilde{B}^{(t)} = \{\widetilde{B}_{\beta'}^{(t)}\}$ (см. п. 2.2.3), где события

$$\widetilde{A}_{I_i}^{(t)} = \bigcup_{\alpha \in I_i} A_\alpha^{(t)}, \quad \widetilde{B}_{I_j^*}^{(t)} = \bigcup_{\beta \in I_j^*} B_\beta^{(t)},$$

причем

$$\bigcup_{i=1}^{a'} I_i = \{1, \ldots, \alpha, \ldots, a\}, \quad I_i \cap I_{i'} = \phi \ (\text{при } i \neq i');$$

$$\bigcup_{j=1}^{b'} I_j^* = \{1, \ldots, \beta, \ldots, b\}, \quad I_j^* \cap I_{j'}^* = \phi \ (\text{при } j \neq j');$$

переход от элементарных значений $A_\alpha^{(t)}$ и $B_\beta^{(t)}$ к элементарным значениям $\widetilde{A}_{I_i}^{(t)}$ и $\widetilde{B}_{I_j^*}^{(t)}$ будем называть группировкой.

Пусты заданы абсолютные вероятности элементарных значений $A_\alpha^{(t)}$, равные $\overline{p}_t = (p_\alpha^{(t)})$; тогда по матрицам $\mathbf{p}(t)$ $(t = \overline{1,s})$ можно вычислить соответствующие матрицы $\widetilde{\mathbf{p}}(t) = \| p_{I_i}^{I_j^*}(t) \|$ для группированных элементарных значений, определяющие группированную СЗ.

Имеют место соотношения, следующие из формулы полной вероятности [25]

$$p_{I_i}^{I_j^*}(t) = \sum_{\alpha \in I_i} \frac{p_\alpha(t)}{\sum\limits_{\tau \in I_i} p_\tau(t)} \sum_{\beta \in I_j^*} p_\alpha^\beta(t), \tag{2.76}$$

где $p_\alpha(t)$ — абсолютные вероятности.

Рассмотрим частный случай бинарной группировки, когда $a' = b' = 2$. Положим $I_1 = I$; $I_2 = \overline{I}$; $I_1^* = I^*$; $I_2^* = \overline{I}^*$; $\sum\limits_{\tau \in I} p_\tau(t) = p_I(t)$; $\sum\limits_{\beta \in I^*} p_\alpha^\beta(t) = p_\alpha^{I^*}(t)$. Тогда из соотношения (2.76) будем иметь

$$\left.\begin{aligned} p_I^{I^*}(t) &= \sum_{\alpha \in I} \frac{p_\alpha(t) \cdot p_\alpha^{I^*}(t)}{p_I^{(t)}}, \quad p_I^{\overline{I}^*}(t) = 1 - p_I^{I^*}(t); \\ p_{\overline{I}}^{I^*}(t) &= 1 - p_{\overline{I}}^{\overline{I}^*}(t), \quad p_{\overline{I}}^{\overline{I}^*}(t) = \sum_{\alpha \in \overline{I}} \frac{p_\alpha(t)\, p_\alpha^{\overline{I}^*}(t)}{p_I(t)}. \end{aligned}\right\}$$

2.7.6. *Обратная стохастическая зависимость.* Пусть абсолютные вероятности событий $A^{(t)} = (A_\alpha^{(t)})$ равны $\overline{p}_t = (p_\alpha(t))$; тогда по СЗ заданной набором матриц вероятностей переходов $\mathbf{p}(t) = \| p_\alpha^\beta(t) \|$ можно определить абсолютные вероятности $\overline{q}_t = (q_\beta(t)) = \overline{p}_t \mathbf{p}(t)$ событий $B^{(t)} = (B_\beta^{(t)})$, а также по формулам Байесса условные вероятности

$$P(A_\alpha^{(t)} / B_\beta^{(t)}) = \frac{p_\alpha(t)\, p_\alpha^\beta(t)}{q_\beta(t)} = q_\beta^\alpha(t) \quad (q_\beta(t) > 0), \tag{2.77}$$

задающие набор марковских матриц перехода

$$\mathbf{q}(t) = \| q_\beta^\alpha(t) \| = (\overline{q}_\beta(t)), \quad (t = \overline{1,s}). \tag{2.78}$$

Легко показать, что $\overline{q}_t \mathbf{q}(t) = \overline{p}(t)$.

Будем говорить, что набор матриц (2.78) определяет обратную стохастическую зависимость $\mathcal{P}(A/B)$, по отношению к СЗ $\mathcal{P}(B/A)$. Впервые для однородного случая обратную стохастическую зависимость (ОСЗ) рассматривал В. И. Романовский [25]. С ней же связано известное тождество Шеннона [2]. В самом деле, в однородном случае имеем из (2.77)

$$p_\alpha p_\alpha^\beta = q_\beta q_\beta^\alpha. \tag{2.79}$$

Взяв натуральный логарифм от обеих частей тождества (2.79), будем иметь

$$\ln p_\alpha + \ln p_\alpha^\beta \equiv \ln q_\beta + \ln q_\beta^\alpha. \tag{2.80}$$

Перемножая соответствующие части соотношений (2.79) и (2.80) и суммируя по всем α и β, получим, используя обозначения п.1.6.1,

$$-h(\bar{p}) - \sum_{\alpha=1}^{a} p_\alpha h(\bar{p}_\alpha) = -h(\bar{q}) - \sum_{\beta=1}^{b} q_\beta h(\bar{q}_\beta),$$

откуда имеем соотношения

$$h(\bar{q}) - h(\bar{p}) \equiv \sum_{\alpha=1}^{a} p_\alpha h(\bar{p}_\alpha) - \sum_{\beta=1}^{b} q_\beta h(\bar{q}_\beta);$$

$$h(\bar{q}) - \sum_{\alpha=1}^{a} p_\alpha h(\bar{p}_\alpha) \equiv h(\bar{p}) - \sum_{\beta=1}^{b} q_\beta h(\bar{q}_\beta),$$

из которых последнее было впервые установлено Шенноном [2] (см. подробнее гл. 5).

§ 2.8. Связь между распределениями частот стохастического аргумента и стохастической функции

2.8.1. *Соотношения между производящими частот СА и СФ.* Рассмотрим случай тождественных в разные моменты множеств значений $A^{(t)} = A = (A_1, \ldots, A_\alpha, \ldots, A_a)$ и $B^{(t)} = B = (B_1, \ldots, B_\beta, \ldots, B_b)$, когда имеет смысл говорить об множествах значений СА $x \in \mathscr{E}_\sigma^{\bar{m}}$ и СФ $y \in \mathscr{E}_\sigma^{\bar{m}^*}$ и соответствующих распределениях частот $\mathcal{P}(x \in \mathscr{E}_\sigma^{\bar{m}}) = P_s(\bar{m})$ и $\mathcal{P}(y \in \mathscr{E}_\sigma^{\bar{m}^*}) = P_s^*(\bar{m}^*)$. Производящие этих вероятностей получаются из производящих (2.9) и (2.62) заменой $\bar{u}_t = \bar{u}$ и $\bar{v}_t = \bar{v}$, соответственно:

$$g(\bar{u}) = g(\underbrace{\bar{u}, \ldots, \bar{u}}_{s}) = \sum_{\bar{m}} P_s(\bar{m}) \bar{u}^{\bar{m}} \quad \text{и} \quad g^*(\bar{v}) = g^*(\underbrace{\bar{v}, \ldots, \bar{v}}_{s}) = \sum_{\bar{m}^*} P_s^*(\bar{m}), \tag{2.81}$$

Из (2.81) и основного соотношения между производящими значений (2.68) следует основное соотношение между производящими частот

$$g^*(\bar{v}) = g(\mathbf{p}(1)\bar{v}, \ldots, \mathbf{p}(t)\bar{v}, \ldots, \mathbf{p}(s)\bar{v}).$$

Отсюда в однородном случае $\mathbf{p}(t) = \mathbf{p} = \| p_\alpha^\beta \|$ получаем особенно простое соотношение между производящими частот СА и СФ

$$g^*(\bar{v}) = g(\mathbf{p}\bar{v}). \tag{2.82}$$

Соотношение (2.82) можно получить и другим путем. В самом деле, при фиксированной частоте $\overline{m}$ СА распределение частоты $\overline{m}^*$ СФ $P_{\overline{m}}(\overline{m}^*)$ может быть представлено в виде композиции вероятностей $\mathbf{p} = (\overline{p}_\alpha) = \| p_\alpha^\beta \|$:

$$P_{\overline{m}}(\overline{m}^*) = \overline{p}_1^{[m_1]} * \ldots * \overline{p}_a^{[m_a]}$$

Производящая вероятностей $\overline{p}_\alpha = (p_\alpha^\beta)$ имеет вид

$$g_\alpha(\overline{v}) = (\overline{p}_\alpha \overline{v}) = \left(\sum_{\beta=1}^{b} p_\alpha^\beta v_\beta \right).$$

Производящая частот $\overline{m} = (m_\alpha)$ имеет вид

$$g(\overline{u}) = g(u_1, \ldots, u_a).$$

Отсюда, используя теорему об итерации производящих [23] в многомерном случае (см. (1.47)), получим для производящей частот $\overline{m}^*$ соотношение (2.82):

$$g^*(\overline{v}) = g(\overline{p}_1\overline{v}, \ldots, \overline{p}_\alpha\overline{v}, \ldots, \overline{p}_a\overline{v}) = g(\mathbf{p}\overline{v}).$$

Рассмотрим вырожденный детерминированный случай распределения

$$P_s(\overline{m}) = \begin{cases} 1, & \text{при } \overline{m} = \overline{m}_0, \\ 0, & \text{при } \overline{m} \neq \overline{m}_0. \end{cases}$$

В этом случае

$$g(\overline{u}) = \sum_{\overline{m}} P_s(\overline{m}) \overline{u}^{\overline{m}} = \overline{u}^{\overline{m}_0}.$$

Далее заметим, что в рассматриваемом случае из соотношения (2.67) для условной производящей частот в однородном случае при $x \in \mathscr{E}_\sigma^{\overline{m}}$ имеем

$$g_x(\overline{v}) = g_{\overline{m}}(\overline{v}) = \prod_{t=1}^{s} (\mathbf{p}\overline{v})^{\overline{\varepsilon}_\alpha(t)} = (\mathbf{p}\overline{v})^{\sum_{t=1}^{s} \overline{\varepsilon}_\alpha(t)} = (\mathbf{p}\overline{v})^{\overline{m}} = \prod_{\alpha=1}^{a} \left(\sum_{\beta=1}^{b} p_\alpha^\beta v_\beta \right)^{m_\alpha}.$$

Итак

$$g_{\overline{m}}(\overline{v}) = \prod_{\alpha=1}^{a} \left(\sum_{\beta=1}^{b} p_\alpha^\beta v_\beta \right)^{m_\alpha} = \sum_{\overline{m}^*} P_{\overline{m}}(\overline{m}^*) \overline{v}^{\overline{m}^*}. \qquad (2.82')$$

Отсюда, используя обобщенную формулу Ньютона, легко получим

$$P_{\overline{m}}(\overline{m}^*) = \sum_{\sum_{\alpha=1}^{a} m_{\alpha\beta} = m_\beta^*} \prod_{\alpha=1}^{a} C_{m_\alpha}^{\overline{m}_\alpha} \prod_{\beta=1}^{b} (p_\alpha^\beta)^{m_{\alpha\beta}},$$

что несложно следует и из прямых комбинаторных рассмотрений.

Пусть частоты $\overline{m}$ СА имеют полиномиальное распределение

$$P_s(\overline{m}) = C_s^{\overline{m}} \overline{p}^{\overline{m}}.$$

Тогда

$$g(\bar{u}) = \sum_{\bar{m}} P_s(\bar{m})\,\bar{u}^{\bar{m}} = \Big(\sum_{\alpha=1}^{a} p_\alpha u_\alpha\Big)^s = (\bar{p}\bar{u})^s$$

и, согласно (2.82),

$$g^*(\bar{v}) = g(\mathbf{p}\bar{v}) = (\bar{p}\mathbf{p}\bar{v})^s = (\bar{q}\bar{v})^s = \sum_{\bar{m}^*} C_s^{\bar{m}^*}\bar{q}^{\bar{m}^*}\bar{v}^{\bar{m}^*},$$

откуда следует, что

$$P_s^*(\bar{m}^*) = C_s^{\bar{m}^*}\bar{q}^{\bar{m}^*}.$$

Итак, если распределение частот $\bar{m}$ СА полиномиально с параметром $\bar{p}$, то распределение частот $\bar{m}^*$ СФ полиномиально с параметром $\bar{q} = \bar{p}\mathbf{p}$.

Следует заметить, что из бернуллиевости СА следует бернуллиевость СФ и полиномиальность частоты СА и СФ. Но из полиномиальности распределений частот СА и СФ не следует бернуллиевость их значений.

2.8.2. *Связь между моментами частот* СА и СФ. Будем исходить из следующей записи производящей СФ в однородном случае

$$g^*(\bar{v}) = g(\bar{p}\bar{v}) = g(\bar{u})\Big|_{u_\alpha = \sum_{\beta=1}^{b} p_\alpha^\beta v_\beta}.$$

Тогда

$$\frac{\partial g^*(\bar{v})}{\partial v_\beta} = \sum_{\alpha=1}^{a} \frac{\partial g(\bar{u})}{\partial u_\alpha}\frac{\partial u}{\partial v_\beta} = \sum_{\alpha=1}^{a} \frac{\partial g(u)}{\partial u_\alpha}\,p_\alpha^\beta, \tag{2.83}$$

$$\frac{\partial^2 g^*(\bar{v})}{\partial v_\beta\,\partial v_{\beta'}} = \sum_{\alpha\alpha'} \frac{\partial^2 g(\bar{u})}{\partial u_\alpha\,\partial u_{\alpha'}}\,p_\alpha^\beta\,p_{\alpha'}^{\beta'}. \tag{2.84}$$

Но

$$\frac{\partial g(\bar{u})}{\partial u_\alpha}\Big|_{\bar{u}=\bar{e}} = \mathrm{E}m_\alpha \quad \text{и} \quad \frac{\partial^2 g(u)}{\partial u_\alpha\,\partial u_{\alpha'}}\Big|_{\bar{u}=\bar{e}} = \begin{cases} \mathrm{E}\,m_\alpha m_{\alpha'}, & \text{при } \alpha \neq \alpha' \\ \mathrm{E}\,m_\alpha(m_\alpha - 1), & \text{при } \alpha = \alpha' \end{cases} \tag{2.85}$$

и так как при $\bar{u} = \bar{e}$, $\bar{v} = \bar{e}$, то, используя соотношения (2.85), получим из (2.83) и (2.84)

$$\mathrm{E}m_\beta^* = \sum_{\alpha=1}^{a} p_\alpha^\beta\,\mathrm{E}\,m_\alpha; \tag{2.86}$$

$$\left.\begin{aligned} \mathrm{E}m_\beta^* m_{\beta'}^* &= \sum_{\alpha=1}^{a} \mathrm{E}m_\alpha(m_\alpha - 1)\,p_\alpha^\beta\,p_\alpha^{\beta'} + \sum_{\alpha\neq\alpha'} \mathrm{E}m_\alpha m_{\alpha'}\,p_\alpha^\beta p_{\alpha'}^{\beta'}; \\ \mathrm{E}m_\beta^*(m_\beta^* - 1) &= \sum_{\alpha=1}^{a} \mathrm{E}m_\alpha(m_\alpha - 1)(p_\alpha^\beta)^2 + \sum_{\alpha\neq\alpha'} \mathrm{E}m_\alpha m_{\alpha'}\,p_\alpha^\beta p_{\alpha'}^\beta. \end{aligned}\right\} \tag{2.87}$$

В векторных обозначениях соотношения (2.86) и (2.87) имеют вид соответственно

$$\mathrm{E}\bar{m}^* = \mathrm{E}\bar{m}\cdot\mathbf{p}; \tag{2.88}$$

$$M_1^* = \mathbf{p}'M_1\mathbf{p}, \tag{2.89}$$

где

$$\mathrm{E}\overline{m} = (\mathrm{E}m_1, \ldots, \mathrm{E}m_\alpha, \ldots, \mathrm{E}m_a), \quad \mathrm{E}\overline{m}^* = (\mathrm{E}m_1^*, \ldots, \mathrm{E}m_\beta^*, \ldots, \mathrm{E}m_b^*);$$

$$M_1 = \| \mathrm{E}m_\alpha m_{\alpha'} \| - D(\mathrm{E}\overline{m}), \quad M_1^* = \| \mathrm{E}m_\beta^* m_{\beta'}^* \| - D(\mathrm{E}\overline{m}^*).$$

В частности, в полиномиальном случае $P_s(\overline{m}) = C_s^{\overline{m}}\overline{p}^{\overline{m}}$ имеем (см. (2.20)

$$M_1 = M_1(\overline{p}) = s(s-1)\overline{p}'\overline{p} = s^2\overline{p}'\overline{p} - s\overline{p}'\overline{p}. \tag{2.90}$$

Рассмотрим матрицу центральных моментов

$$M = \| \mathrm{E}(m_\alpha - \mathrm{E}m_\alpha)(m_{\alpha'} - \mathrm{E}m_{\alpha'}) \| = \| \mathrm{E}m_\alpha m_{\alpha'} \| - \mathrm{E}\overline{m}'\mathrm{E}\overline{m}.$$

В полиномиальном случае матрица $M = M(\overline{p})$ имеет вид (см. (2.90))

$$M(\overline{p}) = sD(\overline{p}) - s\overline{p}'\overline{p}.$$

Составим теперь разности

$$\left.\begin{aligned} M_1 - M &= \mathrm{E}\overline{m}'\mathrm{E}\overline{m} - D(\mathrm{E}\overline{m}), \\ M_1(\overline{p}) - M(\overline{p}) &= s^2\overline{p}'\overline{p} - sD(\overline{p}) \end{aligned}\right\}. \tag{2.91}$$

Если положить $\overline{p} = (1/s)\,\mathrm{E}\overline{m}$, то будем иметь из (2.91)

$$M_1 - M = M_1(\overline{p}) - M(\overline{p})$$

или

$$M_1 - M_1(\overline{p}) = M - M(\overline{p}). \tag{2.92}$$

Используя (2.89) и (2.90) будем иметь

$$\mathbf{p}'(M_1 - M_1(\overline{p}))\,\mathbf{p} = \mathbf{p}'M_1\mathbf{p} - \mathbf{p}'M_1(\overline{p})\,\mathbf{p} =$$

$$= M_1^* - s(s-1)\,\mathbf{p}'\overline{p}'\overline{p}\mathbf{p} = M_1^* - s(s-1)\,\overline{q}'\overline{q} = M_1^* - M_1(\overline{q}),$$

где $\overline{q} = \overline{p}\mathbf{p}$. Так как согласно (2.92)

$$M - M(\overline{p}) = M_1 - M_1(\overline{p}) \text{ и } M^* - M(\overline{q}) = M_1^* - M_1(\overline{q}),$$

то имеем

$$M_1^* - M_1(\overline{q}) = \mathbf{p}'(M_1 - M(\overline{p})\,\mathbf{p};$$

$$M^* - M(\overline{q}) = \mathbf{p}'(M - M(\overline{p})\,\mathbf{p}.$$

Последнее соотношение может быть переписано в следующей форме

$$M^* = M(\overline{q}) + \mathbf{p}'(M - M(\overline{p})\,\mathbf{p}. \tag{2.93}$$

Рассмотрим вырожденный случай, когда

$$P_s(\overline{m}) = \begin{cases} 1, & \text{если } \overline{m} = \overline{m}_0, \\ 0, & \text{если } \overline{m} \neq \overline{m}_0. \end{cases}$$

Тогда $M = O$, т. е. матрица вторых центральных моментов, состоит из нулевых элементов. Кроме того, $\overline{p} = \frac{1}{s}\mathrm{E}\overline{m}$, $\overline{q} = \overline{p}\mathbf{p}$ и в нашем случае

$\mathrm{E}\overline{m} = \overline{m}_0$; поэтому, учитывая, что $M(\overline{p}) = sD(\overline{p}) - \overline{s}\overline{p}'\overline{p}$, имеем из (2.93)

$$M^* = M(\overline{q}) + \mathbf{p}'(M - M(\overline{p}))\mathbf{p} =$$

$$= M(\overline{q}) - \mathbf{p}'\{O - M(\overline{p})\}\mathbf{p} = sD(\overline{q}) - s\overline{q}'\overline{q} - s\mathbf{p}'D(\overline{p})\mathbf{p} +$$

$$+ \mathbf{p}'s\overline{p}'\overline{p}\mathbf{p} = D(\overline{m}_0\mathbf{p}) - \mathbf{p}'D(\overline{m}_0)\mathbf{p}.$$

Итак, обозначая в нашем случае $M^* = M^*_{\overline{m}_0}$, имеем

$$M^*_{\overline{m}_0} = D(\overline{m}_0\mathbf{p}) - \mathbf{p}'D(\overline{m}_0)\mathbf{p}.$$

Далее, представим соотношения (2.88) и (2.93) в форме «отклонений» от вырожденного случая. Прежде всего, из (2.88) имеем

$$\mathrm{E}\overline{m}^* - s\overline{q} = [\mathrm{E}\overline{m} - s\overline{p}]\mathbf{p},$$

где $\overline{q} = \overline{p}\mathbf{p}$.

Положим $\mathbf{p} = \overline{e}\,\overline{r} + R$, где $\overline{r} = (r_1, \ldots, r_b)$, $\overline{r}\,\overline{e} = 1$. Но $\mathrm{E}\overline{m}\overline{e} = s$ и $\overline{p}\overline{e} = 1$; поэтому, обозначив

$$\overline{\Delta} = \mathrm{E}\overline{m} - s\overline{p} \quad \text{и} \quad \overline{\Delta}^* = \mathrm{E}\overline{m}^* - s\overline{q},$$

получим из (2.88)

$$\overline{\Delta}^* = \overline{\Delta}R. \tag{2.94}$$

Заметим, что $M\overline{e} = \overline{e}'M = O$; поэтому, обозначив $\mathbf{\Delta} = M - M(\overline{p})$ и $\mathbf{\Delta}^* = M^* - M(\overline{q})$, будем иметь из (2.93)

$$\mathbf{\Delta}^* = R'\mathbf{\Delta}R. \tag{2.95}$$

§ 2.9. Бинарный случай стохастической функции

В бинарном случае СФ, когда $a = b = 2$, общие многомерные соотношения упрощаются лишь для частот, переходя в одномерные. Рассмотрим основное соотношение между производящими частот в этом случае.

Имеем

$$g^*(v_1, v_2) = g(p_1^1 v_1 + p_1^2 v_2,\ p_2^1 v_1 + p_2^2 v_2), \tag{2.96}$$

где $\mathbf{p} = \begin{pmatrix} p_1^1 p_1^2 \\ p_2^1 p_2^2 \end{pmatrix}$ матрица переходов СЗ и

$$g(u_1, u_2) = \sum_{m_1+m_2=s} P_s(m_1, m_2)\, u_1^{m_1} u_2^{m_2} = u_2^s g\left(\frac{u_1}{u_2}, 1\right), \tag{2.97}$$

$$g^*(v_1, v_2) = \sum_{m_1^*+m_2^*=s} P_s^*(m_1^*, m_2^*)\, v_1^{*m_1} v_2^{*m_2} = v_2^s g^*\left(\frac{v_1}{v_2}, 1\right),$$

производящие функции распределений $P_s(m_1, m_2)$ и $P_s^*(m_1^*, m_2^*)$ частот СА и СФ соответственно.

Рассмотрим случай слабо зависимого СА. Тогда согласно (2.42)

$$g\left(\frac{u_1}{u_2}, 1\right) = \left[p\frac{u_1}{u_2} + q + \delta\frac{s-1}{s}pq\frac{\left(\frac{u_1}{u_2} - 1\right)^2}{p\frac{u_1}{u_2} + q}\right]^s + O(\delta^2)(u_1 - u_2), \tag{2.98}$$

где матрица переходов СА $\mathcal{P} = \begin{pmatrix} p + \delta q & q - \delta q \\ p - \delta p & q + \delta p \end{pmatrix}$. Из (2.97) и (2.98) получим

$$g(u_1, u_2) = u_2^s g\left(\frac{u_1}{u_2}, 1\right) = \left[pu_1 + qu_2 + \delta \frac{s-1}{q} pq \frac{(u_1 - u_2)^2}{pu_1 + qu_2}\right]^s + O(\delta^2)(u_1 - u_2). \quad (2.99)$$

Подставив в (2.96) явное выражение (2.99), после несложных преобразований получим производящую СФ

$$g^*(v_1, v_2) = \left[p^* v_1 + q^* v_2 + \delta' \frac{s-1}{s} p^* q^* \frac{(v_1 - v_2)^2}{p^* v_1 + q^* v_2}\right]^s + O(\delta')^2 (v_1 - v_2)$$

того же типа, что и производящая СА, где

$$p^* = pp_1^1 + qp_2^1;\ q^* = pp_1^2 + qp_2^2;\quad \varepsilon = p_1^1 - p_2^1 = p_2^2 - p_1^2 \text{ и } \delta' = \delta\varepsilon^2 \frac{pq}{p^* q^*}. \quad (2.100)$$

Из последнего соотношения формулы (2.100) следует, что основное влияние на приближение слабо зависимой СФ к бернуллиевской $\left(\delta' = \delta\varepsilon^2 \frac{pq}{p^* q^*} \to 0\right)$ оказывает СЗ (ε^2), а не СА (δ).

Получим соотношения между средними и дисперсиями частот СА и СФ в однородном бинарном случае.

Здесь векторы частот $\overline{m} = (m_1, s - m_1)$ и $\overline{m^*} = (m_1^*, s - m_1^*)$ определяются первыми компонентами, которые мы будем обозначать $m_1 = m$ и $m_1^* = m^*$.

Из соотношений (2.94) и (2.95) после несложных преобразований получим в нашем случае:

$$\Delta^* = \varepsilon\Delta; \quad (2.101)$$

$$\Delta^* = 2(0{,}5 - p_1^1)\varepsilon\Delta + \varepsilon^2(\Delta + \Delta), \quad (2.102)$$

где

$$\Delta = \mathrm{E}m - sp, \quad \Delta = \mathrm{D}m - spq;$$

$$\Delta^* = \mathrm{E}m^* - sp^*, \quad \Delta^* = \mathrm{D}m^* - sp^*q^*.$$

Из соотношений (2.101) и (2.102) следует, что и в общем случае параметры распределения частот СФ ближе к биномиальным, чем соответствующие параметры СА.

Глава 3.

СТАТИСТИЧЕСКИЕ РЕШЕНИЯ

§ 3.1. Вводные замечания

3.1.1. *Решающие процедуры.* Прямая задача теории стохастических функций состоит в изучении вероятностных свойств СФ при заданных СА и СЗ. Обратная задача состоит в изучении вероятностных свойств СА при заданных СФ и СЗ (обратные СФ). Обе эти задачи, как и другие задачи гл. 2, носят описательный характер и качественно не отличаются от физических задач.

Качественно отличными от перечисленных задач являются задачи вынесения в некотором смысле оптимальных решений о значении СА x, если известны значение СФ y и СЗ, определяемая условной вероятностью $p_x(y)$. Задачи такого рода, впервые возникшие в статистике, в последнее время в более общей и разветвленной форме рассматриваются теорией решений, теорией игр, динамическим программированием и др. Эти новые кибернетические дисциплины изучают вопросы оптимальных решений, стратегий, поведения и других целенаправленных действий в бесконфликтных (статистика) и конфликтных (теория игр) ситуациях.

В этой главе выделяется весьма узкий круг таких задач, решение которых используется в последующем изложении.

3.1.2. *Постановка задачи выбора между гипотезами.* Дальнейшее изложение этой главы использует лишь первоначальные идеи указанного направления, связанные с оптимальным выбором между гипотезами. Для постановки задачи можно воспользоваться понятием СЗ, вводя статистическую терминологию. Именно здесь рассматривается несколько более общий случай, когда x и y определены не обязательно на одном и том же множестве $\sigma = (1, \ldots, s)$. В самом деле пусть x определено на $\sigma' = (1, \ldots, t', \ldots, s')$, а y определено на $\sigma = (1, \ldots, t, \ldots, s)$. В статистике $x \in R'$ называется s'-мерным параметром, R' — пространством параметра, $y \in R$ называется выборкой объема s, а R называется выборочным пространством. Условную вероятность $p_x(y)$ называют функцией правдоподобия.

Задача оптимального выбора между гипотезами ставится следующим образом. Выдвигается M гипотез о значении параметра x известной функции правдоподобия $p_x(y)$

$$H_m = H_m(x = x_m)\,(m = \overline{0,\ M-1},\ x_m \neq x_{m'},\ \text{если}\ m \neq m').$$

Пусть имеется выборка y объема s, полученная при каком-то одном нам неизвестном значении $x = x_m$. Если по выборке y выносить решение о значении x_m, при котором она была получена (то есть принять одну из M гипотез), то неизбежны ошибки. Можно принять гипотезу $H_{m'}$, когда на самом деле имела место гипотеза $H_m\,(m \neq m')$. Вероятности такого рода ошибок будем обозначать $\gamma_{m,\,m'}$. Вместе с вероятностями $\gamma_{m,\,m}$ правильных

решений они составляют матрицу $\gamma = \|\gamma_{m,m'}\|$ с суммой элементов по строкам, равной единице.

Геометрически задача сводится к разбиению выборочного пространства R на M заполняющих его непересекающихся областей $\mathcal{E}_m$ ($m = \overline{0, M-1}$), так что если $y \in \mathcal{E}_m$, то принимается гипотеза H_m.

Оптимальной называют такую процедуру разбиения R на M областей $\mathcal{E}_m$, при которой, фиксируя матрицу γ, мы можем провести соответствующий выбор между M гипотезами по выборке y минимального объема s (классическое решение).

Другое решение той же задачи (последовательное решение) состоит в последовательном анализе выборки $y = y_s$ нарастающего объема s. Для каждого s производится разбиение выборочного пространства R_s на $M+1$ заполняющих его и непересекающихся областей $\mathcal{E}_m^{(s)}$ ($m = \overline{0, M}$). При этом, если $y_s \in \mathcal{E}_M^{(s)}$ попадает в индиферентную область, то испытание продолжается, а падение $y_s \in \mathcal{E}_m^{(s)}$ в области с индексом $m < M$ приводит к окончанию процедуры и принятию соответствующей гипотезы H_m. Ясно, что здесь объем выборки s, при котором происходит вынесение окончательного решения, является случайной величиной. Оптимальной называется такая последовательная процедура, при которой для фиксированной матрицы $\gamma = \|\gamma_{m,m'}\|$ обращается в минимум среднее значение Es. В классической постановке сформулированная задача решена без предположения о структуре функции правдоподобия для $M=2$ в [31] и для произвольного M в [32]. В последовательной постановке эта задача решена в [33] лишь при $M = 2$ в предположении СЗ с независимыми переходами.

3.1.3. *Содержание главы.* Излагаемые в этой главе результаты известны [33, 34] для широкого в том числе и непрерывного класса функций правдоподобия $p_x(y)$. Однако далее рассмотрения ограничиваются общим случаем дискретных СЗ с независимыми переходами. Важно отметить, что для описания существенно статистических ситуаций здесь достаточно использовать лишь две специальные функции гл. 1, имеющие комбинаторное происхождение: h-функцию и k-функцию и их асимптотику.

В § 3.2 описаны оптимальные классическая и последовательная процедуры выбора между двумя гипотезами и связь между параметрами этих процедур.

В § 3.3 рассмотрен важнейший для всей главы случай близких гипотез и оценка эффективности последовательной процедуры по сравнению с классической.

В § 3.4 приведены результат асимптотической оценки распределения случайного объема выборки последовательной процедуры, а также рассмотрены практически важные вопросы усечения последовательной процедуры. Так как большинство результатов этой главы приводятся в [34], где имеются ссылки на более ранние публикации, то в этой главе на них не делается специальных ссылок.

§ 3.2. Оптимальный выбор между двумя гипотезами

3.2.1. *Описание оптимальных процедур выбора между двумя гипотезами.* Ограничимся рассмотрением случая оптимального выбора между двумя гипотезами, так как лишь этот случай используется в дальнейшем изложении. Точнее, далее случай выбора между M гипотезами сводится к случаю M выборов между двумя гипотезами (см. гл. 5). В связи с этим упростим обозначения (другая параметризация см. п. 5.2.1).

Пусть задано дискретное распределение $\mathcal{P}(B_\rho)=p_\rho$

$$\bar{p}=(p_1,\ldots,p_\rho,\ldots,p_r)\left(\sum_{\rho=1}^{r}p_\rho=1\right),$$

о котором выдвигаются две гипотезы: $H_0=H_0(\bar{p}=\bar{p_0})$ и $H_1=H_1(\bar{p}=\bar{p_1}\neq\bar{p_0})$, где $\bar{p_0}=(p_1^{(0)},\ldots,p_\rho^{(0)},\ldots,p_r^{(0)})$ и $\bar{p_1}=(p_1^{(1)},\ldots,p_\rho^{(1)},\ldots,p_r^{(1)})$ два различных значения $\bar{p}$.

Выбор между гипотезами H_0 и H_1 производится по выборке

$$y=(B_{\rho_1},B_{\rho_2},\ldots,B_{\rho_t},\ldots,B_{\rho_s})$$

конечного объема s, состоящей из r различных выборочных значений $B_\rho\,(\rho=\overline{1,r})$ произвольной природы. В рассматриваемом случае матрица γ имеет вид

$$\gamma=\begin{pmatrix}1-\gamma & \gamma\\ 1-\delta & \delta\end{pmatrix}$$

и определяется двумя вероятностями γ и δ принятия гипотезы H_1, когда верны гипотезы H_0 и H_1 соответственно. Величины γ и $1-\delta$ называются в статистике вероятностями ошибок первого и второго рода соответственно.

Оптимальный выбор между двумя гипотезами H_0 и H_1 здесь, согласно критерию Неймана и Пирсона, основывается на однопороговом анализе одномерной статистики — отношении правдоподобия.

В самом деле, в рассматриваемом случае независимых B_ρ функция правдоподобия имеет вид

$$p_0(y)=\prod_{t=1}^{s}p_t^{(0)}=\prod_{\rho=1}^{v}(p_\rho^{(0)})^{m_\rho},$$

если верна гипотеза H_0, и

$$p_1(y)=\prod_{t=1}^{s}p_{\rho_t}^{(1)}=\prod_{\rho=1}^{r}(p_\rho^{(1)})^{m_\rho},$$

если верна гипотеза H_1, где m_ρ — частота выборочного значения B_ρ (в выборке y).

Далее составляется так называемое отношение правдоподобия

$$L_s=L_s(y)=p_1(y)/p_0(y)=\prod_{t=1}^{s}(p_{\rho_t}^{(1)}/p_{\rho_t}^{(0)})=\prod_{\rho=1}^{b}(p_\rho^{(1)}/p_\rho^{(0)})^{m_\rho}\qquad(3.1)$$

и назначается некоторое число (порог) $C=C(\gamma,\delta,s)$, так что если $L_s<C$, то принимается гипотеза H_0, а если $L_s\geqslant C$, то принимается гипотеза H_1*.

Геометрически это эквивалентно разбиению пространства выборок на две непересекающиеся области $\mathcal{E}_0$ и $\mathcal{E}_1$ ($\mathcal{E}_0\cup\mathcal{E}_1=R$, $\mathcal{E}_0\cap\mathcal{E}_1=\varnothing$) таких, что если $y\in\mathcal{E}_0$, то принимается гипотеза H_0 и если $y\in\mathcal{E}_1$, то принимается гипотеза H_1.

---

* Если $p_0(y)=p_1(y)=0$, то полагаем $L_s=1$.

Оптимальность указанной процедуры состоит в том, что при фиксации двух параметров из трех γ, $1-\delta$ и s третий не больше соответствующего параметра любой другой процедуры выбора между двумя гипотезами, основанной на заранее назначенном объеме s выборки y. Последовательная процедура выбора между двумя гипотезами является естественным обобщением классической. Она состоит в двухпороговом анализе той же одномерной статистики (функции выборочных значений) — отношении правдоподобия. Однако здесь объем выборки заранее не фиксируется. Именно, последовательно для каждого $s = 1, 2, \ldots$, отношение правдоподобия L_s сравнивается с двумя порогами $A = A(\gamma, \delta)$ и $B = B(\gamma, \delta)$. Если $B \geqslant L_s$, то принимается гипотеза H_0. Если $L_s \geqslant A$, то принимается гипотеза H_1. Наконец, если $B < L_s < A$, то никакого решения не выносится, к выборке y объема s добавляется $s+1$ — е выборочное значение и аналогично анализируется получившаяся выборка объема $s+1$.

Геометрически это эквивалентно разбиению выборочного пространства R_s для каждого $s = 1, 2, 3, \ldots$ на три непересекающиеся области $\mathscr{E}_0^{(s)}$, $\mathscr{E}_*^{(s)}$ и $\mathscr{E}_1^{(s)}$ ($\mathscr{E}_0^{(s)} \cup \mathscr{E}_*^{(s)} \cup \mathscr{E}_1^{(s)} = R_s$, $\mathscr{E}_0^{(s)} \cap \mathscr{E}_*^{(s)} = \mathscr{E}_0^{(s)} \cap \mathscr{E}_1^{(s)} = \mathscr{E}_*^{(s)} \cap \mathscr{E}_1^{(s)} = \varnothing$), таких, что если $y \in \mathscr{E}_0^{(s)}$, то принимается гипотеза H_0, если $y \in \mathscr{E}_0^{(s)}$, то принимается гипотеза H_1 и если $y \in \mathscr{E}_*^{(s)}$, то никакого решения не выносится и испытания продолжаются. Ясно, что в рассматриваемой ситуации объем s выборки, при котором принимается окончательное решение, является случайной величиной. В зависимости от того, какая из двух гипотез (H_0 или H_1) имеет место, s имеет среднее значение $\mathrm{E}_0 s$ или $\mathrm{E}_1 s$ соответственно.

Оптимальность указанной последовательной процедуры состоит в том, что при фиксации двух параметров из трех γ, $1-\delta$, $\mathrm{E}_i s$ третий не больше соответствующего параметра для любой другой последовательной процедуры. И это имеет место как для случая $i = 0$, так и для случая $i = 1$. Классическая процедура, основанная на заранее фиксированном объеме выборки, является частным случаем рассмотренной последовательной процедуры. В самом деле, для этого достаточно положить

$$\mathscr{E}_*^{(t)} = \begin{cases} R_t, & \text{для } 1 \leqslant t < s, \\ \varnothing, & \text{для } t = s. \end{cases}$$

При этом происходит вырождение средних значений $\mathrm{E}_0 s = \mathrm{E}_1 s = s$. Поэтому можно ожидать, что из-за оптимальных свойств последовательной процедуры при одних и тех же γ и δ объем s выборки классической процедуры может оказаться меньше средних $\mathrm{E}_0 s$ и $\mathrm{E}_1 s$. Этот факт имеет место в действительности (см. § 3.4).

3.2.2. *Связь между параметрами оптимальных процедур выбора между двумя гипотезами.* Для получения связей между параметрами рассмотренных процедур удобно прологарифмировать соотношение (3.1) что, очевидно, не изменит сути дела. В этом случае классическая процедура будет основана на однопороговом (с порогом $\ln C$) анализе суммы

$$l_s = \sum_{t=1}^{s} z_t = \sum_{\rho=1}^{r} m_\rho \ln \frac{p_\rho^{(1)}}{p_\rho^{(0)}},$$

где независимые случайные слагаемые z_t одинаково распределены с

$$z = \begin{Bmatrix} \ln (p_\rho^{(1)}/p_\rho^{(0)}) \\ p_\rho^{(i)} \end{Bmatrix} \quad (i = 0, 1)$$

и

$$\overline{m} = \begin{pmatrix} \overline{m}_* \\ C_s^{\overline{m}_*} & p_i^{\overline{m}_*} \end{pmatrix}$$

частоты выборочных значений соответственно с двумя возможными распределениями в зависимости от того, какая из двух гипотез H_0 или H_1 имеет место. Последовательная процедура основывается на двухпороговом анализе (с порогами $\ln A$ и $\ln B$) той же суммы.

Из самого определения вероятностей γ и δ имеем в классическом случае

$$1-\gamma = \mathcal{P}(l_s \leqslant \ln C/H_0) = \sum_{\sum_{\rho=1}^{r} m_\rho \ln \frac{p_\rho^{(1)}}{p_\rho^{(0)}} \leqslant \ln C} C_s^{\overline{m}} p_0^{\overline{m}}, \tag{3.2}$$

$$1-\delta = \mathcal{P}(l_s \leqslant \ln C/H_1) = \sum_{\sum_{\rho=1}^{r} m_\rho \ln \frac{p_\rho^{(1)}}{p_\rho^{(0)}} \leqslant \ln C} C_s^{\overline{m}} p_1^{\overline{m}}. \tag{3.3}$$

Оценим суммы (3.2) и (3.3) в асимптотическом случае, при $s \to \infty$, как это делалось в гл. 1. В самом деле, для этого достаточно положить

$$d = \frac{1}{s}\ln C,\ d_\rho = \ln(p_\rho^{(1)}/p_\rho^{(0)}),\ \overline{d} = (d_\rho),\ \mu_\rho = \frac{1}{s} m_\rho\ (\rho = \overline{1, r}),$$

$$\gamma_i(\lambda) = \ln \sum_{\rho=1}^{r} p_\rho^{(i)} e^{\lambda \ln\left(p_\rho^{(1)} \mid p_\rho^{(0)}\right)}\ (i = 0,1).$$

Откуда

$$\gamma_i'(0) = \mathrm{E}_i z = \sum_{\rho=1}^{r} p_\rho^{(i)} \ln \frac{p_\rho^{(1)}}{p_\rho^{(0)}} \quad \text{и} \quad \gamma_i''(0) = \mathrm{D}_i z =$$

$$= \sum_{\rho=1}^{r} p_\rho^{(i)} \left(\ln \frac{p_\rho^{(1)}}{p_\rho^{(0)}}\right)^2 - (E_i z)^2\ (i = 0,1). \tag{3.4}$$

Выберем $d = \frac{1}{s}\ln C$ так, чтобы

$$E_0 z = -E(\overline{p}_0, \overline{p}_1) < d < \mathrm{E}_1 z = E(\overline{p}_1, \overline{p}_0), \tag{3.5}$$

где в (3.5) обозначено $E(\overline{u}, \overline{v}) = h(\overline{u}, \overline{v}) - h(\overline{u}) \geqslant 0$. Тогда, согласно (1.71), имеем для γ и δ, определяемых соотношениями (3.2) и (3.3), следующие асимптотические выражения:

$$\gamma \approx e^{-sk_{\overline{p}_0 d}^{-1}(d - \mathrm{E}_0 z)} \quad \text{и} \quad \delta \approx 1 - e^{-sk_{\overline{p}_1 d}^{-1}(d - \mathrm{E}_1 z)}. \tag{3.6}$$

При фиксированных $\overline{p}_0$ и $\overline{p}_1$ соотношения (3.6) накладывают две связи на четыре параметра γ, δ, d и s. Эти соотношения можно переписать в виде

$$\left.\begin{aligned} d &= \mathrm{E}_0 z + k_{\overline{p}_0 d}^{-1}\left(\frac{1}{s}\ln\frac{1}{\gamma}\right) = \mathrm{E}_1 z - k_{\overline{p}_1 d}^{-1}\left(\frac{1}{s}\ln\frac{1}{1-\delta}\right) \\ k_{\overline{p}_0 d}^{-1}&\left(\frac{1}{s}\ln\frac{1}{\gamma}\right) + k_{\overline{p}_1 d}^{-1}\left(\frac{1}{s}\ln\frac{1}{1-\delta}\right) = \mathrm{E}_1 z - \mathrm{E}_0 z \end{aligned}\right\}. \tag{3.6'}$$

Заметим, что при фиксированных $\bar{p}_0$, $\bar{p}_1$, $\bar{d}$, γ, δ и $s \to \infty$ аргументы под знаком функции k^{-1} ( ) стремятся к нулю и в этом случае можно воспользоваться приближенным соотношением (1.84), при $\eta \to 0$, откуда $\varepsilon \to 0$

$$\eta = k_{\bar{p}\bar{d}}(\varepsilon) = \frac{\varepsilon^2}{2Dz} + O(\varepsilon^3) \approx \frac{\varepsilon^2}{2Dz}, \quad \text{откуда} \quad \varepsilon = k^{-1}_{\bar{p}\bar{d}}(\eta) \approx \sqrt{2Dz\eta}.$$

Поэтому из (3.6′) имеем

$$d \approx E_0 z + \sqrt{2\ln\frac{1}{\gamma}}\sqrt{D_0 z/s} \approx E_1 z - \sqrt{2\ln\frac{1}{1-\delta}}\sqrt{D_1 z/s}; \tag{3.7}$$

$$s \approx \frac{\left(\sqrt{2\ln\frac{1}{\gamma}}\sqrt{D_0 z} + \sqrt{2\ln\frac{1}{1-\delta}}\sqrt{D_1 z}\right)^2}{(E_1 z - E_0 z)^2}.$$

Аналогичные соотношения можно вывести на основе центральной предельной теоремы и асимптотического выражения p-квантили нормального распределения $|u_p| \approx \sqrt{2\ln\frac{1}{p}}$, при $p \to 0$.

Для получения связи между параметрами последовательной процедуры воспользуемся легко выводимыми оценками (в нужную сторону)

$$\frac{1-\delta}{1-\gamma} \leqslant B < 1 < A \leqslant \frac{\delta}{\gamma},$$

которые, как показал А. Вальд, достаточно хороши, так что можно положить

$$B = \frac{1-\delta}{1-\gamma} \quad \text{и} \quad A = \frac{\delta}{\gamma}.$$

Поэтому из-за логарифмирования соответствующие нижний и верхний пороги имеют вид

$$\ln B = \ln\frac{1-\delta}{1-\gamma} < 0 < \ln A = \ln\frac{\delta}{\gamma}. \tag{3.8}$$

Далее из соотношения (1.47) в одномерном случае получим для производящей l_s

$$g_{l_s}(u) = g_s(g_z(u)), \tag{3.9}$$

где индексы у производящих указывают на соответствующие случайные величины и значения z могут и не быть целочисленными. Далее, дифференцируя обе части соотношения (3.9) и положив $u = 1$, получим известное тождество Вальда

$$El_s = Es \cdot Ez. \tag{3.10}$$

Но из самого определения окончания последовательной процедуры имеем для El_s

$$El_s = \begin{cases} E_0 l_s = (1-\gamma)\ln\frac{1-\delta}{1-\gamma} + \gamma\ln\frac{\delta}{\gamma} = h(\gamma) - h(\gamma, \delta) = E(\gamma, \delta), & (3.11) \\ E_1 l_s = (1-\delta)\ln\frac{1-\delta}{1-\gamma} + \delta\ln\frac{\delta}{\gamma} = h(\delta, \gamma) - h(\delta) = E(\delta, \gamma). & (3.12) \end{cases}$$

Поэтому, используя соотношения (3.11), (3.12) и (3.5), будем иметь из (3.10)

$$\mathrm{E}s = \begin{cases} \mathrm{E}_0 s = \dfrac{E_0 l_s}{\mathrm{E}_0 z} = \dfrac{E(\gamma, \delta)}{E(\bar{p}_0, \bar{p}_1)} \\ \mathrm{E}_1 s = \dfrac{\mathrm{E}_1 l_s}{\mathrm{E}_1 z} = \dfrac{E(\delta, \gamma)}{E(\bar{p}_1, \bar{p}_0)}. \end{cases} \tag{3.13}$$

Соотношения (3.8) и (3.13) при фиксированных $\bar{p}_0$ и $\bar{p}_1$ накладывают связи на параметры γ, δ, $\ln A$, $\ln B$, $\mathrm{E}_0 s$ и $\mathrm{E}_1 s$. Важная положительная особенность последних соотношений состоит в том, что в отличие от порога $\ln C$ классической процедуры пороги $\ln B$ и $\ln A$ не зависят от вида распределений $p_i(y)$ $(i = 0,1)$ и зависят лишь от вероятностей γ и δ.

§ 3.3. Случай близких гипотез

3.3.1. *Упрощение общих соотношений.* В практически интересном случае близких гипотез происходит заметное упрощение соотношений предыдущего параграфа. Будем говорить, что имеет место случай близких гипотез, если при фиксированном $\bar{p}_0$, $\bar{p}_1 \to \bar{p}_1$, что означает

$$\max_{1 \leqslant \rho \leqslant r} |p_\rho^{(1)} - p_\rho^{(0)}| = \Delta \to 0.$$

Используя разложение логарифма в ряд, легко получить следующие соотношения Бартлетта [34]

$$-2\mathrm{E}_0 z \approx 2\mathrm{E}_1 z \approx \mathrm{D}_0 z \approx \mathrm{D}_1 z \approx \chi^2 = \sum_{\rho=1}^{r} \frac{(p_\rho^{(1)} - p_\rho^{(0)})^2}{p_\rho^{(0)}} = O(\Delta^2) > 0, \tag{3.14}$$

верные с точностью до $O(\Delta^3)$.

В этом случае из (3.14) следует, что

$$-(d - \mathrm{E}_1 z) \approx (d - \mathrm{E}_0 z) \approx O(\Delta^2) > 0.$$

Поэтому соотношения (3.7) и (3.13) упростятся

$$s \approx 2\left(\sqrt{\ln\frac{1}{\gamma}} + \sqrt{\ln\frac{1}{1-\delta}}\right)^2 \Big/ \chi^2; \tag{3.15}$$

$$\mathrm{E}s = \begin{cases} \mathrm{E}_0 s \approx 2\,|E(\gamma, \delta)|/\chi^2, \\ \mathrm{E}_1 s \approx 2E(\delta, \gamma)/\chi^2. \end{cases} \tag{3.16}$$

Таким образом, в случае близких гипотез объем выборки классической процедуры и средние объемы выборки последовательной процедуры имеют идентичную структуру и разнятся лишь числителями, зависящими только от вероятностей ошибок γ и $1-\delta$. В знаменателях у них стоит одна и та же величина χ^2, определяемая соотношением (3.14). Это обстоятельство можно использовать для так называемой оптимальной дискретизации, при переходе от непрерывных распределений $p_x(y)$ к дискретным [34].

В рассматриваемом случае дискретных распределений с r различными значениями A_ρ $(\rho = \overline{1, r})$ это эквивалентно оптимальной группировке значений, приводящей к $r' < r$ различным «группированным» значениям. При этом оптимальной называется, как и в непрерывном случае, такая группировка, которая обращает в максимум χ^2 (соответственно обращает в минимум s и $\mathrm{E}s$). Можно предполагать, что оптимальная бинарная группировка

($r = 2$), так же как и в непрерывном случае, не должна приводить к существенному увеличению s и Es (в непрерывном случае для широкого класса распределений s и Es при тех же γ и δ увеличиваются лишь в полтора раза). Какие-либо общие соображения относительно структуры оптимальной бинарной группировки в дискретном случае неизвестны. Простая структура выражений s и Es в случае близких гипотез позволяет легко оценить выигрыш в числе наблюдений при использовании последовательной процедуры вместо классической.

3.3.2. *Эффективность последовательной процедуры.* Естественно определять эффективность последовательной процедуры по сравнению с классической отношением $e'(\gamma, \delta) = \frac{s - Es}{s} = 1 - \frac{Es}{s} = 1 - e(\gamma, \delta)$, которое в случае близких гипотез зависит лишь от γ и δ.

В самом деле, из (3.15) и (3.16) имеем

$$e(\gamma, \delta) = \begin{cases} e_0(\gamma, \delta) \approx \dfrac{|E(\gamma, \delta)|}{\left(\sqrt{\ln\frac{1}{\gamma}} + \sqrt{\ln\frac{1}{1-\delta}}\right)^2}, & \text{если верна гипотеза } H_0, \\ e_1(\gamma, \delta) \approx \dfrac{E(\delta, \gamma)}{\left(\sqrt{\ln\frac{1}{\gamma}} + \sqrt{\ln\frac{1}{1-\delta}}\right)^2}, & \text{если верна гипотеза } H_1. \end{cases} \tag{3.17}$$

Рассмотрим частный случай $\delta = 1 - \gamma$; тогда

$$-E(\gamma, 1-\gamma) = E(1-\gamma, \gamma) = h(1-\gamma, \gamma) - h(1-\gamma) = \\ = 2(0{,}5 - \gamma)\ln\frac{1-\gamma}{\gamma} \geqslant 0.$$

Если $\gamma \to 0$, то

$$-E(\gamma, 1-\gamma) = E(1-\gamma, \gamma) = \ln\frac{1}{\gamma} + O\left(\gamma \ln\frac{1}{\gamma}\right)$$

и из (3.17) получим результат С. А. Айвазяна

$$\lim_{\gamma \to 0} e(\gamma, 1-\gamma) = 0{,}25,$$

который показал, что $e(\gamma, 1-\gamma)$, при $\gamma \to 0$, монотонно убывает чрезвычайно медленно.

Рассмотрим наиболее интересный случай, когда имеет место, вообще говоря, разный порядок малости γ и $1 - \delta$ [34].

Из (3.17) имеем

$$e(\gamma, \delta) = \begin{cases} e_0(\gamma, \delta) \approx \left(1 + \sqrt{\dfrac{\ln\gamma}{\ln(1-\delta)}}\right)^{-2}, & \text{если верна гипотеза } H_0, \\ e_1(\gamma, \delta) \approx \left(1 + \sqrt{\dfrac{\ln(1-\delta)}{\ln\gamma}}\right)^{-2}, & \text{если верна гипотеза } H_1. \end{cases} \tag{3.18}$$

Пусть $\gamma = o(1 - \delta)$, тогда из (3.18) имеем

$$e(\gamma, \delta) = \begin{cases} e_0(\gamma, \delta) \approx \ln(1-\delta)/\ln\gamma, & \text{если верна гипотеза } H_0, \\ e_1(\gamma, \delta) \approx 1, & \text{если верна гипотеза } H_1. \end{cases} \tag{3.19}$$

Пусть $1 - \delta = o(\gamma)$, тогда из (3.18) имеем

$$e(\gamma, \delta) = \begin{cases} e_0(\gamma, \delta) \approx 1, & \text{если верна гипотеза } H_0, \\ e_1(\gamma, \delta) \approx \ln\gamma/\ln(1-\delta), & \text{если верна гипотеза } H_1. \end{cases} \tag{3.20}$$

Коротко результаты (3.19) и (3.20) можно записать так

$$e(\gamma, \delta) \approx \begin{cases} \dfrac{\min\left(\ln\dfrac{1}{\gamma}, \ln\dfrac{1}{1-\delta}\right)}{\max\left(\ln\dfrac{1}{\gamma}, \ln\dfrac{1}{1-\delta}\right)}, & \text{при } \gamma = o(1-\delta) \text{ и } H_0, \\ & \text{либо } 1-\delta = o(\gamma) \text{ и } H_1, \end{cases} \tag{3.21}$$

$$\begin{cases} 1 \ldots\ldots\ldots\ldots, & \text{при } 1-\delta = o(\gamma) \text{ и } H_0, \\ & \text{либо } \gamma = o(1-\delta) \text{ и } H_1. \end{cases} \tag{3.22}$$

Таким образом, применение последовательной процедуры при одном и том же порядке малости γ и $1-\delta$ дает экономию в среднем числе наблюдений по сравнению с классической процедурой не более чем в четыре раза. Однако при выполнении условий (3.21) эффективность последовательной процедуры по сравнению с классической может неограниченно приближаться к единице. При выполнении условий (3.22) эффективность стремится к нулю (нивелируется).

В ряде практических ситуаций [34], например тех, что описываются в гл. 4, имеют место условия (3.21) и поэтому использование последовательных процедур особенно целесообразно.

§ 3.4. Распределение объема выборки последовательной процедуры

3.4.1. *Распределение и дисперсия объема выборки.* Выигрыш в среднем объеме выборки при использовании последовательной процедуры по сравнению с классической не гарантирует того же при проведении конкретной последовательной процедуры. В самом деле, при наличии большой дисперсии объема выборки последовательной процедуры он может с большой вероятностью превзойти среднее значение и даже фиксированный объем классической процедуры. Однако можно показать (см. далее), что указанная дисперсия не велика по сравнению со средним. Ограничимся рассмотрением лишь случая наиболее выгодного использования последовательной процедуры при выполнении соотношения (3.21). В этом случае один из порогов уходит в бесконечность, а знак второго порога совпадает со знаком среднего значения $\mathrm{E}_i z$.

Рассмотрим для определенности случай $\mathrm{E}_0 z < 0$, когда $\gamma = o(1-\delta)$. Тогда в случае близких гипотез имеем $-2\mathrm{E}_0 z \approx \mathrm{D}_0 z$. Найдем распределение объема выборки s, при котором происходит окончание последовательной процедуры, то есть первое пересечение логарифмом отношения правдоподобия l_s нижнего порога $\ln B$. Для этого рассмотрим бинарную случайную величину

$$x = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \text{ где } p = 1-q = -\mathrm{E}_0 z/(2-\mathrm{E}_0 z);$$

тогда случайная величина

$$x' = \frac{2}{1-p}x = \begin{pmatrix} 0 & \dfrac{2}{1-p} \\ q & p \end{pmatrix}$$

имеет среднее и дисперсию

$$\mathrm{E}x' = \frac{2}{1-p}\mathrm{E}x = \frac{2}{1-p}p = -\mathrm{E}_0 z \text{ и } \mathrm{D}x' = \left(\frac{2}{1-p}\right)^2 \mathrm{D}x = \left(\frac{2}{1-p}\right)^2 pq =$$

$$= \frac{4}{1-p}p = -2\mathrm{E}_0 z \approx \mathrm{D}_0 z. \tag{3.23}$$

Итак, случайные величины z и $-x' = -\frac{2}{1-p}x$ имеют одни и те же средние и дисперсии. Можно показать [34], что в рассматриваемом случае близких гипотез случайная величина l_s асимптотически имеет следующее представление

$$l_s = \sum_{i=1}^{s} z_i = -\frac{2}{1-p}\sum_{i=1}^{s} x_i,$$

где z_i и x_i — независимые одинаково распределенные с z и x соответственно случайные величины.

Доказательство этого факта основывается на том, что, группируя слагаемые z_i и x_i в достаточно большие группы, мы придем в силу центральной предельной теоремы к одинаково распределенным в обеих суммах нормальным случайным величинам (из-за совпадения их средних и дисперсий). Поэтому вероятность первого достижения l_s нижнего порога $\ln B$ асимптотически совпадает с вероятностью первого равенства суммы $\sum_{i=1}^{s} x_i$ величины

$$m = \left[\ln B/ - \frac{2}{1-p}\right]. \tag{3.24}$$

Последняя же вероятность, как легко видеть, совпадает с распределением Паскаля (см. (1.37)).

$Q_m(s) = C_{s-1}^{m-1} p^m q^{s-m}$, имеющего производящую функцию $g_m(u) = \left(\frac{pu}{1-qu}\right)^m$. Дифференцируя $g_m(u)$, находим

$$\mathrm{E}\,s = g_m'(1) = \frac{m}{p} = \frac{m}{\mathrm{E}\,x}, \quad \mathrm{D}\,s = g_m''(1) - (g_m'(1))^2 + g_m'(1) =$$

$$= m\frac{q}{p^2} = m\frac{\mathrm{D}\,x}{(\mathrm{E}\,x)^3}.$$

Отсюда, учтя выражения (3.23) и (3.24) m, $\mathrm{E}x$ и $\mathrm{D}x$ через $\ln B$, $\mathrm{E}_0 z$ и $\mathrm{D}_0 z$, получим

$$\mathrm{E}\,s = \frac{\ln B}{\mathrm{E}_0 z} \text{ и } \mathrm{D}\,s = |\ln B|\frac{\mathrm{D}_0 z}{|\mathrm{E}_0 z|^3} \approx |\ln B|\frac{2}{|\mathrm{E}_0 z|^2} = 2\,\mathrm{E}^2 s. \Big/ |\ln \mathrm{B}| \tag{3.25}$$

Введя нормировку $t = \frac{s}{\mathrm{E}\,s}$ и параметр [см. (1.90)]

$$c = m\frac{(\mathrm{E}\,x)^2}{\mathrm{D}\,x} = |\ln B|\frac{(\mathrm{E}_0 z)^2}{\mathrm{D}_0 z} = \frac{(\mathrm{E}\,s)^2}{\mathrm{D}\,s} = \frac{|\ln \mathrm{B}|}{2}, \tag{3.26}$$

получаем асимптотическое выражение для распределения объема выборки последовательной процедуры через распределение Вальда

$$\mathcal{P}(s < t\,\mathrm{E}\,s) \approx W_c(t) = \frac{\sqrt{c}}{\sqrt{2\pi}}\int_0^t v^{-\frac{3}{2}} e^{-\frac{c}{2}(v+v^{-1}-2)}\,dv. \tag{3.27}$$

Заметим, что в рассматриваемом случае близких гипотез из соотношения (3.25) следует, что $\mathrm{E}s$ и $\sqrt{\mathrm{D}s}$ имеют один и тот же порядок роста при $\mathrm{E}_0 z \to 0$. Но в этом случае при $|\ln \mathrm{B}| \to \infty$ согласно (3.26) параметр c

неограниченно растет, что приводит согласно (3.27) к вырождению распределения Вальда в функцию единичного скачка (плотность распределения Вальда вырождается в δ-функцию).

Таким образом, в случае близких гипотез значения объемов выборок конкретных последовательных процедур с большой вероятностью мало отклоняются от своего среднего значения и этим сохраняется эффективность последовательного анализа.

Аналогично можно рассмотреть случай $E_1 z > 0$, когда $1-\delta = o(\gamma)$. Для этого случая в полученных формулах надо заменить $E_0 z$ на $E_1 z$ и $\ln B$ на $\ln A$.

3.4.2. *Усечение.* Если последовательная процедура не может иметь объем выборки s, превосходящий некоторое значение s_0, то такая последовательная процедура называется усеченной. Если $s_0 \gg Es$, то в случае близких гипотез на основании результатов предыдущего пункта можно утверждать, что, как правило, будет происходить окончание последовательной процедуры до усечения на s_0. В частности это будет иметь место при усечении на соответствующем (для тех же γ и δ) классическом объеме выборки.

Часть II

ОПТИМАЛЬНОЕ КОДИРОВАНИЕ ДЛЯ КАНАЛОВ С ШУМАМИ

Глава 4.
ОСНОВНЫЕ ПОНЯТИЯ

§ 4.1. Общая схема передачи сигналов по каналу с шумами

4.1.1. *Вводные замечания.* Общая схема передачи сигналов по каналу с шумами (см. рис. 4.1) не следует из каких-либо математических теорий, являясь обобщением ряда конкретных схем передачи. Оптимальные задачи, связанные с этой схемой, впервые были поставлены в теории потенциальной помехоустойчивости В. А. Котельникова [1] (см. § 4.5) и носили вероятностный характер. Впоследствии, начиная с работы К. Шеннона [2], эта схема явилась основным объектом исследования так называемой теории информации. Последней были установлены возможности надежного различения предельно большого числа сигналов, передаваемых по каналу с шумами (см. § 4.5).

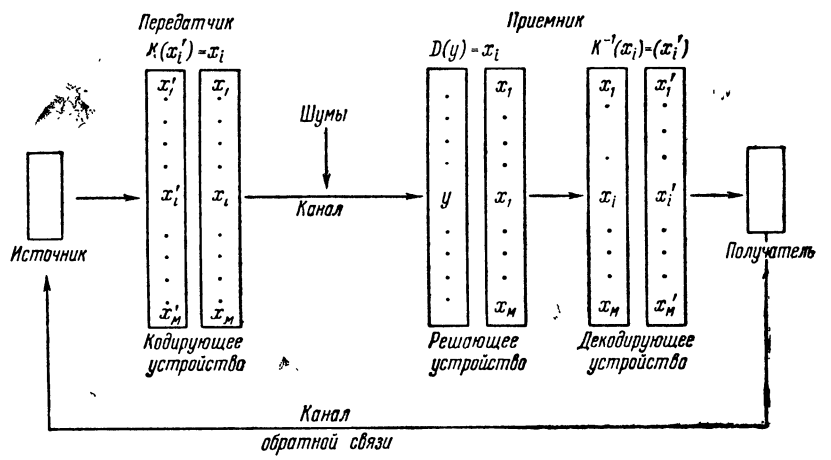


Рис. 41. Общая принципиальная схема передачи сигналов по каналу с шумами

Обе теории основывались на допущении о статистической природе сигналов и шумов. При этом использовались в конечном счете общие статистические закономерности. Именно, единственным радикальным средством борьбы с шумами при фиксированной интенсивности сигналов и шумов является накопление данных до вынесения окончательных решений. В другой терминологии эта тенденция современной техники связи определяется как прием сигнала в целом, в отличие от посимвольного приема, характерного для раннего развития техники связи. Технические трудности осуществления приема сигнала в целом

(необходимо иметь большой объем памяти) могут быть в ряде случаев преодолены использованием электронных машин дискретного действия. С последними связана также общая тенденция автоматизации современной техники связи, а также расширения области ее приложений для связи между людьми и машинами. При этом характерным является повсеместное использование дискретной техники, что в достаточной мере было подготовлено рядом теоретических исследований.

Речь идет об известной теореме В. А. Котельникова о дискретном задании непрерывных во времени сигналов с ограниченным спектром, а также о более поздних исследованиях [34], показавших, что даже наиболее грубая бинарная дискретизация сигналов по амплитуде приводит к сравнительно небольшим потерям. В этой главе подробно комментируется общая схема передачи сигналов по каналу с шумами, а также ряд связанных с ней понятий.

4.1.2. *Технические предпосылки.* Общая принципиальная схема передачи сигналов по каналу с шумами показана на рис. 4.1. Классическая схема Котельникова — Шеннона не предполагает наличия канала обратной связи, однако, как будет ясно из дальнейшего изложения, в общем случае его следует предусмотреть.

Прокомментируем схему, изображенную на рис. 4.1. Пусть источник вырабатывает M сообщений $\{x_i'\}$ $(i = \overline{1, M})$, заранее известных как на входе, так и на выходе канала. На входе канала они перекодируются кодирующим устройством в удобные для передачи входные сигналы $\{x_i\}$, также известные на входе и выходе канала. Другими словами, кодирующее устройство осуществляет взаимно-однозначное преобразование $K(x_i') = x_i$ $(i = \overline{1, M})$. Входящие сигналы передаются по каналу с шумами. В результате на выходе канала появляются искаженные шумами выходные сигналы. Решающее устройство производит идентификацию принятого выходного сигнала y с одним из подлежащих передаче входных сигналов. Другими словами, оно должно по некоторому правилу $D(y) = x_i$ по принятому выходному сигналу y указать, какой входной сигнал x_i передавался. Появляющийся таким образом на выходе решающего устройства входной сигнал декодируется, то есть подвергается обратному по отношению к кодированию преобразованию $K^{-1}(x_i) = x_i'$. В результате получается выходное сообщение.

В рассматриваемой схеме предполагается, что кодирующее, решающее и декодирующее устройства работают безошибочно. Для решающего устройства это означает точное соблюдение заложенных в него решающих правил с заранее предусмотренными вероятностями ошибок (см. гл. 3).

Таким образом, единственным источником ошибок в передаче является канал с шумами. Предполагается, что передача сигнала ведется во времени последовательно отдельными посылками (символами), из которых он состоит. Непосредственно под действием шумов в канале искажаются отдельные посылки. Упоминавшийся выше прием сигнала в целом состоит в вынесении решения не о каждой принятой посылке, а сразу о всех n принятых посылках, составляющих выходной сигнал. Но учитывая последовательную во времени передачу отдельных посылок, это приводит к необходимости ожидать n тактов передачи, при каждом из которых передается одна посылка, пока не накопится весь выходной сигнал y. Другими словами, по мере приема отдельных посылок решения об отдельных входных сигналах выносятся через каждые n посылок (предполагается, что решения выносятся моментально). Поэтому общий темп передачи не снижается, и образуется лишь общая задержка в n тактов передачи, связанная с вынесением решения о пер-

вом из передававшихся входных сигналов. Ясно, что эта задержка не накапливается.

Необходимым условием рассматриваемой схемы без обратной связи является синхронизация входа и выхода канала для установления начал и концов входных сигналов на выходе. Без этого невозможна работа решающего и декодирующего устройств. Если решающее устройство осуществляет идентификацию входного сигнала по выходному сигналу с заранее назначенной длиной n (классическое декодирование), то такая синхронизация возможна. Однако, если решающее устройство выносит решение по некоторой начальной части выходного сигнала случайной длины (последовательное декодирование [34]), то синхронизация входа и выхода возможна лишь при использовании канала обратной связи. Последний в этом случае подает сигнал с выхода на вход о конце приема входного сигнала, что является одновременно вызовом следующего входного сигнала. В дальнейшем изложении предполагается, что канал обратной связи — без шумов.

4.1.3. *Описание с помощью теории стохастических функций.* Как отмечалось в общем введении, теория стохастических функций в своем первоначальном развитии носила чисто математический характер без каких-либо приложений. Покажем, что теория стохастических функций может быть использована для математического описания общей схемы передачи сигналов по каналу с шумами.

В самом деле, в рассматриваемой схеме существенны лишь вероятностные свойства источника сообщений. Поэтому последний может быть математически описан стохастическим аргументом. Канал связи описывается стохастической зависимостью, а выходные сообщения — стохастической функцией. При этом из-за однозначной связи входные и выходные сигналы имеют ту же стохастическую структуру, что и входные и выходные сообщения соответственно. Поэтому при изучении вероятностных вопросов достаточно ограничиться рассмотрением дискретных сигналов и каналов с дискретным входом и выходом. Если, кроме того, ограничиться каналами с независимыми шумами, то такие входные сигналы, шумы и выходные сигналы могут быть описаны дискретными СА, СЗс независимыми переходами и СФ соответственно (подробнее эта терминология описана ниже, в п. 4.3.1).

Отдельные дискретные посылки будем называть символами. Входные символы будем обозначать $\alpha=1, 2, \ldots, a$; выходные — $\beta=1, 2, \ldots, b$; их совокупности будем называть входными A и выходными B алфавитом объема, a и b (вообще говоря, $a \neq b$) соответственно. Входным и выходным символам соответствуют элементарные значения СА и СФ, соответственно.

Совокупности n последовательных входных и выходных символов будем обозначать $x=(\alpha_1, \alpha_2, \ldots, \alpha_t, \ldots, \alpha_n)$ и $y=(\beta_1, \beta_2, \ldots, \beta_t, \ldots, \beta_n)$ и называть входными и выходными словами длины n соответственно. Им соответствуют значения СА и СФ соответственно. Совокупности всевозможных входных и выходных слов длины и будем обозначать R и R^* и называть пространствами входных и выходных слов длины n соответственно.

В рассматриваемом случае дискретного постоянного канала с независимыми шумами последние полностью описываются матрицей переходов отдельных символов

$$\mathbf{p} = \| p_\alpha^\beta \| \left(\sum_{\beta=1}^{b} p_\alpha^\beta = 1 \, (\alpha = \overline{1, a}); \quad \sum_{\alpha=1}^{b} p_\alpha^\beta > 0 \, (\beta = \overline{1, b}) \right),$$

где p_α^β означают условные вероятности появления выходных символов β, если имеют место входные символы α. Условия, наложенные на вероятности p_α^β, прокомментированы в п. 2.7.3. в терминах однородной СЗ с независимыми переходами, соответствующей рассматриваемому каналу.

Схема передачи сигналов по каналу с шумами требует самостоятельного рассмотрения вероятностной схемы стохастических функций. В самом деле, для последней характерны выделение и самостоятельное изучение СА и СЗ, от которых зависит СФ. В схеме передачи сигналов по каналу с шумами им соответствуют источник, канал с шумами и зависящие от последних сообщения на выходе канала. Но естественность независимого рассмотрения источника и канала не вызывает сомнения хотя бы потому, что по одному и тому же каналу можно передавать сообщения различных источников и сообщения одного и того же источника можно передавать по различным каналам.

§ 4.2. Источник сообщений

4.2.1. *Типы источников.* Для последующего изложения не существен семантический смысл сообщений источника. Важна лишь их вероятностная структура. Кроме того, можно ограничиться лишь дискретными или подвергнутыми дискретизации по времени и амплитуде физическими носителями сообщений (текстов, электрических и радиосигналов и т. д.). Как уже неоднократно подчеркивалось, достаточно мелкая дискретизация эквивалентна непрерывным рассмотрениям.

Заметим, что основной для последующего изложения тип источников дискретен по своей природе. Имеются в виду языковые источники сообщений, отображающие дискретную природу человеческого мышления.

Такие физические носители языковых источников, как тексты, остаются дискретными, однако такие, как акустические, электрические, радио и другие сигналы, непрерывны, но могут быть подвергнуты дискретизации.

Другой тип в основном непрерывных источников доставляет природа. Однако ее «сообщения» в основном телеметрического характера не столь разнообразны, как сообщения первого типа источников. Здесь безусловно речь идет не о беспорядочных замерах случайно меняющихся многомерных характеристик процессов, происходящих в природе. Подразумеваются замеры простых параметров известных законов природы: давления, температуры, влажности и т. д. Во временно́м отношении современные физические носители сообщений достаточно гибки.

В самом деле, мысль и речевое ее выражение примерно синхронны, в то время как тексты, магнитофонные записи и другие средства консервации информации являются способом длительного хранения сообщений до их передачи и восприятия и позволяют выбирать нужный темп их передачи. Поэтому в дальнейшем изложении, говоря о темпе передачи по каналу, мы будем подразумевать тот темп, который выбирается из каких-либо технических или других соображений и который вовсе не обязан совпадать с первоначальным темпом воспроизведения сообщений источника. В немалой степени темп передачи диктуется быстродействием передающей и приемной аппаратуры.

4.2.2. *Вероятностные свойства сообщений источника.* Изучение вероятностных свойств, в частности речевых источников, сообщений велось задолго до создания общей теории связи. Практически это отразилось, например, в различных размерах наборных касс, а также в различном числе точек и тире азбуки Морзе, выбираемых в зависимости от

повторяемости в данном языке соответствующих букв. Однако несомненно, что систематическое изучение и использование «статистики языка» для нужд связи было начато с момента появления работы Шеннона [2]. Ему же принадлежат два основных результата в этой области. Речь идет о теореме оптимального кодирования для каналов без шумов и об установлении вырожденного характера асимптотического распределения длинных сообщений.

Прокомментируем эти результаты. Начнем с первого. Короткие сообщения источника — отдельные символы источника, а также их сочетаний небольшой длины — существенно разнятся по их вероятностям. Это обстоятельство можно использовать при их передаче, передавая частые сообщения короткими сигналами, а редкие — длинными. Оптимальным кодом Шеннон называет неравномерный код, т. е. набор входных сигналов различной длины, который, перекодируя сообщения источника, приводит к минимальной средней длине входного слова. Оптимальным кодом оказывается известный код Фано — Шеннона [2], в котором длины входных слов выбираются пропорциональными логарифмам величин, обратных вероятностям соответствующих сообщений.

Второй, несколько неожиданный, результат состоит в следующем. Рассмотрим все $a^n = e^{n \ln a}$ априори возможные сообщения источника длины $n \to \infty$, состоящие из символов алфавита объема a. Все эти сообщения для широкого класса источников разбиваются на две группы сообщений. К первой группе относятся равновероятные высоковероятные сообщения в числе, имеющем порядок e^{nH} с вероятностью каждого e^{-nH}. Основная масса сообщений в числе $e^{n \ln a} - e^{nH} \approx e^{n \ln a}$ второй группы маловероятна (имеет общую вероятность, стремящуюся с ростом n к нулю).

Для случая бернулливского источника, соответствующего бернулливскому СА, этот факт был доказан в гл. 2. Однако, как уже отмечалось, он имеет место для широкого класса источников. При этом асимптотическое вырождение распределения длинных сообщений существенно облегчает изучение статистики источников. В самом деле, последняя определяется при фиксированных a и n лишь одним числом H $(0 \leqslant H \leqslant \ln a)$ *. Фактическое выделение высоковероятной группы сообщений источника при больших n является сложной технической проблемой (см. гл. 9). Примерами высоковероятных групп сообщений небольшой длины источников (языков) могут служить обычные словари.

В заключение остановимся на особенности «передачи» вероятностных свойств сообщений источника кодирующим эти сообщения входным словам длины n. Эта особенность состоит в том, что сообщения наделяют своими вероятностями соответствующие входные слова в целом. При этом отдельные символы, составляющие сообщения и слова, могут иметь различную вероятностную структуру.

§ 4.3. Канал с шумами

4.3.1. *Типы каналов с шумами.* Непрерывный (с непрерывным входом и выходом) канал с нормально флуктуационными аддитивными шумами является классическим каналом, для которого в основном и была построена теория потенциальной помехоустойчивости В. А. Котельникова [1]. Последняя существенно использовала возможность для этого канала геометрической интерпретации шумов и сигналов векторами п-мерного Евклидова пространства. Эта же геометрическая

* Величина H в теории информации называется энтропией источника сообщений.

интерпретация впоследствии использовалась в ряде работ Шеннона. Следует отметить, что малейшие отступления от допущений нормальности и аддитивности шумов в канале сразу же приводят к непреодолимым аналитическим трудностям. Это обстоятельство наряду с ранее указанными оправдывает переход к принятому далее дискретному описанию канала с шумами.

Впервые такое описание содержится у Шеннона [2]. Математически оно эквивалентно дискретной СЗ.

Точнее, рассматриваемый далее случай дискретного, постоянного канала с независимыми шумами математически описывается дискретной однородной СЗ с независимыми переходами. Здесь термин «постоянный канал», соответствующий однородной СЗ, взят из монографии [35]. Мы намеренно отказываемся от принятого термина «канал без памяти» в пользу более подходящего на наш взгляд термина «канал с независимыми шумами», соответствующего СЗ с независимыми переходами [1].

Таким образом, в рассматриваемом случае действие шумов в канале полностью описывается $(a\times b)$-матрицей вероятностей переходов $\mathbf{p}=\|p_\alpha^\beta\|$ входных символов α алфавита $A=(1, 2,...,a)$ в выходные символы β алфавита $B=(1, 2,...,b)$.

Дискретные моменты времени, в которых происходит посимвольная передача, как и в теории стохастических функций, будем обозначать $\sigma=(1, 2, ..., t, ..., n)$.

Более общий случай канала с зависимыми шумами не охватывается развитой теорией. Для описания таких каналов необходимо развитие теории СЗ с зависимыми переходами.

4.3.2. *Связь распределений на входе и выходе канала с шумами.* Найденные в гл. 2 связи между распределениями СА и СФ можно использовать для изучения связей между распределениями на входе и выходе канала с шумами. Эти связи прямые и через производящие функции касаются распределений входных и выходных слов (значений СА и СФ), а также входных и выходных частот символов (частот элементарных значений СА и СФ).

Кроме того, в гл. 2 приведены связи между моментами частот символов (элементарных значений). Следует подчеркнуть, что в ряде практических ситуаций невозможно проводить кодирование сообщений на входе канала и тогда схема стохастических функций описывает общую схему передачи сообщений по каналу с шумами в точности. При этом могут быть поставлены обычные физические задачи определения распределений на выходе канала, если известны распределения шумов в канале и распределения на его входе. Содержание гл. 2 можно рассматривать как решение такого рода задач для дискретного канала с независимыми шумами и, вообще говоря, произвольного дискретного источника.

§ 4.4. Кодирование и декодирование

4.4.1. *Вводные замечания.* Кодирование и декодирование, включающее решающее устройство, являются кибернетическими элементами общей схемы передачи сигналов по каналу с шумами. Без них эта схе-

[1] Слова «без памяти» в термине «канал без памяти» часто неправильно воспринимаются как указание на возможность осуществления процедур оптимального кодирования и декодирования без большой машинной памяти, что до сих пор является нерешенной технической проблемой (см. гл. 9).

ма, как уже отмечалось выше, не отличалась бы ничем от обычных физических схем, где ничто целенаправленно не вмешивается в естественное течение процесса. В данном случае процесс передачи состоял бы в подключении (если это возможно) выхода источника непосредственно к входу канала с последующей передачей по нему сообщений с неизбежными искажениями на выходе. Оптимальное кодирование и декодирование, вклиниваясь в естественный ход передачи и приема сообщений по каналу с шумами, имеет целью, насколько это возможно, уменьшить искажающее влияние шумов при фиксированной интенсивности шумов и сигналов [1].

Одной из основных задач общей теории связи является задача установления предельных возможностей такого рода.

Следует сразу же оговорить, что техническое осуществление процедур оптимального кодирования и декодирования (см. гл. 9) требует в ряде случаев тяжелых условий передачи (малое отношение сигнал/шум) затрат, соизмеримых со стоимостью энергии для соответствующего увеличения интенсивности сигнала. Однако далее эти технико-экономические вопросы не обсуждаются, тем более что многие из них носят в связи с прогрессом техники преходящий характер. Вместе с тем установление предельных возможностей оптимального кодирования имеет для кибернетики такое же значение, как общие естественнонаучные законы для физики, такие, например, как законы сохранения и др.

Интересно заметить, что оптимальные кодирование и декодирование, включающее решающее устройство, являются примерами из двух различных по своей природе групп оптимальных кибернетических задач. К первой группе относятся задачи долговременной оптимизации кибернетических устройств. Другими словами, здесь при создании кибернетического устройства производится оптимизация его структуры, обеспечивающая долговременное оптимальное функционирование кибернетического устройства. К такого рода задачам наряду с задачами оптимального кодирования относятся известное шенновское построение сложных релейных систем из минимального числа реле, оптимальные схемы надежных систем из ненадежных элементов, оптимальные иерархические структуры самоконтролирующихся систем и т. д.

Ко второй группе задач (особенно характерных для кибернетики) относятся задачи оптимальных единовременных решений в ходе функционирования кибернетических устройств. К таким задачам наряду с оптимальным декодированием, включающим решающее устройство, относятся задачи обнаружения и воспроизведения сигналов на фоне шумов (связанные со статистическими задачами оптимального выбора между гипотезами и оценки параметров), теоретико-игровые задачи отыскания оптимальных стратегий, задачи определения оптимального поведения теории динамического программирования и т. д.

Задачи еще не созданной статистической теории обучения являются в некотором смысле синтезом двух рассмотренных групп кибернетических задач. В самом деле, здесь на всем периоде функционирования кибернетической системы происходит оптимизация ее структуры в результате оценки исходов серии единовременных решений и основанных на них действий системы.

Остановимся подробнее на функциях оптимального кодирования и декодирования в общей системе передачи сигналов по каналу с шумами.

---

[1] Как отмечалось в п. 4.2.2, оптимальное кодирование для каналов без шумов имеет другую цель, состоящую в использовании «статистики» источника для предельного сокращения средней длины входных слов. кодирующих сообщения.

4.4.2. *Кодирование.* Воспроизведение физическим носителем высоковероятной группы сообщений источника, как правило, не предусматривает последующей их передачи по каналу с шумами. Поэтому среди физических воспроизведений сообщений источника может оказаться часть таких, которые в некотором смысле похожи друг на друга. Это обстоятельство делает их наиболее уязвимыми при передаче по каналу с шумами, так как на выходе канала из-за действия шумов они могут перейти друг в друга.

Целью оптимального кодирования является соотнесение сообщениям источника входных сигналов канала, предельно разнящихся друг от друга в некотором смысле.

Дискретные рассмотрения приводят к конечному пространству R входных слов длины n, состоящему из a^n входных слов. Выбор среди них M в некотором смысле оптимальных входных слов $(M\times n)$ — кодовая таблица) является сложной комбинаторной задачей. Большая часть гл. 1 содержала подготовительный материал для ее решения. Окончательное решение содержится в гл. 8.

4.4.3. *Декодирование.* Существенной частью декодирования является решающая схема, после которой собственно декодирование состоит в преобразовании, обратном преобразованию кодирования.

Решающая схема, как уже указывалось выше, должна идентифицировать искаженное шумами выходное слово y с передававшимся входным словом x — одним из слов $(M\times n)$-таблицы, известной на входе и выходе канала. Эту задачу можно рассматривать как чисто статистическую (см. п. 3.1.2). Задача имеет и комбинаторное решение (гл. 7). Оба решения оказываются эквивалентными друг другу.

Заметим, что статистические постановки и решения задач обнаружения сигнала на фоне шума достаточно разработаны [34]. В этих задачах наличие одного шума принимается за гипотезу H_0, а наличие смеси сигнала и шума принимается за гипотезу H_1. Необходимо произвести выбор между H_0 и H_1 по конечному числу наблюдений.

Таким образом, задача обнаружения сводится к задаче выбора между двумя гипотезами. Аналогично задачу различения M сигналов (или M — 1-го сигнала на фоне шума и одного шума) можно рассматривать как статистическую задачу выбора между M гипотезами (см. п. 3.1.2), где гипотеза H_m $(m=\overline{1,M})$ означает наличие m-го сигнала (наличие одного шума можно считать «нулевым» сигналом). Однако в отличие от задач обнаружения задачи различения хуже разработаны из-за неэффективности соответствующей статистической теории при $M=\text{const}>2$.

В гл. 5 будут рассмотрены задачи различения M сигналов при $M\to\infty$, когда оказывается возможным эффективное решение.

§ 4.5. Предельный случай идеального приемника В. А. Котельникова

4.5.1. *Идеальный приемник В. А. Котельникова.* В качестве примера, иллюстрирующего характер закономерностей, изучению которых посвящена вся вторая часть книги, рассмотрим хронологически первый пример такого рода. Речь идет об идеальном приемнике В. А. Котельникова [1], точнее об одном его предельном случае.

Ранее уже упоминалось, что нормально-флуктуационный характер аддитивных шумов канала и ограниченность спектра сигнала и шума с полосой в пределах $(-W, W)$ приводят к тому, что вход и выход канала описываются n-мерными Евклидовыми пространствами с размерностью

$n=T/2W$, где T — время наблюдения входного сигнала. Зафиксируем произвольные M n-мерных векторов $\{x_i\}$ $(i=\overline{1, M})$ (описывающих входные сигналы) на входе канала. Будем считать их для простоты кодирующими M равновероятных сообщений источника (при больших T, как это отмечалось выше, для широкого класса источников все высоковероятные сообщения действительно равновероятны). Разобьем выход канала на M соответствующих входным сигналам непересекающихся областей $\mathcal{E}_i$ $(i=\overline{1, M})$ и будем считать, что передавался входной сигнал x_i, если соответствующий выходной сигнал $y\in\mathcal{E}_i$ попал в область $\mathcal{E}_i$ (решающая схема). Вероятность P того, что y попадет в область $\mathcal{E}_i$, соответствующую действительно передававшемуся входному сигналу x_i $(i=\overline{1,M})$, назовем вероятностью правильного приема. Соответствующее разбиение выходного пространства на множества $\mathcal{E}_i$ $(i=\overline{1,M})$ назовем приемником. Ясно, что вероятность правильного приема P при фиксированных входных сигналах $\{x_i\}$ зависит от приемника. Идеальным приемником называется приемник с максимальной вероятностью правильного приема P.

Для случая M ортогональных, равновероятных, расположенных на сфере радиуса $\sqrt{T}N_c$ сигналов В. А. Котельников [1] нашел выражение для вероятности правильного приема

$$P = P(n, M, c) = \int_{-\infty}^{\infty} \Phi^{M-1}(x + \sqrt{2cn})\,\Phi'(x)\,dx, \qquad (4.1)$$

где $c = N_c / N_ш$ — отношение сигнал/шум по мощности и

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

Вывод формулы (4.1) существенно использует ортогональность сигналов $\{x_i\}$ $(i=1, M)$ для вероятностной независимости сумм сигналов и шума. Отсюда обязательно условие $M \leqslant n$.

Однако можно пользоваться соотношением (4.1) и для случая неортогональных сигналов в числе $M > n$, если они являются независимыми в вероятностном смысле.

4.5.2. *Предельный случай идеального приемника.* Итак, пусть $M > n$ и $n \to \infty$. Тогда $M \to \infty$. Но при любых фиксированных $|x| < K_1 = \text{const}$ и $c > 0$

$$\Phi(x + \sqrt{2cn}) \to \Phi(\sqrt{2cn}) \to 1 - e^{-cn}\ *.$$

Далее,

$$[\Phi(x + \sqrt{2cn})]^{M-1} \to (1 - e^{-cn})^{M-1} \to e^{-Me^{-cn}}.$$

Итак, при $n \to \infty$

$$[\Phi(x + \sqrt{2cn})]^{M-1} \to e^{-Me^{-cn}}. \qquad (4.2)$$

Из соотношения (4.2) ясно, что при $M = K_2 n^k$, где K_2 и k — произвольные константы; $\Phi(x + \sqrt{2cn})^{M-1} \to 1$.

---

* Требование выполнения для произвольной функции распределения $\Phi(x)$ аналогичного соотношения «на хвосте» делает для нее справедливыми все последующие результаты этого пункта.

Рассмотрим случай $M = e^{Hn}$, где $H > 0$ — некоторая константа. Тогда при $n \to \infty$ из (4.2) имеем

$$[\Phi(x + \sqrt{2cn})]^{M-1} \to e^{-e^{-(c-H)n}} \to \begin{cases} 1, & \text{если } H < c, \\ 0, & \text{если } H > c. \end{cases} \tag{4.3}$$

Переходя к пределу (4.3) под знаком интеграла (4.1), что возможно для интеграла (4.1), легко получим для вероятности правильного приема P предельные значения

$$P = P(n,\ e^{Hn},\ c) \to e^{-e^{-(c-H)n}} \to \begin{cases} 1, & \text{если } H < c \\ 0, & \text{если } H > c. \end{cases} \tag{4.4}$$

Таким образом, для рассматриваемого канала существует предельно большое число $M_c = e^{cn}$ надежно различимых ($P \to 1$) n-мерных сигналов.

Если число сигналов $M < M_c$, то их надежное различение возможно, если $M > M_c$, то надежное различение невозможно ($P \to 0$). Все последующее изложение связано с исследованием такого рода закономерностей для дискретного постоянного канала с независимыми шумами.

В гл. 7 мы вернемся к разобранному примеру в связи с результатами теории информации Шеннона для рассматриваемого канала.

Глава 5.

ОПТИМАЛЬНОЕ СТАТИСТИЧЕСКОЕ ДЕКОДИРОВАНИЕ

§ 5.1. Статистическая постановка задачи

5.1.1. *Вводные замечания.* В гл. 4 рассматривалась задача обнаружения сигнала на фоне шума. Она сводится к задаче оптимального выбора между двумя гипотезами, решаемой в классической и последовательной постановке. Более общая задача различения $M>2$ сигналов сводится к задаче оптимального выбора между M гипотезами. До сих пор неизвестно[1] эффективное решение общей задачи для фиксированного $M>2$.

Однако, если рассмотреть предельный случай растущего с объемом выборки n числа гипотеза M, то возможен ряд интересных оптимальных постановок задач и их решений, связанных с теорией потенциальной помехоустойчивости В. А. Котельникова (см. § 4.5). В частности, решение таких задач выясняет предельные возможности надежного различения растущего с объемом выборки числа сигналов, передаваемых по каналу с шумами, впервые установленные Шенноном [2]. Специфика задач такого рода состоит в том, что в них производится оптимальный выбор между M гипотезами (декодирование) о M оптимально заданных значениях параметра x распределения (кодирование), причем истинное значение параметра всегда точно совпадает с одним из гипотетических значений.

Обычно же в статистических приложениях мы не располагаем возможностью произвольно задавать все гипотетические значения параметра, и истинное значение не всегда совпадает с гипотетическим. Например, в задачах обнаружения параметр сигнал/шум при наличии одного шума всегда имеет определенное независящее от нас значение, а истинное значение этого параметра при наличии сигнала, как правило, не совпадает с гипотетическим.

В этой главе на основе статистических результатов гл. 3 решается задача надежной передачи предельно большого числа M сигналов по дискретному постоянному каналу с независимыми шумами.

В § 5.2 выводятся простые (особенно для высокого уровня шумов) оптимальные соотношения, оценивающие максимальный порядок роста числа M надежно различаемых сигналов с ростом объема выборки n.

При этом приводится эффективное построение процедуры оптимального декодирования. Относительно процедуры оптимального кодирования доказывается лишь существование последней, а также недостаточность статистических средств для ее эффективного построения.

---

[1] См. работу А. Вальда «Основные идеи общей теории статистических решений» помещенную в [33].

В § 5.3 рассматривается случай высокого уровня шумов (соответствующего случаю близких гипотез). В этом практически важном случае упрощаются основные соотношения.

В § 5.4 оценивается выигрыш от использования последовательного декодирования вместо классического в сокращении средней задержки передачи. Кроме того, рассматриваются принципиальные варианты реализации оптимальных последовательных процедур декодирования, для которых необходимо наличие надежного канала с обратной связью.

В заключение приведен известный результат теории информации о неизменности предельно большого числа M надежно передаваемых сигналов при добавлении к рассмотренному каналу канала обратной связи.

5.1.2. *Статистическая постановка задачи.* Рассмотрим дискретную во времени передачу сигналов по каналу с дискретным входом и выходом, при наличии шумов независимо воздействующих на входные посылки в различные моменты времени. Как уже отмечалось в гл. 4, непрерывную во времени передачу сигналов по непрерывному каналу можно свести к указанному выше дискретному случаю выбором достаточно малого интервала дискретизации по амплитуде и достаточно большого интервала дискретизации по времени.

В рассматриваемой дискретной схеме вход канала состоит из a входных символов алфавита $A = (1, \ldots, \alpha, \ldots, a)$, нумирующих в непрерывном случае a интервалов дискретизации входа канала. Аналогично выход канала состоит из b выходных символов алфавита $B = (1, \ldots, \beta, \ldots, b)$, причем, вообще говоря, $a \neq b$.

Пусть действие шумов в каждый момент времени описывается $(a \times b)$-матрицей вероятностей переходов

$$\mathbf{p} = \| p_\alpha^\beta \| \left(\sum_{\beta=1}^{b} p_\alpha^\beta = 1 \; (\alpha = \overline{1,a}), \sum_{\alpha=1}^{a} p_\alpha^\beta > 0 \; (\beta = \overline{1,b}) \right),$$

элементы которой $p_\alpha^\beta = \mathcal{P}(\beta/\alpha)$ являются условными вероятностями появления выходного символа β при условии появления входного символа α. Так определенный канал был назван в гл. 4 дискретным постоянным каналом с независимыми шумами.

Поступающие на вход кодирующего устройства высоковероятные равновероятные сообщения источника кодируются входными кодовыми словами $x = (\alpha_1, \ldots, \alpha_t, \ldots, \alpha_n)$ длины n, состоящими из n входных символов.

Последние под действием шумов переходят в выходные кодовые слова длины n $y = (\beta_1, \ldots, \beta_t, \ldots, \beta_n)$, состоящие из n выходных символов, которые подвергаются дальнейшему декодированию.

Пусть передаче подлежат M сообщений, пронумерованных целыми числами от 1 до M.

Так как каждому v-му сообщению однозначно сопоставляется свое входное слово $x_v (v = \overline{1,M})$, то всего необходимо иметь M различных входных слов

$$\boldsymbol{A} = (x_1, \ldots, x_v, \ldots, x_M),$$

составляющих $(M \times n)$-кодовую таблицу.

Лишь слова кодовой таблицы $\boldsymbol{A}$ используются для передачи, и они предполагаются известными на входе и выходе канала.

Пусть при передаче одного из них (какого именно — на выходе неизвестно) на выходе получено выходное слово y. Операция декодирования состоит в определении по y, какое из входных слов кодовой таблицы $\boldsymbol{A}$ было передано, чему соответствует определение передаваемого сообщения. Из-за наличия шумов при декодировании можно ошибиться, приняв непе-

редававшееся входное слово за передававшееся, и существует лишь некоторая вероятность правильного декодирования P. Мы будем допускать кодовые таблицы с повторяющимися входными словами, но при этом будем полагать, что для них $P = 0$.

Используя матрицу $\mathbf{p} = \| p_\alpha^\beta \|$, можно вычислить условную вероятность $p_x(y)$ появления y на выходе при условии передачи x. Вероятность $p_x(y)$ можно рассматривать как вероятность выборки y объема n при фиксации значения параметра x (функция правдоподобия).

Тогда задачу различения M сообщений, передаваемых по рассматриваемому каналу с шумами, на основании принятого входного слова y можно свести к задаче выбора между M гипотезами $H_v(x = x_v)(v = \overline{1, M})$ о значении параметра x вероятности $p_x(y)$ по выборке y (см. гл. 3). Специфика рассматриваемой задачи состоит в точном совпадении истинного значения параметра x с одним из M назначаемых нами многомерных гипотетических значений $x_v(v = \overline{1, M})$.

Представляет интерес определение связей между параметрами n, M и P (вероятностью правильного декодирования), которые зависят от кодовой таблицы $\boldsymbol{A}$ процедуры декодирования и матрицы $\mathbf{p} = \| p_\alpha^\beta \|$. Последняя определяется независящими от нас шумами в канале. Кодовую таблицу $\boldsymbol{A}$ и процедуру декодирования можно выбирать произвольно. Естественно выбирать их так, чтобы иметь в некотором смысле оптимальные соотношения между параметрами n, M и P. В общем случае можно считать, что параметры M и P зависят от n.

Оптимальной будем считать такую зависимость M и P от n, при которой с ростом n предельно быстро растет M, и P еще стремится к единице. Такое определение соответствует естественному желанию надежно передавать как можно больше сообщений, что уменьшает вероятность правильного декодирования.

В гл. 3 рассматривались ситуации, когда M с ростом n оставалось фиксированным (например, $M = 2$ в задачах выбора между двумя гипотезами). Однако в ряде случаев, например при передаче сообщений языкового источника, общее число M сообщений растет по экспоненциальному закону с ростом их длины. Естественно передавать экспоненциально растущее число M сообщений надежно различаемыми входными словами возможно меньшей длины n, чего можно добиться использованием оптимального соотношения между M и n (заранее ясно, что $M \leqslant a^n = e^{n \ln a}$). Поэтому описанная предельная постановка задачи помимо теоретического имеет также и практический интерес.

Итак, поставим задачу отыскания оптимальных соотношений между параметрами n, M и P, а также отыскания приводящих к ним кодовой таблицы $\boldsymbol{A}$ (процедуры кодирования) и процедуры декодирования (их мы будем называть оптимальными).

§ 5.2. Оптимальные соотношения между параметрами n, M и P и оптимальное декодирование

5.2.1. *Сведение задачи к M выборам между двумя гипотезами.* Приступим к решению поставленной в § 5.1 задачи. Будем составлять кодовую таблицу $\boldsymbol{A}$ из входных слов, считая, что каждое входящее в $\boldsymbol{A}$ входное слово $x = (\alpha_1, \ldots, \alpha_s, \ldots, \alpha_n)$ возникает с вероятностью

$$p(x) = \prod_{s=1}^{n} p_{\alpha_s} = \prod_{\alpha=1}^{a} p_\alpha^{m_\alpha(x)}, \tag{5.1}$$

равной произведению вероятностей $\bar{p} = (p_1, \ldots, p_\alpha, \ldots, p_a)\,(p_\alpha > 0)$ входящих в нее символов $\alpha\,(\alpha = \overline{1, a})$, где частота $m_\alpha(x)$ — число повторений символа α во входном слове $x\left(\sum_{\alpha=1}^{a} m_\alpha(x) = n\right)$. Другими словами, можно считать x выборками объема n из полиномиальной генеральной совокупности с параметрами $\bar{p} = (p_1, \ldots, p_\alpha, \ldots, p_a)$ (повторяющиеся входные слова также вносятся в кодовую таблицу). Если теперь передавать по дискретному каналу с независимыми шумами входные слова, то абсолютные вероятности q_β выходных символов вычисляются по формуле

$$q_\beta = \sum_{\alpha=1}^{b} p_\alpha p_\alpha^\beta \left(\sum_{\beta=1}^{b} q_\beta = 1,\ q_\beta > 0\,(\beta = \overline{1, b})\right).$$

Абсолютные вероятности выходных слов

$$y = (\beta_1, \ldots, \beta_s, \ldots, \beta_n)$$

имеют вид

$$p(y) = \prod_{s=1}^{n} q_{\beta_s} = \prod_{\beta=1}^{b} q_\beta^{m_\beta(y)},$$

где частота $m_\beta(y)\left(\sum_{\beta=1}^{b} m_\beta(y) = n\right)$ — число повторений символа β в выходном слове.

Вычислим теперь условные вероятности

$$p_{x_v}(y) = \mathcal{P}(y/x_v)$$

выходных слов y, при фиксации входных слов x_v кодовой таблицы $\boldsymbol{A}$.

Легко показать [35], что если передавалось x_u и было принято y, то

$$p_{x_v}(y) = \begin{cases} p_1(y) = \prod_{\alpha\beta} (p_\alpha^\beta)^{m_{\alpha\beta}(x_u, y)}, & \text{при } x_v = x_u; \quad (5.2) \\ p_0(y) = \prod_{\alpha\beta} q^{m_{\alpha\beta}(x_v y)} = \prod_{\beta} q_\beta^{m_\beta(y)} = p(y), & \text{при } x_v \neq x_u, \quad (5.3) \end{cases}$$

где частота $m_{\alpha\beta}(x, y)\left(\sum_{\alpha=1}^{a} m_{\alpha\beta}(x, y) = m_\beta(y),\ (\beta = \overline{1, b})\right)$ означает число пар вида (α, β) при посимвольном сопоставлении x и y (их символы сопоставляются для одних и тех же моментов времени).

В самом деле, соотношение (5.2) следует из независимого в различные моменты действия шумов в канале и определения условных вероятностей p_α^β. Соотношение (5.3) следует из того, что в этом случае вероятности появления выходных символов β при фиксации непередававшегося входного слова $x_v\,(x_v \neq x_u)$ не зависят от символов последнего и равны абсолютным вероятностям $q_\beta\,(\beta = \overline{1, b})$.

Проще всего показать это вычислением искомых вероятностей через совместные вероятности

$$\mathcal{P}(x_v, x_u, y) = p(x_v)\ p(x_u)\ p_{x_u}(y);$$

$$\mathcal{P}(x_u, y) = \sum_{x_v} \mathcal{P}(x_v, x_u, y) = p(x_u)\, p_{x_u}(y);$$

$$\mathcal{P}(x_v,\ y) = p(x_v)\,p_{x_u}(y) = \sum_{x_u} \mathcal{P}(x_v,\ x_u,\ y) = p(x_v) \sum_{x_u} p(x_u)\,p_{x_u}(y).$$

Откуда

$$p_{x_v}(y) = \frac{\mathcal{P}(x_v,\ y)}{p(x_v)} = \begin{cases} p_{x_u}(y), & \text{при } x_v = x_u, \\ p(y), & \text{при } x_v \neq x_u, \end{cases}$$

$$\text{где } p(y) = \sum_{x_u} p(x_u)\,p_{x_u}(y) = \prod_{\beta=1}^{b} q_\beta^{m_\beta(y)},$$

что и доказывает соотношения (5.2), (5.3).

Задача декодирования, трактуемая до сих пор как задача выбора между M гипотезами, допускает дальнейшие упрощения.

В самом деле, рассмотрим случай, когда в кодовой таблице $\boldsymbol{A}$ все входные слова различны. Тогда из соотношений (5.2) и (5.3) видно, что каждая из $M-1$ гипотез $H_v(x = x_v)$ о значениях x_v, не соответствующих истинному ($v \neq u$), приводит к одной и той же вероятности $p_{x_u}(y) = p_0(y)$, не зависящей от x_v.

Поэтому, так же как и в п.3.2.1, естественно перейти к другой параметризации задачи, характеризуя гипотезы самими параметрами $\mathcal{P}(\alpha, \beta) = p_\alpha p_\alpha^\beta$ в случае (5.2) и $\mathcal{P}(\alpha, \beta) = p_\alpha q_\beta$ в случае (5.3).

Тогда задача выбора между M гипотезами $H_v(a = a_v)$ $(v = \overline{1, M})$ сведется к M задачам выбора между двумя гипотезами

$$\widetilde{H}_1 = \widetilde{H}_1[\mathcal{P}(\alpha,\ \beta) = p_\alpha p_\alpha^\beta (\alpha = \overline{1, a},\ \beta = \overline{1, b})]$$

и

$$\widetilde{H}_0 = \widetilde{H}_0[\mathcal{P}(\alpha,\ \beta) = p_\alpha q_\beta\ (\alpha = \overline{1, a},\ \beta = \overline{1, b})].$$

Оптимальное решение об одном таком выборе содержится в § 3.2 и основывается на однопороговом (классическом) и двухпороговом (последовательном) анализе логарифма отношения правдоподобия, который в нашем случае имеет вид

$$l_n = \ln \frac{p_1(y)}{p_0(y)} = \sum_{s=1}^{n} z_s = \sum_{\alpha\beta} m_{\alpha\beta} \ln \frac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta}, \tag{5.4}$$

где частота $m_{\alpha\beta}$ является полиномиально распределенной случайной величиной с параметрами $\overline{\mathcal{P}}_1 = \{p_\alpha p_\alpha^\beta\}$, если верна гипотеза $\widetilde{H}_1$, и параметрами $\overline{\mathcal{P}}_0\{p_\alpha q_\beta\}$, если верна гипотеза $\widetilde{H}_0$, а z_s—независимые случайные величины, одинаково распределенные со случайной величиной z. Последняя имеет вид:

$$z = \begin{cases} \left\{ \ln \dfrac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta} \right\}, & \text{если верна гипотеза } \widetilde{H}_1 \\ p_\alpha p_\alpha^\beta \\ \left\{ \ln \dfrac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta} \right\}, & \text{если верна гипотеза } \widetilde{H}_0. \\ p_\alpha q_\beta \end{cases}$$

Отсюда с учетом соотношений гл. 3 (3.4) и (3.5) имеем

$$\mathrm{E}_1 z = \sum_{\alpha\beta} p_\alpha p_\alpha^\beta \ln \frac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta} = E(\overline{\mathscr{P}}_1, \overline{\mathscr{P}}_0) = h(\overline{q}) -$$

$$- \sum_\alpha p_\alpha h(\overline{p}_\alpha) = h(\overline{p}) - \sum_\beta q_\beta h(\overline{q}_\beta), \tag{5.5}$$

если верна гипотеза $\widetilde{H}_1$, и

$$\mathrm{E}_0 z = \sum_{\alpha\beta} p_\alpha q_\beta \ln \frac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta} = -E(\overline{\mathscr{P}}_0, \overline{\mathscr{P}}_1) = h(\overline{q}) -$$

$$- \sum_\alpha p_\alpha h(\overline{q}, \overline{p}_\alpha) = h(\overline{p}) - \sum_\beta q_\beta h(\overline{p}, \overline{q}_\beta), \tag{5.6}$$

если верна гипотеза $\widetilde{H}_0$, где

$$\overline{p} = (p_\alpha), \ \overline{p}_\alpha = (p_\alpha^\beta), \ \overline{q} = (q_\beta) \text{ и } \overline{q}_\beta = (q_\beta^\alpha) = \left(\frac{p_\alpha p_\alpha^\beta}{q_\beta}\right).$$

Из соотношений (5.5) и (5.6) следует двойное представление важной для дальнейшего изложения величины

$$2R = \mathrm{E}_1 z - \mathrm{E}_0 z = \sum_{\alpha=1}^{a} p_\alpha [h(\overline{q}, \overline{p}_\alpha) - h(\overline{p}_\alpha)] = \sum_{\beta=1}^{b} q_\beta [h(\overline{p}, \overline{q}_\beta) - h(\overline{q}_\beta)] > 0. \tag{5.7}$$

5.2.2. *Оптимальное декодирование.* Теорема 5.1. Пусть $(M \times n)$-кодовые таблицы $\boldsymbol{A}$ составляются из $M = e^{nH}$ входных слов $x_v (v = \overline{1, M})$ длины n, выбранных из полиномиальной генеральной совокупности с параметрами $\overline{p} = (p_1, \ldots, p_\alpha, \ldots, p_a)\left(\sum_{\alpha=1}^{a} p_\alpha = 1, \ p_\alpha > 0\right)$. Передача входных слов ведется по дискретному постоянному каналу с независимыми шумами, определяемому матрицей переходов $\mathbf{p} = \| p_\alpha^\beta \|$.

Пусть для каждой конкретно выбранной кодовой таблицы $\boldsymbol{A}$ процедура декодирования выходного слова y состоит в M однопороговых классических процедурах выбора между двумя гипотезами; $\widetilde{H}_1 (\overline{\mathscr{P}} = \overline{\mathscr{P}}_1)$, где $\overline{\mathscr{P}}_1 = \{p_\alpha p_\alpha^\beta\}$, и $\widetilde{H}_0 (\overline{\mathscr{P}} = \overline{\mathscr{P}}_0)$, где $\overline{\mathscr{P}}_0 = \{p_\alpha q_\beta\}$ $\left(\text{здесь } q_\beta = \sum_\beta p_\alpha p_\alpha^\beta\right)$, когда эти процедуры основаны на M нормированных n логарифмах отношений правдоподобия

$$l^{(v)} = \frac{1}{n} \sum_{\alpha\beta} m_{\alpha\beta}^{(v)} \ln \frac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta} (v = \overline{1, M})$$

с одним и тем же порогом $X = X(H)$; этот порог является корнем уравнения

$$k_{\overline{\mathscr{P}}_0 \overline{d}}(X - \mathrm{E}_0 z) - k_{\overline{\mathscr{P}}_1 \overline{d}}(X - E_1 z) = H \quad \left(\overline{d} = \left\{\ln \frac{p_\alpha^\beta}{q_\beta}\right\}\right), \tag{5.8}$$

а частота $m_{\alpha\beta}^{(v)} = m_{\alpha\beta}(x_v, y)$ означает число пар (α, β) при посимвольном сопоставлении x_v с y.

При этом решающая процедура состоит в следующем. Принимается решение о том, что передавалось x_u, если для одного $v = u$ $l^{(u)} > X$. Случаи выполнения таких неравенств для нескольких различных значений v рассматриваются как ошибки декодирования.

При наличии в кодовой таблице A совпадающих x_v также считается, что происходит ошибка декодирования.

Тогда вероятность P правильного декодирования для указанных кодовых таблиц A при $H < \frac{1}{2}[h(\bar{p}) - k_{\bar{\mathcal{P}}_1\bar{d}}(X - \mathrm{E}_1 z)]$ с ростом n имеет следующие асимптотические оценки:

$$P \approx 1 - e^{-k_{\bar{\mathcal{P}}_1\bar{d}}[X(H)-\mathrm{E}_1 z]n}, \text{ при } H < k_{\bar{\mathcal{P}}_0\bar{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z), \tag{5.9}$$

и

$$P < e^{-k_{\bar{\mathcal{P}}_1\bar{d}}[X(H)-\mathrm{E}_1 z]n}, \text{ при } H > k_{\bar{\mathcal{P}}_0\bar{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z); \tag{5.10}$$

Доказательство. Вероятность правильного декодирования P рассмотренных кодовых таблиц A, составленных из M входных слов x (выборок объема n из полиномиальной генеральной совокупности с параметрами $\bar{p} = (p_1 \ldots, p_\alpha, \ldots, p_a)$), является полной вероятностью

$$P = \sum_{A \in U} P(A) P_A, \tag{5.11}$$

где $P(A)$ — абсолютная вероятность кодовой таблицы A; P_A — условная вероятность правильного декодирования при условии использования кодовой таблицы A; U — совокупность всевозможных кодовых таблиц A.

Разобьем совокупность U на две непересекающиеся части

$$U = U' \cup \overline{U'},$$

где U' состоит из всех кодовых таблиц A с M различными входными словами x.

Согласно условию теоремы, в случае, если в A входят повторяющиеся x, то считаем, что имеет место ошибка декодирования, то есть для $A \in \overline{U'}$

$$P_A \equiv 0.$$

Тогда из (5.11) имеем

$$P = \sum_{A \in U'} P(A) P_A. \tag{5.12}$$

Пусть вероятности P_A для $A \in U'$ оцениваются снизу величиной $P^{(1)} \leqslant P_A$. Тогда из (5.12) имеем

$$P \geqslant P^{(1)} \sum_{A \in U'} P(A) = P^{(1)} P^{(2)},$$

где $P^{(2)}$ — вероятность того, что A состоит из M различных входных слов. Итак

$$P \geqslant P^{(1)} P^{(2)}.$$

Оценим вероятности, стоящие в правой части этого неравенства.

Начнем с оценки вероятности $P^{(1)}$. Используем результаты п.3.2.2. В нашем случае нормированный n логарифм отношения правдоподобия имеет вид

$$l = l(x, y) = \frac{1}{n} l_n = \frac{1}{n} \sum_{\alpha\beta} m_{\alpha\beta} \ln \frac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta} = \frac{1}{n} \sum_{s=1}^{n} z_s, \tag{5.13}$$

где частоты $m_{\alpha\beta} = m_{\alpha\beta}(x, y)$ определены на паре (x, y); z_s — независимые одинаково полиномиально распределенные случайные величины, принимающие значения $\ln(p_\alpha p_\alpha^\beta / p_\alpha q_\beta)$ с вероятностями $p_\alpha q_\beta$, если верна гипотеза $\widetilde{H}_0$ (x не передавалось, принято y), и вероятностями $p_\alpha p_\alpha^\beta$, если верна гипотеза $\widetilde{H}_1$ (x передавалось, принято y).

Пусть единственный порог X классической процедуры выбора между гипотезами $\widetilde{H}_0$ и $\widetilde{H}_1$ находится в пределах $\mathrm{E}_0 z < X < \mathrm{E}_1 z$. Тогда для вероятностей ошибок первого и второго рода имеем из (3.6) соответственно следующие асимптотические выражения

$$\gamma \approx e^{-nk\overline{\mathscr{P}}_0 \overline{d}(X - \mathrm{E}_0 z)} \tag{5.14}$$

и

$$1 - \delta \approx e^{-nk\overline{\mathscr{P}}_1 \overline{d}(X - \mathrm{E}_1 z)}, \tag{5.15}$$

где положено

$$\overline{\mathscr{P}}_0 = \{p_\alpha q_\beta\};\ \overline{\mathscr{P}}_1 = \{p_\alpha p_\alpha^\beta\};\ \overline{d} = \left\{\ln \frac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta}\right\}.$$

Если $X > \mathrm{E}_1 z$, то легко показать, что γ по-прежнему выражается соотношением (5.14), а

$$1 - \delta = 1 - e^{-nk\overline{\mathscr{P}}_1 \overline{d}(X - \mathrm{E}_1 z)}. \tag{5.16}$$

Декодирование кодовой таблицы $\boldsymbol{A}$, состоящей из M полиномиально распределенных входных слов $x_v (v = \overline{1, M})$, было сведено в п. 5.2.1 к M выборам между гипотезами $\widetilde{H}_0$ и $\widetilde{H}_1$, по M величинам $l^{(v)} = l(x_v, y)$ (см. (5.13)).

Для оценки вероятности $P^{(1)}$ рассмотрим случай, когда все $x_v (v = \overline{1, M})$ различные. Тогда правильное декодирование соответствует совмещению M событий, а именно, — превышению величиной $l^{(u)}$ порога X, если передавалось x_u, и M — 1-му непревышению порога X для остальных $l^{(v)}$ $(v \neq u)$, соответствующих непередававшимся x_v.

Выбор единого порога X для всех M выборов между $\widetilde{H}_0$ и $\widetilde{H}_1$ естествен из-за симметрии задачи. Рассматриваемые события являются зависимыми ввиду того, что соответствующие им случайные величины $z_s^{(v)}$, входящие в выражения $l^{(v)}$, определены на одном и том же[1] y. Можно привести лишь оценки (впрочем, довольно хорошие) для вероятности их совмещения $P_{\boldsymbol{A}}$, где $\boldsymbol{A} \in U'$.

Эти оценки имеют вид

$$\delta + \sum_{v=1}^{M-1} (1 - \gamma) - (M - 1) \leqslant P^{(1)} \leqslant \delta. \tag{5.17}$$

Верхняя оценка очевидна, так как вероятность совмещения нескольких событий не может превзойти вероятности одного из них. Нижняя оценка является частным случаем неравенства Буля (см. (2.12)).

[1] Это обстоятельство не учитывается в [35], что впрочем не сказывается на окончательных результатах из-за близости оценки Буля в рассматриваемом случае к вероятности совмещения независимых событий (см. далее).

Подставив в неравенства (5.17) выражения для γ и $1-\delta$ из (5.14), (5.15) и (5.16), а также учитывая, что $M = e^{nH}$ $(0 < H < \ln a)$ будем иметь

$$\left.\begin{aligned} 1 - e^{-nk_{\overline{\mathcal{P}}_1\bar{d}}(X-\mathrm{E}_1 z)} - e^{-n[k_{\overline{\mathcal{P}}_0\bar{d}}(X-\mathrm{E}_0 z)-H]} \leqslant P^{(1)} \leqslant 1 - e^{-nk_{\overline{\mathcal{P}}_1\bar{d}}(X-\mathrm{E}_1 z)}, \\ \text{при } \mathrm{E}_0 z < X < \mathrm{E}_1 z \end{aligned}\right\} \quad (5.18)$$

и

$$P^{(1)} \leqslant e^{-nk_{\overline{\mathcal{P}}_1\bar{d}}(X-\mathrm{E}_1 z)}, \quad \text{при } X > \mathrm{E}_1 z.$$

Положим $X = X(H)$ равным корню уравнения

$$k_{\overline{\mathcal{P}}_1\bar{d}}(X - \mathrm{E}_1 z) = k_{\overline{\mathcal{P}}_0\bar{d}}(X - \mathrm{E}_0 z) - H.$$

Тогда

$$1 - 2e^{-nk_{\overline{\mathcal{P}}_1\bar{d}}(X-\mathrm{E}_1 z)} \leqslant P^{(1)} \leqslant 1\, e^{-nk_{\overline{\mathcal{P}}_1\bar{d}}(X-\mathrm{E}_1 z)},$$

и неравенства (5.18) переходят в соотношения

$$P^{(1)} \approx 1 - e^{-nk_{\overline{\mathcal{P}}_1\bar{d}}(X-\mathrm{E}_1 z)}, \quad \text{при } 0 \leqslant X < \mathrm{E}_1 z, \tag{5.19}$$

$$P^{(1)} \leqslant e^{-nk_{\overline{\mathcal{P}}_1\bar{d}}(X-\mathrm{E}_1 z)}, \quad \text{при } X > \mathrm{E}_1 z. \tag{5.20}$$

Приступим теперь к оценке вероятности $P^{(2)}$ того, что все M выборок объема n различны. Пронумеруем всевозможные входные слова x в порядке невозрастания их вероятностей (см. (5.1)

$$P_t = p(x_t) = \prod_{\alpha=1}^{a} p_\alpha^{m_\alpha(x_t)} \qquad (t = \overline{1, a^n}),$$

где $m_\alpha(x_t)$ означает частоту символов α во входном слове x_t длины n.

Имеем соотношение

$$P^{(2)} = \sum_{t=1}^{a^n} P_{t_1} \sum_{t_2 \neq t_1} P_{t_2} \dots \sum_{t_M \neq t_1, \dots, t_{M-1}} P_{t_M} \geqslant$$

$$\geqslant 1\,(1 - P_1)(1 - P_1 - P_2) \dots \left(1 - \sum_{t=1}^{M-1} P_t\right), \tag{5.21}$$

где оценка снизу получена последовательными оценками сумм справа налево с помощью соотношений (следуют из принятой нумерации)

$$\sum_{s=1}^{S} P_t \leqslant \sum_{t=1}^{S} P_t \quad (t_s \neq t_{s'}, \text{ при } 1 \leqslant s \neq s' \leqslant S \leqslant a^n).$$

Дальнейшие оценки $P^{(2)}$ приведем в предположении

$$H \leqslant \frac{1}{2}[h(\bar{p}) - k_{\overline{\mathcal{P}}_1\bar{d}}(X - \mathrm{E}_1 z)]. \tag{5.22}$$

При $n \to \infty$ (см. (2.26)) $P_t \approx e^{-nh(\bar{p})}$, поэтому с учетом (5.21) и (5.22) имеем

$$(M-1)\sum_{t=1}^{M-1} P_t \approx e^{-n(h(\bar{p})-2H)} < e^{-nk_{\overline{\mathcal{P}}_1\bar{d}}(X-\mathrm{E}_1 z)}$$

и

$$P^{(2)} \geqslant 1\,(1-P_1)(1-P_1-P_2)\ldots\left(\sum_{t=1}^{M-1} P_t\right) \geqslant \left(1-\sum_{t=1}^{M-1} P_t\right)^{M-1} \geqslant$$

$$\geqslant 1-(M-1)\sum_{t=1}^{M-1} P_t \geqslant 1-e^{-nk_{\overline{\mathcal{P}}_1\overline{d}}(X-\mathrm{E}_1 z)}.$$

Поэтому из оценок (5.19) и (5.20) для $P^{(1)}$ следуют аналогичные оценки для $P \geqslant P^{(1)}P^{(2)}$ при дополнительном условии (5.22)

Но можно показать, что $X = X(H)$ монотонно растет с ростом H, и из (5.8) следует, что при $X \to \mathrm{E}_1 z$, $H \to k_{\overline{\mathcal{P}}_0\overline{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z)$. Тогда окончательные вероятности P имеют оценки (5.9) и (5.10), указанные в условии теоремы. Что и требовалось доказать.

Прокомментируем результаты теоремы 5.1. Прежде всего отметим, что ограничение (5.22) связано лишь со способом доказательства теоремы. Однако верхняя граница H, при которой имеет место оценка (5.9), для вероятности P не снижается, если

$$\frac{1}{2}[h(\overline{p}) - k_{\overline{\mathcal{P}}_1\overline{d}}(X - \mathrm{E}_1 z) \geqslant k_{\overline{\mathcal{P}}_0\overline{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z).$$

Для этого достаточно выполнение условия

$$2k_{\overline{\mathcal{P}}_0\overline{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z) + k_{\overline{\mathcal{P}}_1\overline{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z) < h(\overline{p}). \tag{5.23}$$

Итак, в случае выполнения условий (5.23) соотношения (5.9) и (5.10) имеют место без каких-либо ограничений.

Оптимизируем соотношения между параметрами n, M и P, приведенные в теореме 5.1, при фиксированных $\overline{p} = (p_1, \ldots, p_\alpha, \ldots, p_a)$ и $\mathbf{p} = \| p_\alpha^\beta \|$. Соотношения (5.9) и (5.10) показывают, что для выполнения условия $P \to 1$ (надежной передачи) параметр H, определяющий асимптотический рост числа $M = e^{nH}$ входных слов, с ростом n должен быть меньше параметра $k_{\overline{\mathcal{P}}_0\overline{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z)$. Поэтому для оптимальности связи между параметрами n, M и P необходимо иметь максимально большое значение $k_{\overline{\mathcal{P}}_0\overline{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z)$. Из соотношения (5.7) заключаем, что последнее зависит от вероятностей $\overline{p} = (p_1, \ldots, p_\alpha, \ldots, p_a)$ и матрицы вероятностей переходов $\mathbf{p} = \| p_\alpha^\beta \|$, причем лишь вероятности $\overline{p}$ мы можем выбирать по своему усмотрению. В соответствии с этим найдем максимальное значение

$$C = \max_{\overline{p}} k_{\overline{\mathcal{P}}_0\overline{d}}(\mathrm{E}_1 z - \mathrm{E}_0 z)$$

при фиксированной матрице $\mathbf{p} = \| p_\alpha^\beta \|$ по всем произвольным вероятностям $\overline{p} = (p_1 \ldots, p_\alpha, \ldots, p_a)$, удовлетворяющим условиям $\sum_{\alpha=1}^{a} p_\alpha = 1$ $(p_\alpha \geqslant 1 (\alpha = \overline{1, a}))$.

Полученное значение C, зависящее только от матрицы $\mathbf{p} = \| p_\alpha^\beta \|$, то есть лишь от вероятностных свойств шумов канала, является фундаментальной константой канала[1]. Приводящее к нему распределение

[1] В теории информации Шеннона [2] эта величина называется пропускной способностью канала.

$\overline{p}^{(0)} = (p_1^{(0)}, \ldots, p_a^{(0)}, \ldots, p_a^{(0)})$ [оно обращает в максимум $k_{\overline{\mathscr{P}_0, \overline{d}}}(\mathrm{E}_1 z - \mathrm{E}_0 z)$] задает параметры полиномиальной генеральной совокупности, из которой выбираются входные слова. Распределение $\overline{p}^{(0)}$ и соответствующую генеральную совокупность будем называть оптимальными.

Таким образом, если имеет место (5.23), то оптимальные соотношения между параметрами n, M и P при фиксированной матрице переходов $\mathbf{p} = \| p_a^{\beta} \|$ имеют вид:

$$P \approx 1 - e^{-nk_{\overline{\mathscr{P}_1, \overline{d}}}[X(H) - \mathrm{E}_1 z]}, \quad M = e^{nH} \tag{5.24}$$

если

$$0 < H < C = \max_{\overline{p}} k_{\overline{\mathscr{P}_0 d}}(\mathrm{E}_1 z - \mathrm{E}_0 z). \tag{5.25}$$

Оптимальным оказывается (см. гл. 6) и приводящий к соотношениям (5.24) и (5.25) способ декодирования, указанный в теореме 5.1. Что касается оптимального способа кодирования, то теорема 5.1 не дает никаких указаний об эффективном способе построения конкретной оптимальной кодовой таблицы $\boldsymbol{A}$.

В самом деле, при указанном оптимальном способе декодирования в теореме 5.1 оценивалась по существу средняя вероятность правильного декодирования

$$P = \sum_{\mathbf{A}} P(\boldsymbol{A}) P_{\mathbf{A}} \left(\sum_{\mathbf{A}} P(\boldsymbol{A}) = 1 \right) \tag{5.26}$$

по множеству всевозможных кодовых таблиц, где $P(\boldsymbol{A})$ — вероятность кодовой таблицы $\boldsymbol{A}$; $P_{\boldsymbol{A}}$ — вероятность правильного декодирования при ее использовании. Поэтому из соотношения (5.26) следует лишь существование хотя бы одной кодовой таблицы $\boldsymbol{A}^*$, для которой $P_{\boldsymbol{A}^*} \geqslant P$, то есть имеют место оптимальные соотношения[1] (5.24) и (5.25).

Однако из теоремы 5.1 не следует способ построения кодовой таблицы $\boldsymbol{A}^*$; более того, из нее не следует, что не найдется такая кодовая таблица $\boldsymbol{A}^{**}$, при которой соотношения (5.24) выполняются, когда $H > C$ (обратная теорема).

5. 2. 3. *Недостаточность статистического подхода.* По сути дела, трудность использования полученного результата для каких-либо заключений о вероятности правильного декодирования $P_{\boldsymbol{A}}$ для конкретной случайно выбранной кодовой таблицы $\boldsymbol{A}$ состоит в грубости оценки распределения $\{P_{\boldsymbol{A}}\}$ по его среднему значению P.

Покажем, что в нашем случае такая оценка принципиально неудовлетворительна.

В самом деле, случайный алгоритм построения кодовых таблиц $\boldsymbol{A}$, изложенный в предыдущем пункте, позволяет рассматривать их как выборочные значения Mn-мерной генеральной совокупности определяемой распределением $\{P_{\boldsymbol{A}}\}$. Тогда вероятности правильного декодирования $P_{\boldsymbol{A}}$ можно рассматривать как значения случайной величины

$$\Pi = \left\{ \frac{P_{\boldsymbol{A}}}{P(\boldsymbol{A})} \right\}$$

ограниченной условиями $0 \leqslant \Pi \leqslant 1$. Оцененная соотношением (5.24) вероятность P является средним значением $\mathrm{E}\Pi = P$ случайной величины Π.

---

[1] Последнее соображение стало после работы [2] традиционным в современной теоретико-информационной литературе.

Оценка случайной величины П по ее среднему значению P носит вероятностный характер и основывается на неравенстве Чебышева.

В самом деле, имеем для случайной величины $1-\text{П}$ и произвольного $T>0$

$$\mathscr{P}[1-\text{П}<T\text{E}(1-\text{П})]>1-T^{-1}.$$

Откуда

$$\mathscr{P}[\text{П}>1-T(1-P)]>1-T^{-1}. \tag{5.27}$$

Полагая $T=e^{-nr}$ $(r>0)$ и используя для P оценку (5.24) получим из (5.27)′ оценку

$$P'=\mathscr{P}[\text{П}>1-e^{-n(k-r)}]>P''=1-e^{-nr}, \tag{5.28}$$

где $k=k_{\underset{\mathscr{P}_{1d}}{\ }}[X(H)-\text{E}_1 z]$.

Итак, мы получили экспоненциальную оценку искомой вероятности P' того, что вероятность П правильного декодирования для случайно выбранной конкретной кодовой таблицы больше некоторой экспоненциально стремящейся к единице величины.

Для того, чтобы оценка (5.28) не теряла смысла, произвольный параметр r нужно выбирать в пределах $0<r<k$, что уже приводит к неравенству $P''<P$. Выбирая r близким к k, мы существенно нарушаем оптимальность соотношений между параметрами кодирования. В самом деле, при этом оценка вероятности правильного декодирования П имеет меньший, чем это должно быть в оптимальном случае, экспоненциальный порядок стремления к единице. С другой стороны, выбор r близким к нулю понижает порядок экспоненциального стремления вероятности P' к единице. А это требует для достижения практически приемлемых значений P', близких к единице, неоправданного по сравнению с оптимальным случаем завышения длины n входных слов. Выбор любых промежуточных значений r между нулем и единицей приводит одновременно к обоим указанным недостаткам.

Не избавляет от них и неравенство Чебышева, учитывающее оценку неизвестной дисперсии DП. В самом деле, имеем из-за того, что $0\leqslant\text{П}\leqslant 1$,

$$\text{DП}=\text{EП}^2-(\text{EП})^2\leqslant\text{EП}-(\text{EП})^2=\text{EП}(1-\text{EП}),$$

откуда

$$\mathscr{P}[\text{П}>\text{EП}-T\sqrt{\text{EП}(1-\text{EП})}]\geqslant\mathscr{P}[\text{П}>\text{EП}-T\sqrt{\text{DП}})>1-T^{-1}$$

и с учетом приведенных выше обозначений имеем оценку

$$P'(\text{П}>1-e^{-n(\frac{k}{2}-r)}\sqrt{1-e^{-nk}}-e^{-nk})>1-e^{-nr},$$

еще более слабую, чем оценка (5.28).

Таким образом, очевидна недостаточность имеющихся статистических средств для оценки оптимальности кода, образованного случайным алгоритмом.

Построение оптимальных кодовых таблиц, а также доказательство обратной теоремы, требуют использования комбинаторных методов гл. 1, не имеющих прямого отношения к используемым в этой главе статистическим методам. Поэтому мы здесь не будем их касаться. Заметим лишь, что использование этих комбинаторных методов в гл. 7 позволяет оценить вероятность P' того, что конкретная кодовая таблица, состоящая из входных слов, выбранных из оптимальной генеральной совокупности, окажется оптимальной в указанном выше смысле. Можно показать (см. гл. 7), что с

ростом n вероятность P' стремится к единице быстрее вероятности правильного декодирования P. Поэтому указанный алгоритм построения оптимальной кодовой таблицы практически всегда приводит к цели и приведенный в теореме 5.1 способ декодирования оказывается оптимальным.

Однако осуществление оптимального кодирования сопряжено с рядом технических трудностей. Они связаны вовсе не со случайной природой алгоритма (она в некоторых случаях может оказаться даже полезной; см. гл. 9). Важнейшая техническая трудность состоит в необходимости иметь экспоненциально растущий с длиной n входных слов объем памяти для хранения M входных слов.

Минимальную длину n входных слов при заданном порядке нерастания числа сигналов $M = e^{nH}$ (он определяется константой H), а также при заданной практически приемлемой вероятности правильного декодирования P можно вычислить по формуле

$$n \approx [\ln(1-P)^{-1}]/k_{\overline{\mathcal{P}_1 d}}(X(H) - \mathrm{E}_1 z),$$

которая следует из соотношения (5.24)

§ 5.3. Случай высокого уровня шумов

В наиболее интересном практически случае при большом уровне шумов, математически соответствующем случаю близких гипотез, заметно упрощаются соотношения, приведенные в предыдущем параграфе. В рассматриваемом случае дискретного постоянного канала с независимыми шумами высокий уровень шумов означает требование малости модулей разностей

$$\max_{1 \leqslant \beta \leqslant b} | p_\alpha^\beta - r_\beta | = \Delta \to 0 \quad (\alpha = \overline{1, a}),$$

где

$$\sum_{\beta=1}^{b} r_\beta = 1 \text{ и } r_\beta > 0.$$

Другими словами, требуется поэлементная близость матрицы $\mathbf{p} = (\bar{p}_\alpha) = \| p_\alpha^\beta \|$ вероятностей переходов к матрице $\bar{e}\bar{r} = (\bar{r}) = \| r_\beta \|$ с постоянными элементами в столбцах.

Начнем упрощения соотношений с рассмотрения формулы Бартлетта (3.14) в нашем случае. Имеем, пренебрегая величиной $O(\Delta^3)$,

$$-2\mathrm{E}_0 z \sim 2\mathrm{E}_1 z \sim \mathrm{E}_1 z - \mathrm{E}_0 z \sim D_0 z \sim \mathrm{D}_1 z \sim 2R \sim O(\Delta^2), \qquad (5.29)$$

где [см. 5.5]

$$R = \mathrm{E}_1 z = \mathrm{E}(\bar{\mathcal{P}}_1, \bar{\mathcal{P}}_0) = h(\bar{q}) - \sum_\alpha p_\alpha h(\bar{p}_\alpha) = h(\bar{p}) - \sum_\beta q_\beta h(\bar{q}_\beta) \to 0.$$

Таким образом, случай высокого уровня шумов можно определять как случай, при котором $R \to 0$.

Далее, согласно соотношению (1.84), имеем, пренебрегая величиной $o(\varepsilon^2)$,

$$k_{\overline{pd}}(\varepsilon) \sim \frac{\varepsilon^2}{2\gamma''(0)},$$

$$\text{где } \gamma''(0) = \sum_\rho p_\rho d_\rho^2 - \Big(\sum_\rho p_\rho d_\rho\Big)^2,$$

откуда [см. (5.29)]

$$k_{\overline{\mathscr{P}_{0d}}}(\varepsilon) \sim \frac{\varepsilon^2}{2D_0 z} \sim k_{\overline{\mathscr{P}_{1d}}}(\varepsilon^2) \sim \frac{\varepsilon^2}{2D_1 z} \sim \frac{\varepsilon^2}{4R}. \tag{5.30}$$

Используем соотношения (5.30) для решения трансцендентного уравнения (5.8)

$$K_{\overline{\mathscr{P}_{0d}}}(X - \mathrm{E}_0 z) - k_{\overline{\mathscr{P}_{1d}}}(X - \mathrm{E}_1 z) = H$$

в рассматриваемом случае. Имеем $\frac{(X+R)^2}{4R} - \frac{(X-R)^2}{4R} \sim H$, откуда $X \sim H$. Далее, из (5.29) и (5.30) имеем предельную для H константу

$$k_{\overline{\mathscr{P}_{0d}}}(\mathrm{E}_1 z - \mathrm{E}_0 z) \sim k_{\overline{\mathscr{P}_{1d}}}(\mathrm{E}_1 z - \mathrm{E}_0 z) \sim R. \tag{5.31}$$

Покажем, что в рассматриваемом случае высокого уровня шумов ограничения (5.22) теоремы 5.1 всегда выполняются. В самом деле, из (5.30) следует, что условие

$$H \leqslant \frac{1}{2}[h(\bar{p}) - k_{\overline{\mathscr{P}_{1d}}}(X - \mathrm{E}_1 z)]$$

эквивалентно условию

$$h(\bar{p}) - 2H \geqslant \frac{(H-R)^2}{4R}.$$

Последнее имеет место, если

$$h(\bar{p}) - 2H \geqslant \frac{(H - h(\bar{p}))^2}{4h(\bar{p})}$$

из-за монотонного возрастания выражения $\frac{(H-R)^2}{4R}$ с ростом R и того, что $R \leqslant h(\bar{p})$ (см. (5.5).

Но последнее неравенство всегда имеет место, так как после несложных преобразований получаем эквивалентное ему неравенство

$$H < \frac{\sqrt{3}}{2-\sqrt{3}} h(\bar{p}),$$

заведомо выполняемое (рассматривается случай $H < R < h(\bar{p})$).

Если искать ограничений R, то они следуют из (5.23) и имеют вид[1]

$$R < \frac{1}{3} h(\bar{p}).$$

Можно переформулировать с учетом приведенных выше упрощений результаты теоремы 5.1 для случая высокого уровня шумов.

Следствие 5. 1. 1 [34]. Пусть $(M \times n)$-кодовые таблицы $\boldsymbol{A}$ составляются из $M = e^{nH}$ $(H > 0)$ входных слов $x_v (v = \overline{1, M})$ длины n, выбран-

[1] В [34] (стр. 337) вместо условия $R < \frac{1}{3} h(\bar{p})$ требуется лишь $R < \frac{4}{9} h(\bar{p})$, что формально надостаточно, однако, так как $R \to 0$, то это не оказывает влияния на окончательное правильное утверждение предельной теоремы.

ных из полиномиальной генеральной совокупности с параметрами $\overline{p} = (p_1, \ldots, p_\alpha, \ldots, p_a)$ $\left(\sum_{\alpha=1}^{a} p_\alpha = 1,\ p_\alpha > 0\right)$. Передача входных слов ведется по дискретному постоянному каналу с высоким уровнем шумов, определяемых матрицей переходов $\mathbf{p} = \| p_\alpha^\beta \|$ с элементами, удовлетворяющими условиям

$$\max_{1 \leqslant \beta \leqslant b} (p_\alpha^\beta - r_\beta) = \Delta \to 0, \text{ где } \sum_{\beta=1}^{b} r_\beta = 1;\ r_\beta \geqslant 0.$$

Пусть для каждой конкретно выбранной кодовой таблицы $\boldsymbol{A}$ процедура декодирования выходного слова y состоит в M однопороговых классических процедур выбора между двумя гипотезами $\widetilde{H}_1 (\overline{\mathcal{P}} = \overline{\mathcal{P}}_1)$, где $\overline{\mathcal{P}}_1 = \{ p_\alpha p_\alpha^\beta \}$, и $\widetilde{H}_0 (\overline{\mathcal{P}} = \overline{\mathcal{P}}_0)$, где $\overline{\mathcal{P}}_0 = \{ p_\alpha q_\beta \}$ (здесь $q_\beta = \Sigma\, p_\alpha p_\alpha^\beta$), когда эти процедуры основаны на M нормированных n логарифмах отношений правдоподобия

$$l^{(v)} = \frac{1}{n} \sum_{\alpha\beta} m_{\alpha\beta}^{(v)} \ln \frac{p_\alpha p_\alpha^\beta}{p_\alpha q_\beta} \quad (v = \overline{1, M})$$

с одним и тем же порогом H и частота $m_{\alpha\beta}^{(v)} = m_{\alpha\beta} (x_v, y)$ означает число пар (α, β) при посимвольном сопоставлении x_v и y.

Решающая процедура состоит в следующем. Принимается решение о том, что передавалось x_v, если для одного $v = u$ $l^{(u)} > H$. Случаи выполнения таких неравенств для нескольких различных значений v рассматриваются как ошибка декодирования.

При наличии в кодовой таблице $\boldsymbol{A}$ совпадающих x_v также считается, что происходит ошибка декодирования.

Тогда вероятности P правильного декодирования для указанных кодовых таблиц $\boldsymbol{A}$ с ростом n имеет следующие асимптотические оценки:

$$P \approx 1 - e^{-\frac{(H-R)^2}{4R} n}, \text{ при } H < R,$$

и

$$P < e^{-\frac{(H-R)^2}{4R} n}, \text{ при } H > R,$$

где

$$R = h(\overline{p}) - \sum_{\alpha} p_\alpha h(\overline{p}_\alpha) = h(\overline{q}) - \sum_{\beta} q_\beta h(q_\beta).$$

Используем результат следствия 5.1.1. для получения оптимального соотношения между параметрами n, M и P. В рассматриваемом случае оно следует из соотношений (5.24) и (5.25) и имеет вид

$$P \approx 1 - e^{-\frac{(H-C)^2}{4C} n}, \quad M = e^{nH}, \tag{5.32}$$

если

$$0 < H < C = \max_{\overline{p}} \left[h(\overline{q}) - \sum_{\alpha} p_\alpha h(\overline{p}_\alpha) \right] (\overline{q} = \overline{p}\mathbf{p}).$$

Минимальную длину входных слов n при заданных параметрах P, H и C получим из (5.32)

$$n \approx \frac{4C \ln \frac{1}{1-P}}{(H-C)^2}.$$

Прокомментируем полученные результаты. В случае высокого уровня шумов особенно прозрачны оптимальные соотношения между параметрами n, M и P, общий экспоненциальный характер которых для общего случая

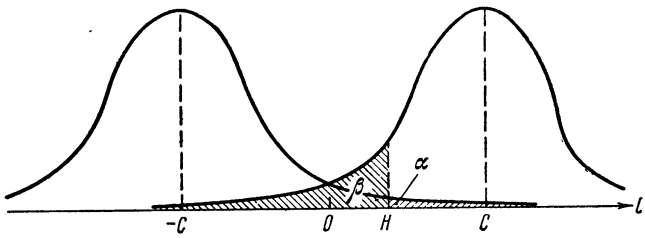


Рис. 5.1. Статистическое декодирование

рассматриваемого канала хорошо известен в литературе [10]. Однако здесь выяснены статистический смысл параметров H и R и, что особенно важно, статистическая структура процедуры декодирования.

Заметим, что можно предвидеть соотношения (5.32), исходя из следующих соображений. Согласно (5.13) и (5.29) нормированный n логарифм отношения правдоподобия $l^{(v)}$ имеет представление

$$l^{(v)} = \frac{1}{n} \sum_{\alpha\beta} m_{\alpha\beta}(x_v, y) = \frac{1}{n} \sum_{i=1}^{n} z_i^{(v)}, \tag{5.33}$$

среднее

$$\mathrm{E}l^{(v)} \approx \begin{cases} R, & \text{если имеет место гипотеза } H_1 (x_v = x_u) \ (x_u \text{ передавалось}) \\ -R, & \text{если имеет место гипотеза } H_0 (x_v \neq x_u) \ (x_u \text{ не передавалось}) \end{cases}$$

и дисперсию $\mathrm{D}l^{(v)} \approx \frac{2R}{n} (v = \overline{1, M})$.

Будучи, согласно (5.33), средним арифметическим независимых одинаково распределенных случайных величин, случайная величина $l^{(v)}$ при $n \to \infty$ асимптотически нормальна и, так как с ростом n $\mathrm{D}l^{(v)} \to 0$, то в пределе ее распределение вырождается в δ-функцию.

Все эти рассуждения вполне строги, однако получение выражений вероятностей $\alpha = \gamma$ и $\beta = 1 - \delta$ ошибок первого и второго рода с помощью асимптотических выражений для «хвостов» нормального распределения [$\Phi(x) \approx \approx 1 - e^{-x^2/2}$] при $x \to \infty$ в общем случае не обосновано, хотя и приводит к правильным результатам в рассматриваемом случае канала. Получающаяся при этом ситуация выбора между двумя гипотезами изображена на рис. 5.1, где параметр H играет роль единственного порога, а величины $-R$ и R являются средними значениями распределений l при конкурирующих гипотезах (в оптимальном случае $R = C$).

§ 5.4. Последовательное декодирование

До сих пор в оптимальную схему надежного различения большого числа сигналов наряду с оптимальным кодированием включалась процедура оптимального декодирования, основанная на M классических выборах между двумя гипотезами (классическое декодирование). Возможна ли замена в этой схеме классического декодирования последовательным декодированием, основанным на M последовательных выборах между двумя гипотезами, и если возможна, то к какому положительному эффекту она приводит?

Прежде всего покажем возможность такой замены, указанной в [34], наряду с возможностью применения последовательного анализа в системах посимвольного приема по каналу с шумами. При этом существенно использование надежного канала обратной связи. Передача в рассматриваемом случае входных сигналов (слов) в целом не исключает такой возможности.

В самом деле, пусть в рассматриваемом случае имеется надежный (без шумов) канал обратной связи. Пусть по-прежнему на входе и выходе имеется $(M \times n)$-кодовая таблица, но вместо классического декодирования, основанного на фиксированной длине n выходного слова y, будем проводить последовательное декодирование, усеченное длиной n. По каналу обратной связи будем передавать с выхода на вход сигнал об окончании декодирования принятого выходного слова y. Оно осуществляется, как правило, по некоторой начальной части y, содержащей $\nu < n$ первых символов (ν — случайная величина).

Такой сигнал явится одновременно сигналом для передачи нового входного слова и т. д.

Использование последовательного декодирования вместо классического приведет лишь к изменению той части теоремы 5.1 и следствия 5.1.1, где говорится об одном из M классических выборов между гипотезами $\widetilde{H}_1$ и $\widetilde{H}_0$. В новом варианте последовательный выбор между гипотезами $\widetilde{H}_1$ и $\widetilde{H}_0$ в оптимальном случае должен основываться на последовательном анализе накопленной суммы [см. (5.33)].

$$l_n^{(v)} = nl^{(v)} = \sum_{s=1}^{n} z_s \, (n = 1,2, \ldots)$$

с двумя порогами, выбираемыми согласно (3.7) гл. 3,

$$\left.\begin{aligned} &\ln B \approx \ln \frac{1-\delta}{1-\gamma} \approx -k_{\overline{\mathcal{P}}_{1d}}(X(H) - \mathrm{E}_1 z) \underset{C \to 0}{\to} -\frac{C}{4}\left(1 - \frac{H}{C}\right)^2 n \\ &\text{и} \\ &\ln A \approx \ln \frac{\delta}{\gamma} \approx k_{\overline{\mathcal{P}}_{0d}}(X(H) - \mathrm{E}_0 z) \underset{C \to 0}{\to} \frac{C}{4}\left(1 + \frac{H}{C}\right)^2 n. \end{aligned}\right\}.$$

Их вид определяется существенно разными по порядку малости вероятностями ошибок первого и второго рода

$$\gamma \approx e^{-k_{\overline{\mathcal{P}}_{0d}}[X(H) - \mathrm{E}_0 z]n} \underset{C \to 0}{\longrightarrow} e^{-\frac{(H+C)^2}{4C}n} \ll 1 - \delta \approx e^{-k_{\overline{\mathcal{P}}_{1d}}[X(H) - \mathrm{E}_1 z]n} \underset{C \to 0}{\longrightarrow} e^{-\frac{(H-C)^2}{4C}n}. \tag{5.34}$$

Для случая высокого уровня шумов (соответствует случаю близких гипотез) из соотношений (5.34) и (3.18) гл. 3 получим отношение среднего $\mathrm{E}\nu$

длины части выходного слова до вынесения решения к соответствующей фиксированной ее длине n в классическом случае декодирования [34]:

$$e = \frac{E\nu}{n} \approx \begin{cases} \left(1 + \sqrt{\frac{\ln\gamma}{\ln(1-\delta)}}\right)^{-2} = \frac{1}{4}\left(1 - \frac{H}{C}\right)^2, & \text{если верна гипотеза } \widetilde{H}_0 \quad (5.35) \\ \left(1 + \sqrt{\frac{\ln(1-\delta)}{\ln\gamma}}\right)^{-2} = \frac{1}{4}\left(1 + \frac{H}{C}\right)^2 & \text{если верна гипотеза } \widetilde{H}_1 \quad (5.36) \end{cases}$$

Усечением в случаях близкой к 1 эффективности $1 - e$ можно пренебречь.

Соотношения (5.35) и (5.36) показывают, что использование последовательного анализа в одном из M выборов между двумя гипотезами приводит к тем большему эффекту, чем H ближе к C при гипотезе $\widetilde{H}_0$ и чем H меньше C при гипотезе $\widetilde{H}_1$.

Приступим теперь к оценке эффективности последовательного декодирования в целом. Для этого заметим, что, вообще говоря, имеются два способа осуществления декодирования, состоящего из M элементарных процедур выбора между двумя гипотезами: 1) можно производить одним устройством M выборов между двумя гипотезами один за другим во времени; 2) можно одновременно осуществлять M таких выборов M устройствами. (Технически это по-видимому неоправдано, особенно при больших M.)

Следует сразу же подчеркнуть, что существенный выигрыш в среднем времени декодирования при использовании последовательного декодирования по сравнению с классическим декодированием имеет место лишь при первом способе его осуществления (считаем время, идущее на осуществление одного выбора между гипотезами, пропорциональным необходимой для этого длине выходного слова).

В самом деле, при первом способе осуществления декодирования идущее на него время равно сумме M времен, идущих на элементарные выборы между двумя гипотезами. Все они, за исключением одного, происходят когда верна гипотеза $\widetilde{H}_0$. Поэтому отношение среднего времени для осуществления последовательного декодирования к времени для осуществления классического декодирования будет совпадать при больших M с соответствующим отношением для элементарных выборов, когда верна гипотеза $\widetilde{H}_0$, то есть оно равно $e = \frac{1}{4}\left(1 - \frac{H}{C}\right)^2$ [см. (5.35)]. Как уже отмечалось, особый эффект имеет место при H, близком к C.

При втором способе осуществление декодирования для рассматриваемого случая больших M выигрыш от последовательного декодирования нивелируется из-за необходимости ждать для вынесения окончательного решения окончания всех элементарных последовательных процедур до первого принятия гипотезы $\widetilde{H}_1$.

В самом деле, не говоря уже о затяжке одной из них, когда верна гипотеза $\widetilde{H}_1$, с ростом M увеличивается вероятность затяжки до классического усечения хотя бы одной из $M - 1$ последовательных процедур, когда верна гипотеза $\widetilde{H}_0$.

Ясно, что последовательное декодирование не может привести к увеличению значения фундаментальной константы канала C, так как оно касается лишь повышения темпа передачи, а не увеличения числа $M = e^{nH}$

надежно передаваемых входных слов, определяемых при фиксированном n константой C.

Но может быть, можно более рационально использовать обязательный для последовательного декодирования канал обратной связи? В теории информации доказывается, что использование канала обратной связи не может привести к увеличению фундаментальной константы C для рассматриваемого канала с независимыми шумами.

Это обстоятельство имеет интуитивное оправдание. В самом деле, описанная в § 5.2 процедура оптимального декодирования в предельном при $n \to \infty$ случае состоит в разбиении пространства R^* выходных слов на предельно большое число M предельно малочисленных непересекающихся множеств $\mathscr{E}_i (i = \overline{1, M})$, при котором вероятность правильного декодирования P еще стремится к 1. Процедура решения состоит в принятии решения о том, что передавалось x_i, если $y \in \mathscr{E}_i$. Добавим еще M' аналогичных множеств $\mathscr{E}_i (i = \overline{M+1,\ M+M'})$. При этом множества $\mathscr{E}_i (i = \overline{1,\ M+M'})$ окажутся пересекающимися. Вероятность правильного декодирования сохранится за исключением случая попадания y в пересечения $\mathscr{E}_i \cap \mathscr{E}_j$ В этом случае с помощью надежного канала обратной связи можно как-то разрешать неопределенность, например запросами о повторении передачи. Однако в случае $n \to \infty$ распределение $\mathscr{P}(\mathscr{E}_i/x_i) \to 1$ имеет характер δ-функции и пересечения $\mathscr{E}_i$ при сохранении условия $\mathscr{P}(\mathscr{E}_i/x_i) \to 1$ приводят к числу M' с меньшим экспоненциальным порядком, чем у числа M.

В случае же, когда экспоненциальный порядок у M' больше, чем у M, то $\mathscr{P}(\mathscr{E}_i[x_i) \to 0$.

Глава 6

КОМБИНАТОРНАЯ КОНСТРУКЦИЯ ОПТИМАЛЬНОГО ДЕКОДИРОВАНИЯ. ДОСТАТОЧНЫЕ УСЛОВИЯ P-РАЗЛИЧИМОСТИ

§ 6.1. Вводные замечания

В предыдущей главе были исчерпаны статистические средства в продвижении к окончательной цели — построению оптимального в смысле Шеннона кода для рассматриваемого канала с шумами. В этой и последующей главах для окончательного решения проблемы необходимо привлечение новых математических средств в основном комбинаторного характера. Весь вспомогательный математический аппарат для этого был развит в первых двух главах. Поэтому остановимся на идейном плане главы.

Прежде всего следует отметить, что в этой главе впервые появляется ряд новых понятий, назначение которых состоит лишь в кратком обозначении ситуаций с длинным словесным описанием.

Далее дана формулировка задачи, использующая введенные понятия (§ 6.2).

Приводятся конструкции совокупностей выходных слов, в которые с вероятностью P переходят входные слова (P-представляют последние) (§ 6.3). Формулируются необходимые условия для числа элементов множества выходных слов произвольной структуры, чтобы оно P-представляло входное слово (§ 6.4).

На основании необходимых условий P-представимости в § 6.5 формулируются условия так называемой P-различимости M входных слов (вероятность правильного их декодирования равна P). Эти условия формулируются в терминах частот $M_{x_l}^{\mathbf{m}}$ входных слов x_j $(j \neq i)$, имеющих с x_l матричное расстояние $\mathbf{m}$ (x_i, x_j), в точности равное $\mathbf{m}$. Приводится пример $\{M_{x_l}^{\mathbf{m}}\}$ выполнения достаточных условий, который имеет и самостоятельный интерес для дальнейшего изложения. В нем $M_{x_l}^{\mathbf{m}}$ выступают как математические ожидания случайных величин.

В заключение (§ 6.6) показана эквивалентность найденного оптимального декодирования, которое можно назвать комбинаторным, с найденным в предыдущей главе оптимальным статистическим декодированием.

§ 6.2. Определения и постановка задачи

Рассмотрим передачу в дискретные моменты времени по дискретному постоянному каналу с независимыми шумами.

Это означает, что в каждый дискретный момент времени $t = 1,2, \ldots, s$ на вход канала подается один из a возможных входных символов α $(\alpha = \overline{1, a})$. На выходе канала в тот же момент возникает один из b возможных выходных символов β $(\beta = \overline{1, b})$. При этом, вообще говоря, $a \neq b$.

Действие шумов в канале описывается матрицей условных вероятностей $\mathbf{p} = \| p_\alpha^\beta \| = (\overline{p_\alpha})$, где элемент матрицы $p_\alpha^\beta = \mathcal{P}(\beta/\alpha)$ означает условную вероятность β при условии, что имеет место α.

Вероятности $p_\alpha^\beta \geqslant 0$ должны удовлетворять условиям

$$\sum_{\beta=1}^{b} p_\alpha^\beta = 1 \ (\alpha = \overline{1,a}); \ \sum_{\alpha=1}^{a} p_\alpha^\beta > 0 \ (\beta = 1,b),$$

из которых первое очевидно, а второе означает, что на входе канала учитываются лишь символы, возникающие с ненулевой вероятностью.

Постоянство канала означает неизменность матрицы $\mathbf{p}$ в различные моменты времени. Независимость шумов означает независимость перехода символов α в символы β в различные моменты времени.

Зафиксируем n последовательных моментов времени $\sigma = (1,2, \ldots, n)$ и рассмотрим входные и выходные слова длины n $x = (\alpha_1, \ldots, \alpha_s, \ldots \alpha_n)$ и $y = (\beta_1, \ldots, \beta_s, \ldots, \beta_n)$. Они являются элементами множеств R и R^* входных и выходных слов длины n, содержащих тех и других в числе a^n и b^n соответственно. R (R^*) будем называть пространством входных (выходных) слов длины n.

С помощью элементов матрицы $\mathbf{p}$ легко вычисляются условные вероятности выходных слов $y \in R^*$, если на вход было подано входное слово $x \in R$

$$\mathcal{P}(y/x) = \prod_{s=1}^{n} p_{\alpha_s}^{\beta_s} = \prod_{\alpha=1}^{a} \prod_{\beta=1}^{b} (p_\alpha^\beta)^{m_{\alpha\beta}}, \tag{6.1}$$

где $m_{\alpha\beta}$ — число пар символов (α, β) в паре слов (x, y) в соответствующих моментах времени (при $p_\alpha^\beta = 0$, $m_{\alpha\beta} = 0$ и полагаем $0^0 = 1$).

Если ввести векторы $\overline{m} = (m_\alpha)$ и $\overline{m^*} = (m_\beta^*)$ с компонентами m_α ($\alpha = \overline{1,a}$) и m_β^* ($\beta = \overline{1,b}$), означающими число символов α и β в x и y соответственно, то:

$$\sum_{\beta=1}^{b} m_{\alpha,\beta} = m_\alpha; \ \sum_{\alpha=1}^{a} m_{\alpha,\beta} = m_\beta^*; \ \sum_{\alpha=1}^{a} m_\alpha = \sum_{\beta=1}^{b} m_\beta^* = n.$$

Зная вероятности $\mathcal{P}(y/x)$, можно вычислять вероятности появления одного из y-ов произвольного множества $\mathcal{E} \subset R^*$, при фиксации на входе $x \in R$. Эти вероятности имеют вид

$$\mathcal{P}(\mathcal{E}/x) = \sum_{y \in \mathcal{E}} \mathcal{P}(y/x). \tag{6.2}$$

Для дальнейшего изложения удобны следующие определения, использующие вероятности (6.2)

Определение 1. Будем говорить, что множество $\mathcal{E} \subset R^*$ выходных слов длины n P-представляет входное слово $x \in R$ той же длины, если

$$P = \mathcal{P}(\mathcal{E}/x) \to 1, \text{ при } n \to \infty.$$

Если $P = \mathcal{P}(\mathcal{E}/x) \to 0$, при $n \to \infty$, то будем говорить, что $\mathcal{E}$ не представляет x.

Определение 2. Входные слова длины n

$$x_1, \ldots, x_i, \ldots, x_j, \ldots, x_M \tag{6.3}$$

будем называть P-различимыми, если существуют такие P_i-представляющие их множества выходных слов той же длины

$$\mathscr{E}_1, \ldots, \mathscr{E}_i, \ldots, \mathscr{E}_j, \ldots, \mathscr{E}_M \tag{6.4}$$

соответственно, что при $n \to \infty$ множества

$$\mathscr{E}_i' = \mathscr{E}_i \cap \bigcup_{j \neq i} \mathscr{E}_j \ (i = \overline{1, M}) \tag{6.5}$$

не представляют x_i; то есть при $n \to \infty$

$$Q_i = \mathscr{P}(\mathscr{E}_i'/x_i) \to 0. \tag{6.6}$$

При этом

$$P = \min_{i=\overline{1, M}} (P_i - Q_i) \to 1. \tag{6.7}$$

Последнее определение оправдано следующими соображениями. Заполним, вообще говоря, пересекающимися множествами $\mathscr{E}_i \ (i = \overline{1, M})$ все R^* и произвольно перераспределим между ними элементы их пересечений $\mathscr{E}_i'$. Полученные множества $\widetilde{\mathscr{E}}_i$ заполняют все R^*, не пересекаются и являются по-прежнему $\widetilde{P}$-различимыми, где $\widetilde{P} \geqslant P \to 1$ при $n \to \infty$. Тогда появившееся на выходе выходное слово y обязательно попадет в одно из множеств $\widetilde{\mathscr{E}}_j \ (j = \overline{1, M})$. Если передавалось x_i, то y попадет с вероятностью не меньшей, чем P в $\mathscr{E}_i$. Поэтому будем считать, что передавалось x_i, если $y \in \mathscr{E}_i$, при этом вероятность ошибки не превосходит $1 - P$.

С помощью входных слов $x_i \ (i = \overline{1, M})$ можно закодировать высоковероятные равновероятные слова источника сообщений (см. гл. 4) и передавать их по рассматриваемому каналу. Далее, по принятому выходному слову y согласно изложенному выше правилу выносится решение о передававшемся входном слове x_i (правило декодирования). Заметим, что можно считать передававшимся входное слово x_i, если $y \in \mathscr{E}_i - \mathscr{E}_i'$ и считать ошибкой декодирования попадание y в $\mathscr{E}_i'$ В этом случае вероятность правильного решения, как легко видеть, будет не меньше вероятности P, определяемой соотношением (6.7). Эту вероятность будем называть вероятностью правильного декодирования.

Совокупность входных слов, определяемая $(n \times M)$-таблицей (6.3), назовем (n, M)-кодом. Назовем (n, M, P)-кодом совокупность P-различимых входных слов (6.3), соответствующих им множеств выходных слов (6.4) с процедурой декодирования. Ясно, что параметры n, M и P нельзя задать произвольно (см, гл. 5). Шеннон [2] открыл, что при $n \to \infty$ существует наибольший экспоненциальный порядок роста числа $M = M_C = e^{nC}$, при котором P еще стремится к единице.

Если же $M = C^{nH}$, где $H > C$, то $P \to 0$.

(n, M, P)-код, у которого при достаточно большом n M может быть сколь угодно близко к M_C, будем называть оптимальным в смысле Шеннона. Наша цель состоит в построении такого (n, M, P)-кода.

В заключение параграфа заметим, что можно дать развернутое толкование определений 1 и 2 подчеркиванием того, что $x_i = x_i^{(n)}$ и $\mathscr{E}_i = \mathscr{E}_i^{(n)}$ являются n-ми членами бесконечных при $n \to \infty$ последовательностей. Однако мы этого явно нигде не будем делать, тем более что в последующем изложении конкретная структура x_i и $\mathscr{E}_i$ делают такое развернутое толкование излишним.

§ 6.3. *P*-представляющие множества

Рассмотрим введенное в гл. 1 множество $\mathscr{E}_\sigma^{\overline{m}} \subset R$, которое состоит из всех x-ов, содержащих m_α символов α ($\alpha = \overline{1, a}$), где $\overline{m} = (m_\alpha)$. Тогда

$$R = \bigcup_{\overline{m}} \mathscr{E}_\sigma^{\overline{m}} \; (\mathscr{E}_\sigma^{\overline{m}_1} \cap \mathscr{E}_\sigma^{\overline{m}_2} = \varnothing, \text{ при } \overline{m}_1 \neq \overline{m}_2).$$

Легко подсчитать (см. (1.20), (1.36)) число элементов

$$N(\mathscr{E}_\sigma^{\overline{m}}) = C_n^{\overline{m}} = \frac{n!}{\prod\limits_{\alpha=1}^{a} m_\alpha!} \tag{6.8}$$

и вероятность

$$\mathscr{P}(\mathscr{E}_\sigma^{\overline{m}}) = C_n^{\overline{m}} \overline{p}^{\overline{m}} = \frac{n!}{\prod\limits_{\alpha=1}^{a} m_\alpha!} \prod_{\alpha=1}^{a} p_\alpha^{m_\alpha}, \tag{6.9}$$

где p_α — абсолютная вероятность символа α ($\alpha = \overline{1, a}$).

Запишем $x \in \mathscr{E}_\sigma^{\overline{m}}$ в форме $x = (\tau_1, \ldots, \tau_\alpha, \ldots, \tau_a)$, где $\tau_\alpha \subset \sigma$ — подмножества моментов времени, в которых имеют место символы α ($N(\tau_\alpha) = m_\alpha$).

Представление $x = \{\tau_\alpha\}$ индуцирует представление пространства R^* в виде прямого произведения $R^* = \underset{\alpha=1}{\overset{a}{\times}} R^*_{\tau_\alpha}$ подпространств $R^*_{\tau_\alpha}$, элементы которых образуют составные элементы R с общим упорядочением, определенным в σ.

В силу независимости переходов символов α в символы β вероятностные рассмотрения можно проводить независимо в каждом из подпространств $R^*_{\tau_\alpha}$ в отдельности.

Рассмотрим подмножество y-ов $\mathscr{E}_{\tau_\alpha}^{\overline{m}_\alpha} \subset R^*_{\tau_\alpha}$, определяемое аналогично множеству $\mathscr{E}_\sigma^{\overline{m}}$. Вектор $\overline{m}_\alpha = (m_{\alpha\beta})$ является строкой матрицы $\mathbf{m} = \| m_{\alpha\beta} \| = (\overline{m}_\alpha)$, элементы которой $(m_{\alpha\beta})$, как уже упоминалось в предыдущем параграфе, означают число пар (α, β) в паре (x, y).

Имеем

$$N(\mathscr{E}_{\tau_\alpha}^{\overline{m}_\alpha}) = C_{m_\alpha}^{\overline{m}_\alpha} \text{ и } \mathscr{P}(\mathscr{E}_{\tau_\alpha}^{\overline{m}_\alpha}/\tau_\alpha) = C_{m_\alpha}^{\overline{m}_\alpha} \overline{p}_\alpha^{\overline{m}_\alpha}. \tag{6.10}$$

Рассмотрим совокупность $[m_\alpha \mathfrak{M}_\alpha]$ векторов $\overline{m}_\alpha$ и определим множество y-ов $\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha} = \bigcup\limits_{\overline{m}_\alpha \in [m_\alpha \mathfrak{M}_\alpha]} \mathscr{E}_{\tau_\alpha}^{\overline{m}_\alpha}$.

Его вероятность имеет вид

$$\mathscr{P}(\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha}/\tau_\alpha) = \sum_{\overline{m}_\alpha \in [m_\alpha \mathfrak{M}_\alpha]} C_{m_\alpha}^{\overline{m}_\alpha} \overline{p}_\alpha^{\overline{m}_\alpha}.$$

Изучим ее асимптотическое поведение при $n \to \infty$. Введем обозначения

$$\left.\begin{aligned} &m_\alpha = n\mu_\alpha;\ m_{\alpha\beta} = m_\alpha \mu_\alpha^\beta = n\mu_\alpha\mu_\alpha^\beta;\ m_\beta^* = n\nu_\beta; \\ &\mathbf{m} = \| m_{\alpha\beta} \| = n \| \mu_\alpha \mu_\alpha^\beta \|;\ \bar{\mu} = (\mu_\alpha);\ \boldsymbol{\mu} = \| \mu_\alpha^\beta \|;\ \bar{\nu} = \bar{\mu}\boldsymbol{\mu} = (\bar{\nu}_\beta); \\ &\sum_{\alpha=1}^{a} \mu_\alpha = \sum_{\mu=1}^{b} \mu_\alpha^\beta = \sum_{\beta=1}^{b} \nu_\beta = 1. \end{aligned}\right\} \quad (6.11)$$

Обозначим (см. п. 1.61) множество векторов $\bar{\mu}$ a-мерного единичного куба лежащих на плоскости $\sum_{\alpha=1}^{a} \mu_\alpha = 1$ через $\mathfrak{A}_a$.

Аналогичное множество векторов $\bar{\nu}$ обозначим $\mathfrak{A}_b$. Матрица $\boldsymbol{\mu}$ отображает точки $\mathfrak{A}_a$ в точки $\mathfrak{A}_b$. Все векторные строки матрицы $\boldsymbol{\mu} = (\bar{\mu}_\alpha)$ лежат в $\mathfrak{A}_b$. Из-за целочисленности величин m_α, $m_{\alpha\beta}$ и m_β^* следует, что к ним приводят целые ε-окрестности чисел $\mu_\alpha = m_\alpha/n$, $\mu_\alpha^\beta = m_{\alpha\beta}/m_\alpha$ и $\nu_\beta = m_\beta^*/n$ в $\mathfrak{A}_a$ и $\mathfrak{A}_b$, где $\varepsilon = \frac{1}{2n}$. Поэтому всюду ниже для ε-окрестностей векторов в $\mathfrak{A}_a$ и $\mathfrak{A}_b$ стремление $\varepsilon \to 0$ считается более медленным, чем $1/n$, при $n \to \infty$. Сделанные оговорки разъясняют смысл описания множеств целочисленных векторов $\bar{m}$, $\bar{m}^*$ и матриц $\mathbf{m}$ с помощью множеств непрерывных векторов $\bar{\mu}$, $\bar{\nu}$ и матриц $\boldsymbol{\mu}$ соответственно.

С помощью формулы Стирлинга, при $n \to \infty$, используя обозначения (6.11), получим из (6.8), (6.9) и (6.10)

$$N(\mathscr{E}_\sigma^{\bar{m}}) \approx e^{h(\bar{\mu}_\alpha)},\ \mathscr{P}(\mathscr{E}_\sigma^{\bar{m}}) \approx e^{n[h(\bar{\mu}, \bar{p}) - h(\bar{\mu})]}$$

и

$$\mathscr{P}(\mathscr{E}_{\tau_\alpha}^{\bar{m}_\alpha}/\tau_\alpha) \approx e^{-\mu_\alpha[h(\bar{\mu}_\alpha, \bar{p}_\alpha) - h(\bar{\mu}_\alpha)]n}, \quad (6.12)$$

где

$$h(\bar{\mu}_\alpha, \bar{p}_\alpha) = -\sum_{\beta=1}^{b} \mu_\alpha^\beta \ln p_\alpha^\beta \text{ и } h(\bar{\mu}_\alpha) = -\sum_{\beta=1}^{b} \mu_\alpha^\beta \ln \mu_\alpha^\beta.$$

Из-за того, что число элементов $[m_\alpha \mathfrak{M}_\alpha]$ растет не экспоненциально с ростом n ($C_{m_\alpha+b-1}^{b-1} = e^{b \ln n + O(n^{-1})}$), имеем из (6.12) следующие асимптотические соотношения

$$\mathscr{P}(\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha}/\tau_\alpha) \approx \begin{cases} 1 - e^{-\mu_\alpha[h(\bar{\mu}_\alpha^{(0)}, \bar{p}_\alpha) - h(\bar{\mu}_\alpha^{(0)})]n}, & \text{если } \bar{p}_\alpha \in \mathfrak{M}_\alpha, \\ e^{-\mu_\alpha[h(\bar{\mu}_\alpha^{(0)}, \bar{p}_\alpha) - h(\bar{\mu}_\alpha^{(0)})]n}, & \text{если } \bar{p}_\alpha \notin \mathfrak{M}_\alpha, \end{cases} \quad (6.13)$$

где

$$0 \leqslant h(\bar{\mu}_\alpha^{(0)}, \bar{p}_\alpha) - h(\bar{\mu}_\alpha^{(0)}) = \min_{\bar{\mu}_\alpha \in \mathfrak{M}_\alpha} [h(\bar{\mu}_\alpha, \bar{p}_\alpha) - h(\mu_\alpha)], \quad (6.14)$$

и включение и невключение вектора $\bar{p}_\alpha$ в $\mathfrak{M}_\alpha$ в (6.13) рассматривается с некоторыми его ε-окрестностями.

В дальнейшем примем конкретный вид множества $\mathfrak{M}_\alpha = \mathfrak{M}_\alpha(\varepsilon_\alpha)$, определяемый следующими линейными относительно μ_α^β условиями

$$h(\bar{\mu}_\alpha, \bar{p}_\alpha) - h(\bar{p}_\alpha) \leqslant \varepsilon_\alpha. \tag{6.15}$$

В этом случае легко находится минимум (6.14) при условии (6.15) (см. (1.69))

$$k_\alpha(\varepsilon_\alpha) = h(\bar{\mu}_\alpha^{(0)}, \bar{p}_\alpha) - h(\bar{\mu}_\alpha^{(0)}) = \lambda[\gamma_\alpha'(0) + \varepsilon_\alpha] - \gamma_\alpha(\lambda) \geqslant 0, \tag{6.16}$$

где λ является корнем уравнения

$$\gamma_\alpha'(0) + \varepsilon_\alpha = \gamma_\alpha'(\lambda) \text{ и } \gamma_\alpha(\lambda) = \ln \sum_{\beta=1}^{b} (p_\alpha^\beta)^{1-\lambda}, d_\alpha = -\ln p_\alpha^\beta \ (\mu_\alpha^\beta = 0, \text{ при } p_\alpha^\beta = 0).$$

Функция $k_\alpha(\varepsilon_\alpha) \geqslant k_\alpha(0)$ и уравнение $k_\alpha(\varepsilon_\alpha) = v > 0$ имеет два корня: $-\varepsilon_\alpha' = k_\alpha^{-1}(v) < 0$ и $\varepsilon_\alpha = k_\alpha^{-1}(v) > 0$.

Аналогичные оценки, полученные в [20,35] с помощью производящих, здесь менее естественны. Заметим, что при $\varepsilon_\alpha > 0$, $\bar{p}_\alpha \in \mathfrak{M}_\alpha(\varepsilon_\alpha)$ со своей ε-окрестностью. Если $\varepsilon_\alpha < 0$, то $\bar{p}_\alpha$ не входит в $\mathfrak{M}_\alpha$ ($\bar{p}_\alpha \notin \mathfrak{M}_\alpha(\varepsilon_\alpha)$) со своей ε-окрестностью. Поэтому[1] имеем из (6.13) и (6.16)

$$\mathcal{P}(\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha(\varepsilon_\alpha)}/\tau_\alpha) \approx \begin{cases} 1 - e^{-\mu_\alpha k_\alpha(\varepsilon_\alpha) n}, & \text{при } \varepsilon_\alpha > 0 \\ e^{-\mu_\alpha k_\alpha(\varepsilon_\alpha) n}, & \text{при } \varepsilon_\alpha < 0. \end{cases} \tag{6.17}$$

Введем в рассмотрение множество y-ов

$$\mathscr{E}_x^{\mathbf{m}} = \bigtimes_{\alpha=1}^{a} \mathscr{E}_{\tau_\alpha}^{m_\alpha},$$

где $\mathbf{m} = \mathbf{m}(x, y) = \| m_{\alpha\beta} \| = (\bar{m}_\alpha)$ и $x = (\tau_\alpha)$, а также множество

$$\mathscr{E}_x^{\mathfrak{M}(\varepsilon)} = \bigcup_{\mathbf{m} \in [n\mu_{\mathfrak{M}(\varepsilon)}]} \mathscr{E}_x^{\mathbf{m}} = \bigtimes_{\alpha=1}^{a} \bigcup_{\bar{m}_\alpha \in [m_\alpha \mathfrak{M}_\alpha(\varepsilon_\alpha)]} \mathscr{E}_{\tau_\alpha}^{\bar{m}_\alpha} = \bigtimes_{\alpha=1}^{a} \mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha(\varepsilon_\alpha)}, \tag{6.18}$$

где $\mathfrak{M}(\varepsilon) = \{\mathfrak{M}_\alpha(\varepsilon_\alpha)\}$.

Теорема 6.1. Пусть $x \in \mathscr{E}_\sigma^{\bar{m}}$, тогда множество y-ов $\mathscr{E}_x^{\mathfrak{M}(\varepsilon)}$, где $\mathfrak{M}(\varepsilon) = \{\mathfrak{M}_\alpha(\varepsilon_\alpha)\}$ и $\varepsilon_\alpha = k_\alpha^{-1}(\varepsilon^2/\mu_\alpha) > 0$ ($\alpha = \overline{1, a}$) P-представляет x, причем

$$P \approx 1 - e^{-\varepsilon^2 n}. \tag{6.19}$$

Доказательство. Из соотношений (6.17) и (6.18) имеем

$$P = \mathcal{P}(\mathscr{E}_x^{\mathfrak{M}(\varepsilon)}/x) \approx \prod_{\alpha=1}^{a} (1 - e^{-\mu_\alpha k_\alpha(\varepsilon_\alpha)}) \approx 1 - e^{-\min\limits_{\alpha=\overline{1,a}} \mu_\alpha k_\alpha(\varepsilon_\alpha) n}.$$

Отсюда при $\varepsilon_\alpha = k_\alpha^{-1}(\varepsilon^2/\mu_\alpha) > 0$ следует (6.19).

[1] Здесь по-прежнему приближение $K(s) \approx e^{ks}$ понимается в смысле $K(s) = e^{sk + O(\ln s)}$.

Для дальнейшего изложения необходимо введение множества $\mathscr{E}_x^{\widetilde{\mathfrak{M}}} = \mathscr{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}$, где компоненты $\widetilde{\mathfrak{M}}(\varepsilon) = \{\mathfrak{M}_\alpha(\varepsilon_\alpha, \varepsilon'_\alpha)\}$ определяются следующими условиями: $-\varepsilon'_\alpha < h(\bar{\mu}_\alpha, \bar{p}_\alpha) - h(\bar{p}_\alpha) < \varepsilon_\alpha$, причем — ε'_α и ε'_α (ε_α, $\varepsilon'_\alpha > 0$) корни уравнения $k_\alpha(u) = \varepsilon^2/\mu_\alpha$.

Следствие 6.1.1. Множество $\mathscr{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}$ P-представляет $x \in \mathscr{E}_\sigma^{\bar{m}}$, где

$$P \approx 1 - e^{-\varepsilon^2 n}. \tag{6.20}$$

Доказательство. Из определения множества $\widetilde{\mathfrak{M}}(\varepsilon) = \{\mathfrak{M}_\alpha(\varepsilon_\alpha, \varepsilon'_\alpha)\}$ следует, что множества $\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha(\varepsilon_\alpha, \varepsilon'_\alpha)}$ выражаются через множества $\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha(\varepsilon_\alpha)}$

$$\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha(\varepsilon_\alpha, \varepsilon'_\alpha)} = \mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha(\varepsilon_\alpha)} - \mathscr{E}^{\mathfrak{M}_\alpha(-\varepsilon'_\alpha)}. \tag{6.21}$$

Из (6.21), учитывая (6.17) и выражения ε'_α и ε_α через ε, легко показать, что

$$\mathscr{P}(\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha(\varepsilon_\alpha, \varepsilon'_\alpha)}/\tau_\alpha) \approx \mathscr{P}(\mathscr{E}_{\tau_\alpha}^{\mathfrak{M}_\alpha(\varepsilon_\alpha)}/\tau_\alpha) \approx 1 - e^{-\varepsilon^2 n},$$

откуда следует утверждение следствия.

Заметим, что при $\varepsilon \to 0$, для $\mu \in \widetilde{\mathfrak{M}}(\varepsilon)$ имеем $\mu \to \mathbf{p}$ (где стремление матриц друг к другу понимается как их поэлементное стремление). Отсюда следует, что если $\mu \in \widetilde{\mathfrak{M}}(\varepsilon)$, то при $\varepsilon \to 0$, $\bar{\nu} = \bar{\mu}\mu \to \bar{\nu} = \bar{\mu}\,\mathbf{p}$. Пусть $x \in \mathscr{E}_\sigma^{\bar{m}} \subset R$, и $y \in \mathscr{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)} \subset (x, R^*)$.

Рассмотрим множество y-ов $\mathscr{E}_\sigma^{\bar{m}^*} \subset R^*$, определяемое аналогично множеству $\mathscr{E}_\sigma^{\bar{m}}$. Введем в рассмотрение множество $\{\bar{m}^*\}_{\bar{\mu},\varepsilon}$ векторов $\bar{m}^*$, удовлетворяющих соотношению $\frac{1}{n}\bar{m}^* = \bar{\nu} = \bar{\mu}\mu$, где μ пробегает все элементы множества $\mathfrak{M}(\varepsilon)$.

Определим множество y-ов

$$\mathscr{E}_\sigma^{\{\bar{m}^*\}_{\bar{\mu},\varepsilon}} = \bigcup_{\bar{m}^* \in \{\bar{m}^*\}_{\bar{\mu},\varepsilon}} \mathscr{E}_\sigma^{\bar{m}^*}. \tag{6.22}$$

Это множество обладает тем свойством, что, каково бы ни было $x \in \mathscr{E}_\sigma^{\bar{m}}$, ему принадлежат все y из множества $\mathscr{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}$ P-представляющего x.

Из (6.22) и из предыдущих замечаний следует, что при $\varepsilon \to 0$

$$N(\mathscr{E}_\sigma^{\{\bar{m}^*\}_{\bar{\mu},\varepsilon}}) \to e^{nh(\bar{\nu}_1) + O(\ln n)}, \tag{6.23}$$

где $\bar{\nu}_1 = \bar{\mu}\mathbf{p}$.

§ 6.4. Необходимые условия P-представимости

Теорема 6.2. Каковы бы ни были $x \in \mathscr{E}_\sigma^{\bar{m}} \subset R$ и $\mathscr{E} \subset R^*$, если число элементов $\mathscr{E}$

$$N(\mathscr{E}) \leqslant e^{n\left[\sum_{\alpha=1}^{a} \mu_\alpha h(\bar{p}_\alpha) - \eta'\right]},$$

где

$$\eta' \geqslant \varepsilon^2 + \overline{\varepsilon}' \left(\overline{\varepsilon}' = \sum_{\alpha=1}^{a} \mu_\alpha \varepsilon'_\alpha,\ \varepsilon'_\alpha = -k^{-1}\ (\varepsilon^2/\mu_\alpha) \right), \tag{6.24}$$

то

$$\mathcal{P}(\mathcal{E}/x) \approx e^{-n\varepsilon^2},$$

т. е. $\mathcal{E}$ не представляет x.

Доказательство. Произвольное множество $\mathcal{E} \subset R^*$ всегда можно представить в виде

$$\mathcal{E} = \mathcal{E} \cap \overline{\mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}} \cup \mathcal{E} \cap \mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)},$$

где $\overline{\mathcal{E}} = R^* - \mathcal{E}$ — дополнение $\mathcal{E}$ до всего R^*.

Далее, используя соотношение (2.1) и условия теоремы 2, получим следующую цепочку оценок

$$\mathcal{P}(\mathcal{E}/x) = \mathcal{P}(\mathcal{E} \cap \mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}/x) + \mathcal{P}(\mathcal{E} \cap \overline{\mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}}/x) \leqslant$$

$$\leqslant \max_{y \in \mathcal{E} \cap \mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}} \mathcal{P}(y/x) \cdot N(\mathcal{E} \cap \mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}) + \mathcal{P}(\overline{\mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}}/x) \leqslant$$

$$\leqslant \max_{y \in \mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}} \mathcal{P}(y/x) \cdot N(\mathcal{E}) + \mathcal{P}(\overline{\mathcal{E}_x^{\widetilde{\mathfrak{M}}(\varepsilon)}}/x) \approx$$

$$\approx \max_{\mu \in \widetilde{\mathfrak{M}}(\varepsilon)} e^{-n \sum_{\alpha=1}^{a} \mu_\alpha h(\overline{\mu}_\alpha, \overline{p}_\alpha)} \cdot e^{n\left[\sum_{\alpha=1}^{a} \mu_\alpha h(\overline{p}_\alpha) - \eta'\right]} + e^{-n\varepsilon^2} =$$

$$= e^{-n \sum_{\alpha=1}^{a} \mu_\alpha (h(\overline{p}_\alpha) - \varepsilon'_\alpha) + n\left[\sum_{\alpha=1}^{a} \mu_\alpha h(\overline{p}_\alpha) - \eta'\right]} + e^{-n\varepsilon^2} \approx$$

$$\approx e^{-n\left(\eta' - \sum_{\alpha=1}^{a} \mu_\alpha \varepsilon'_\alpha\right)} + e^{-n\varepsilon^2} \approx e^{-n\varepsilon^2},$$

из которой следует утверждение теоремы.

Из теоремы 6.2 можно вывести ряд следствий.

Следствие 6.2.1. Если $\mathcal{E} \subset R^*$ P-представляет $x \in \overline{\mathcal{E}}_\sigma^m \subset R$, то

$$N(\mathcal{E}) \geqslant e^{n \sum_{\alpha=1}^{a} \mu_\alpha h(\overline{p}_\alpha)}. \tag{6.25}$$

Доказательство. Пусть $N(\mathcal{E}) < e^{n \sum_{\alpha=1}^{a} \mu_\alpha h(\overline{p}_\alpha)}$; тогда найдется такое малое число $\eta' > 0$, что $N(\mathcal{E}) < e^{n\left(\sum_{\alpha=1}^{a} \mu_\alpha h(\overline{p}_\alpha) - \eta'\right)}$. Но тогда, согласно теореме 2, $\mathcal{E}$ не представляет $x \in \overline{\mathcal{E}}_\sigma^m \subset R$. Полученное противоречие доказывает утверждение следствия 6.2.1.

Следствие 6.2.2. Число M P-различимых x_i $(i = \overline{1, M})$ не может превзойти числа

$$M_C = e^{nC},$$

где

$$C = \max_{\bar{\mu}} \left[h(\bar{\nu}_1) - \sum_{\alpha=1}^{a} \mu_\alpha h(\bar{p}_\alpha) \right] = h(\bar{q}) - \sum_{\alpha=1}^{a} p_\alpha h(\bar{p}_\alpha), \qquad (6.26)$$

$$\bar{\nu}_1 = \bar{\mu}\,\mathbf{p},\ \bar{q} = \bar{p}\,\mathbf{p}.$$

Доказательство. Рассмотрим произвольный (n, M)-код $R' = (x_1, \ldots, x_M) \subset R$. Пусть $\widetilde{\mathscr{E}}_\sigma^{\bar{m}} = R' \cap \mathscr{E}_\sigma^{\bar{m}}$ множество x_i, попавших в $\mathscr{E}_\sigma^{\bar{m}}$. Обозначим $N(\widetilde{\mathscr{E}}_\sigma^{\bar{m}}) = M_{\bar{m}}$. Ясно, что $\sum_{\bar{m}} M_{\bar{m}} = M$. Построим (n, M, P)-код, соотнося x_i непересекающие множества y-ов $\widetilde{\mathscr{E}}_i \subset \mathscr{E}_i = \mathscr{E}_{x_i}^{\mathfrak{M}(\varepsilon)}$, получающиеся перераспределением общих частей $\mathscr{E}_i$. Тогда, если $x_i \in \widetilde{\mathscr{E}}_\sigma^{\bar{m}} \subset \mathscr{E}_\sigma^{\bar{m}} \subset R$, то соответствующие ему $y \in \widetilde{\mathscr{E}}_i \subset \mathscr{E}_i = \mathscr{E}_{x_i}^{\mathfrak{M}(\varepsilon)}$, согласно замечанию в конце предыдущего параграфа, будут принадлежать множеству $\mathscr{E}_\sigma^{\{\bar{m}^*\}_{\bar{\mu}, \varepsilon}}$, определяемому соотношением (6.22). Так как $\widetilde{\mathscr{E}}_i$ не пересекаются, то

$$\mathscr{E}_\sigma^{\{\bar{m}^*\}_{\mu, \varepsilon}} \supseteq \bigcup_{x_i \in \widetilde{\mathscr{E}}_\sigma^{\bar{m}}} \widetilde{\mathscr{E}}_i.$$

Откуда, используя соотношение (6.25) для P-представляющих множеств x_i, имеем

$$N(\mathscr{E}_\sigma^{\{\bar{m}^*\}_{\bar{\mu}, \varepsilon}}) \geqslant \sum_{i=1}^{M_{\bar{m}}} N(\widetilde{\mathscr{E}}_i) \geqslant M_{\bar{m}} e^{n \sum_{\alpha=1}^{a} \mu_\alpha h(\bar{p}_\alpha)}.$$

Полученная оценка верна при любых ε.

В частности, при $\varepsilon \to 0$, когда левая часть оценки согласно (6.23) стремится к $e^{nh(\bar{\nu}_1) + O(\ln n)}$, где $\bar{\nu}_1 = \bar{\mu}\,\mathbf{p}$, имеем

$$e^{nh(\bar{\nu}_1) + O(\ln n)} \geqslant M_{\bar{m}} e^{n \sum_{\alpha=1}^{a} \mu_\alpha h(\bar{p}_\alpha)}.$$

Откуда

$$M_{\bar{m}} \leqslant e^{n\left[h(\bar{\nu}_1) - \sum_{\alpha=1}^{a} \mu_\alpha h(\bar{p}_\alpha)\right] + O(\ln n)}.$$

Суммируя обе части полученного неравенства по всем $\bar{m}$ и учитывая неэкспоненциальную зависимость числа слагаемых от n, получим

$$M = \sum_{\bar{m}} M_{\bar{m}} \leqslant \sum_{\bar{m}} e^{n\left[h(\bar{\nu}_1) - \sum_{\alpha=1}^{a} \mu_\alpha h(\bar{p}_\alpha)\right] + O(\ln n)} \approx e^{n \max_{\bar{\mu}} \left[h(\bar{\nu}_1) - \sum_{\alpha=1}^{a} \mu_\alpha h(\bar{p}_\alpha)\right]}.$$

Откуда следует утверждение следствия.

Величина C, определяемая соотношением (6.26), называется в теории информации пропускной способностью рассматриваемого канала. Она зависит лишь от матрицы $\mathbf{p}$, характеризующей вероятностные свойства шумов в канале.

Следствие 6.2.3. Если удалить из множества $\mathscr{E} = \mathscr{E}_x^{\mathfrak{M}(\varepsilon)} \subset R^*$, P-представляющего $x \in \mathscr{E}_\sigma^{\overline{m}} \subset R$, некоторую его часть $\mathscr{E}'$ с числом элементов

$$N(\mathscr{E}') \leqslant e^{n\left(\sum_{\alpha=1}^{a} \mu_\alpha h(\overline{p}_\alpha) - \eta'\right)},$$

где η' определено согласно (6.24), то оставшаяся часть $\mathscr{E}'' = \mathscr{E} - \mathscr{E}'$ по-прежнему P-представляет x, причем P определяется (6.20).

Доказательство очевидно.

§ 6.5. Достаточные условия P-различимости

В § 6.3 были найдены множества $\mathscr{E}_i = \mathscr{E}_{x_i}^{\mathfrak{M}(\varepsilon)}$, $\mathscr{E}_{x_i}^{\widetilde{\mathfrak{M}}(\varepsilon)}$ P-представляющие $x_i \in \mathscr{E}_\sigma^{\overline{m}}$. Для построения (n, M, P)-кода, состоящего из P-различимых входных слов $x_i (i = \overline{1, M})$, необходимо выполнение условий (6.6), касающихся пересечений $\mathscr{E}_i$ при различных $i = \overline{1, M}$. Исследование пересечений $\mathscr{E}_i$ носит чисто комбинаторный характер (см. гл. 1 п. 1.7.2). Ниже будут сформулированы условия, достаточные для того, чтобы входные слова $x_i (i, = \overline{1, M})$ были P-различимыми. Пусть задан произвольный (n, M)-код $R' = (x_1, \ldots, x_i, \ldots, x_j, \ldots, x_M) \subset R$. Тогда каждое входное слово x_i индуцирует разбиение R вида $R = \bigcup_{\mathbf{m}} \mathscr{E}_{x_i}^{\mathbf{m}}$, где $\mathbf{m} = \| m_{\alpha\alpha'} \| = \mathbf{m}(x_i, x)$ матричное расстояние между x_i и x с элементами $m_{\alpha\alpha'}$, означающими числа пар (α,α') в паре (x_i, x) и $\mathscr{E}_{x_i}^{\mathbf{m}} \subset R$ множество входных слов x, имеющих с x_i расстояние $\mathbf{m}(x_i, x)$, в точности равное $\mathbf{m}$. Пусть $\widetilde{\mathscr{E}}_{x_i}^{\mathbf{m}} = R' \cap \mathscr{E}_x^{\mathbf{m}}$ — множество входных слов $x_j (j \neq i)$, попавших в $\mathscr{E}_{x_i}^{\mathbf{m}}$. Обозначим $N(\widetilde{\mathscr{E}}_{x_i}^{\mathbf{m}}) = M_{x_i}^{\mathbf{m}}$ Ясно, что $\sum_{\mathbf{m}} M_{x_i}^{\mathbf{m}} = M - 1$. Другими словами, $M_{x_i}^{\mathbf{m}}$ — это число $x_j (j \neq i)$ с расстоянием от x_i, в точности равным $\mathbf{m}$. Достаточные условия P-различимости дает следующая теорема.

Теорема 6.3. Пусть (n, M)-код $R' = (x_1, \ldots, x_i, \ldots, x_M)$, где $x_i \in \mathscr{E}_\sigma^{\overline{m}(i)}$ определяет наборы чисел $\{M_{x_i}^{\mathbf{m}}\}$ $(i = \overline{1, M})$; тогда для P-различимости R' при сопоставлении x_i P-представляющих их множеств $\mathscr{E}_{x_i}^{\mathfrak{M}(\varepsilon)}$ достаточно одновременное выполнение M следующих неравенств

$$\sum_{\mathbf{m}} M_{x_i}^{\mathbf{m}} D_{\mathbf{m}}(\mathbf{m}_1^{(0)}, \mathbf{m}_2^{(0)}) \leqslant e^{n\left[\sum_{\alpha=1}^{a} \mu_\alpha^{(i)} h(\overline{\mu}_\alpha, \overline{p}_\alpha) - \eta'\right]}, \quad (i = \overline{1, M}), \qquad (6.27)$$

где η' определяется согласно (6.24) P — согласно (6.20),

$$D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2) = N(\mathscr{E}_{x_1}^{\mathbf{m}_1} \cap \mathscr{E}_{x_2}^{\mathbf{m}_2}),$$

причем $\mathbf{m} = \mathbf{m}(x_1, x_2) = \| m_{\alpha\alpha'} \|$; $\mathbf{m}_1 = \mathbf{m}(x_1, y) = \| m_{\alpha\beta} \|$; $\mathbf{m}_2 = \mathbf{m}(x_2, y) = \| m_{\alpha'\beta} \|$;

$$D_{\mathbf{m}}(\mathbf{m}_1^{(0)}, \mathbf{m}_2^{(0)}) = \max_{\mathbf{m}_1, \mathbf{m}_2 \in [n\,\mathfrak{M}(\varepsilon)]} D(\mathbf{m}_1, \mathbf{m}_2). * \tag{6.28}$$

Доказательство. Имеем для множества $\mathscr{E}_i'$, определяемого (6.5)

$$\mathscr{E}_i' = \mathscr{E}_i \bigcap_{j=i} \mathscr{E}_j = \bigcup_{j \neq i} \mathscr{E}_i \bigcap \mathscr{E}_j,$$

откуда, учитывая возможность пересечений $\mathscr{E}_i \bigcap \mathscr{E}_j$ при различных $j \neq i$, получим

$$N(\mathscr{E}_i') \leqslant \sum_{j \neq i} N(\mathscr{E}_i \bigcap \mathscr{E}_j) =$$

$$= \sum_{\mathbf{m}} M_{x_i}^{\mathbf{m}} N(\mathscr{E}_{x_i}^{\mathfrak{M}(\varepsilon)} \bigcap \mathscr{E}_x^{\mathfrak{M}(\varepsilon)} / \mathbf{m}(x_i, x) = \mathbf{m}) \leqslant$$

$$\leqslant \sum_{\mathbf{m}} M_{x_i}^{\mathbf{m}} D_{\mathbf{m}}(\mathbf{m}_1^{(0)}, \mathbf{m}_2^{(0)}),$$

где $\mathbf{m}_1^{(0)}$ и $\mathbf{m}_2^{(0)}$ определены условием (6.28).

Отсюда, используя (6.27) и следствие 6.2.3, получим утверждение теоремы.

Приведем частный случай выполнения условий (6.27), имеющий самостоятельный интерес для дальнейшего изложения. В самом деле [см. (1.107)], имеет место оценка (не асимптотическая),

$$\sum_{\mathbf{m}} C_n^{\mathbf{m}} \prod_{\alpha\alpha'} (\mu_\alpha \mu_{\alpha'})^{m_{\alpha\alpha'}} D_{\mathbf{m}}(\mathbf{m}_1, \mathbf{m}_2) \leqslant e^{n(h(\mu_1, u_1) + h(\mu_2, u_1))}, \tag{6.29}$$

где

$$C_n^{\mathbf{m}} = \frac{n!}{\prod_{\alpha\alpha'} m_{\alpha\alpha'}!}; \quad u_1 = \left\| \frac{\mu_\alpha^\beta}{\sqrt{\nu_\beta}} \right\|; \quad \mu_1 = \| \mu_\alpha \mu_\alpha^\beta(1) \|;$$

$$\mu_2 = \| \mu_\alpha \mu_\alpha^\beta(2) \| \text{ и } h(\mu_1, u_1) = -\sum_{\alpha\beta} \mu_\alpha \mu_\alpha^\beta(1) \ln \frac{\mu_\alpha^\beta}{\sqrt{\nu_\beta}}.$$

В частности, для $\mathbf{m}_1 = \mathbf{m}_1^{(0)}$ и $\mathbf{m}_2 = \mathbf{m}_2^{(0)}$ имеем, так как $\| \mu_\alpha^\beta(1) \|^{(0)}$ и $\| \mu_\alpha^\beta(2) \|^{(0)} \in \mathfrak{M}(\varepsilon)$

$$h(\mu_1^{(0)}, u_1) = -\frac{1}{2} h(\bar{\nu}, \bar{\nu}_1) + \sum_\alpha \mu_\alpha h(\bar{\mu}_\alpha(1), \bar{\mu}_\alpha) \leqslant$$

$$\leqslant -\frac{1}{2} h(\bar{\nu}_1) + \sum_{\alpha=1}^{a} \mu_\alpha [h(\bar{p}_\alpha) + \varepsilon_\alpha].$$

* Смысл символа $[n\,\mathfrak{M}(\varepsilon)]$ определен на стр 29 (1.59).

Ту же оценку имеет и $h(\mu_2^{(0)}, \mathbf{u}_1)$. Поэтому из (6.29) имеем окончательную оценку

$$\sum_{\mathbf{m}} C_n^{\mathbf{m}} \prod_{\alpha\alpha'} (\mu_\alpha \mu_{\alpha'})^{m_{\alpha\alpha'}} D_{\mathbf{m}}(\mathbf{m}_1^{(0)}, \mathbf{m}_2^{(0)}) \leqslant e^{n\left\{-\left[h(\nu_1) - \sum\limits_{\alpha} \mu_\alpha h(\overline{p}_\alpha)\right] + \sum\limits_{\alpha} \mu_\alpha h(\overline{p}_\alpha) + 2\overline{\varepsilon}\right\}}, \tag{6.30}$$

где

$$\overline{\varepsilon} = \sum_{\alpha=1}^{a} \mu_\alpha \varepsilon_\alpha \quad \text{и} \quad \varepsilon_\alpha = k_\alpha^{-1}(\varepsilon^2/\mu_\alpha).$$

Положим теперь

$$M_{x_i}^{\mathbf{m}} = (M-1) C_n^{\mathbf{m}} \prod_{\alpha\alpha'} (\mu_\alpha \mu_{\alpha'})^{m_{\alpha\alpha'}}, \tag{6.31}$$

$$M = e^{n\left[h(\overline{\nu}_1) - \sum\limits_{\alpha=1}^{a} \mu_\alpha h(\overline{p}_\alpha) - \eta\right]} \tag{6.32}$$

$$\eta \geqslant \eta' + 2\overline{\varepsilon} \geqslant \varepsilon^2 + \overline{\varepsilon}' + 2\overline{\varepsilon}, \tag{6.33}$$

где η' и $\overline{\varepsilon}'$ определены соотношением (6.24). Не будем различать M и $M-1$, что оправдано, так как от правых частей соотношений (6.31) и (6.32) предполагается взятие целой части. Тогда, подставив в (6.28) значения $C_n^{\mathbf{m}} \prod\limits_{\alpha\alpha'} (\mu_\alpha \mu_{\alpha'})^{m_{\alpha\alpha'}}$, выраженные через $M_{x_i}^{\mathbf{m}}$ и M, согласно (6.31) и (6.32) получим окончательную оценку

$$\sum_{\mathbf{m}} M_{x_i}^{\mathbf{m}} D_{\mathbf{m}}(\mathbf{m}_{(1)}^{(0)}, \mathbf{m}_{(2)}^{(0)}) \leqslant e^{n\left[\sum\limits_{\alpha=1}^{a} \mu_\alpha h(\overline{p}_\alpha) - \eta'\right]},$$

что указывает, согласно теореме 6.3, на P-различимость соответствующих x_i $(i = \overline{1, M})$.

Сформулируем теперь более «грубые», чем приведенные выше, достаточные условия P-различимости, использующие хэмминговское расстояние между словами длины n (см. п. 1.7.2).

В самом деле, рассмотрим случай $a = b$ (если $a \neq b$, то соответствующей группировкой можно прийти к рассматриваемому случаю). Пусть дан (n, M)-код

$$R' = (x_1, \ldots, x_i, \ldots, x_j, \ldots, x_M) \subset R$$

с хэмминговским расстоянием

$$d = d(\widetilde{\mathbf{m}}) = d(x_i, x_j) = \sum_{\alpha \neq \alpha'} m_{\alpha\alpha'} = n - \sum_{\alpha} m_{\alpha\alpha}$$

между входными словами, где $m_{\alpha\alpha'}$ число пар (α, α') в паре (x_i, x_j).

Аналогично определяется хэмминговское расстояние $d = d(\mathbf{m}) = d(x, y)$ между входным x и выходным $y \in R^*$ словами длины n.

Рассмотрим множества $[n\mathfrak{M}'(\varepsilon)]$ матричных расстояний $\mathbf{m} = \mathbf{m}(x, y)$, для которых $d(\mathbf{m}) \leqslant [n\varepsilon] = d_0$. Определим множество y-ов (см. п. 1.7.2)

$$\mathscr{G}_x^{\mathfrak{M}'(\varepsilon)} = \bigcup_{\mathbf{m}(x,y) \in [n\mathfrak{M}'(\varepsilon)]} \mathscr{G}_x^{\mathbf{m}(x,y)} = \bigcup_{d=0}^{d_0} \mathscr{G}_x^{d(x,y)}.$$

Примем теперь в качестве множеств $\mathscr{E}_i \subset R^*$ соответствующего (n, M, P)-кода построенные множества $\mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)} = \mathscr{E}_i$. Оценим прежде всего вероятность

$$P = \mathscr{P}(\mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}/x_i) \tag{6.34}$$

в предположении стремления матрицы вероятностей переходов символов $\mathbf{p}$ к единичной матрице E (случай низкого уровня шумов, когда $p_\alpha^\alpha \to 1$, $\alpha = \overline{1, a}$ и $C \to \ln a$). Рассмотрим, кроме того, случай $n \to \infty$.

Выше уже отмечалось, что матрица $\mathbf{m} = \mathbf{m}(x, y) = \| m_{\alpha\beta} \| = (\bar{m}_\alpha)$ распадается на a независимых полиномиально распределенных с параметраметриями $\bar{p}_\alpha$ векторов $\bar{m}_\alpha (\alpha = \overline{1, a})$, если $m_\alpha = \sum\limits_\beta m_{\alpha\beta} = \mu_\alpha n$, где $\mu_\alpha > 0$. В рассматриваемом случае $p_\alpha^\alpha \to 1$ $(\alpha = \overline{1, a})$ (см. п. 1.6.3), последние распадаются на независимые, распределенные по Пуассону с параметрами λ_α^β компоненты $m_{\alpha\beta}$ $(\alpha \neq \beta)$, если $p_\alpha^\beta = \lambda_\alpha^\beta / n \to 0$ $(\alpha \neq \beta)$. Тогда сумма $d = d(\mathbf{m}) = d(x, y) = \sum\limits_{\alpha \neq \beta} m_{\alpha\beta}$ снова имеет распределение Пуассона с параметром $\lambda = \sum\limits_{\alpha \neq \beta} \lambda_\alpha^\beta$. Но в этом случае, согласно асимптотическому соотношению (1.86), имеем для вероятности (6.34)

$$P = \mathscr{P}(\mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}/x_i) = 1 - e^{-\varepsilon n \ln n + O(n)} \sim 1 - e^{-\varepsilon n \ln n}.$$

Теорема 6.4. Пусть (n, M, P)-код $R' = (x_1, \ldots, x_i, \ldots, x_j, \ldots x_M)$, у которого входному слову x_i сопоставляется множество выходных слов $\mathscr{E}_i = \mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}$. Тогда при $\mathbf{p} \to \mathrm{E}$ $(p_\alpha^\alpha \to 1, \alpha = \overline{1, a})$ и $n \to \infty$ достаточным условием P-различимости R' является выполнение условий

$$d(x_i, x_j) > 2[n\varepsilon] \quad (1 \leqslant i < j \leqslant M),$$

или в терминах частот $M_{x_i}^d \equiv 0$ для $d \leqslant 2[n\varepsilon]$ $(i = \overline{1, M})$. При этом

$$P = 1 - e^{-\varepsilon n \ln n + O(n)} \sim 1 - e^{-\varepsilon n \ln n}. \tag{6.35}$$

Доказательство. Условие теоремы согласно следствию 1.2.1 приводит к непересекающимся множествам $\mathscr{E}_i = \mathscr{E}_{x_i}$, а это вместе со стремлением вероятности P к единице при $n \to \infty$ приводит к утверждению теоремы.

Заметим, что определение P-различимости допускает не пустые пересечения множеств $\mathscr{E}_i$ P-представляющих x_i, требуется лишь, чтобы их пересечения не представляли x_i. В указанных в теореме 6.4 достаточных условиях P-различимости пересечения множеств $\mathscr{E}_i$ оказываются пустыми, что достаточно для P-различимости соответствующих x_i. Такое требование приводит к тому, что условия теоремы 6.4, в общем случае дискретного канала, приводят к неоптимальному в смысле Шеннона кодированию, так как число входных слов M не может иметь экспоненциального порядка, близкого к предельно большому экспоненциальному порядку $M_C = e^{nC}$ (см. оценки M в бинарном случае гл. 8 (8.42)).

Лишь в канале с малым уровнем шумов достаточные условия P-различимости теоремы 6.4 приводят к оптимальному по Шеннону кодированию.

§ 6.6. Эквивалентность обеих процедур декодирования

Приведенная в предыдущей главе статистическая процедура оптимального декодирования, основанная на отношении правдоподобия, эквивалентна разбиению пространства R^* выходных слов на M, вообще говоря, пересекающихся множеств $\mathscr{E}_i$ $(i = \overline{1,M})$.

Попадание выходного слова в одно из них приводит к принятию решения о передаче соответствующего входного слова. Попадание в их пересечения приводит к неопределенности, которую мы уже весьма произвольно разрешили в § 6.2. С этими же обстоятельствами связана и эквивалентность обоих методов оптимального декодирования, приведенных в предыдущей и в этой главе.

В самом деле, зафиксируем (n, M)-код на входе канала. Оба из указанных способов декодирования сопоставляют входным словам $\{x_i\}$ $(i = \overline{1,M})$ каждый свою последовательность соответствующих множеств выходных слов $\{\mathscr{E}_i\}$ и $\{\mathscr{E}'_i\}$ $(i = \overline{1,M})$. Последние P-представляют x_i. Но если два каких-либо множества $\mathscr{E}_i$ и $\mathscr{E}'_i$ P-представляют x, то легко показать, что их пересечение $\mathscr{E}^*_i = \mathscr{E}_i \cap \mathscr{E}'_i$ снова будет P-представлять x $(\mathscr{E}^*_i = R^* - \overline{\mathscr{E}}_i \cup \overline{\mathscr{E}}'_i)$. В самом деле, пусть пересечение $\mathscr{E}^*_i = \mathscr{E}_i \cap \mathscr{E}'_i$ не P-представляет x, то есть соответствующая вероятность $P(\mathscr{E}^*_i/x) \underset{n\to\infty}{\not\to} 0$. Тогда множества y-ов $\overline{\mathscr{E}_i - \mathscr{E}^*_i}$ и $\overline{\mathscr{E}'_i - \mathscr{E}^*_i}$ не будут пересекаться и вероятности попадания y в них будут по-прежнему асимптотически P-представлять x-ы и $P \not\to 1$ с ростом n. Но тогда вероятность попадания входного слова в одно из указанных множеств стремится с ростом n к двум, что приводит к противоречию.

Поэтому пересечение P-представляющих один и тот же x снова P-представляет. Далее, если даны два способа декодирования $\{\mathscr{E}_i\}$ и $\{\mathscr{E}'_i\}$, каждый из которых состоит из множеств y-ов, P-представляющих соответствующие $\{x_i\}$, то соответствующие пересечения $\{\mathscr{E}_i \cap \mathscr{E}'_i\}$ снова будут P-представлять $\{x_i\}$. Поэтому оба способа декодирования в точности эквивалентны, если в качестве множеств выходных слов взять $\{\mathscr{E}^*_i\} = \{\mathscr{E}_i \cap \mathscr{E}'_i\}$. Они асимптотически при $n \to \infty$ эквивалентны, если брать в качестве способа декодирования сами $\{\mathscr{E}_i\}$ и $\{\mathscr{E}'_i\}$, так как последние P-представляют $\{x_i\}$ как и их пересечение.

Глава 7

ПОЛУЧЕНИЕ ОПТИМАЛЬНОГО КОДА СЛУЧАЙНЫМ ВЫБОРОМ

§ 7.1. Вводные замечания

В этой главе завершается конструкция оптимального в смысле Шеннона кода для рассматриваемого дискретного канала, начатая в предыдущей главе. Используемое в теории информации Шеннона как способ доказательства существования оптимального кода так называемое случайное кодирование здесь применяется в качестве способа построения конкретного оптимального кода. Задача сводится к вычислению вероятности P' того, что предлагаемый случайный алгоритм построения приведет к цели. Далее будет показано, что с ростом длины n входных слов, выбираемых из определенной генеральной совокупности, вероятность P' существенно быстрее стремится к единице, чем, например, вероятность правильного декодирования P. Поэтому практически можно считать, что предлагаемый алгоритм вполне достоверно приводит к цели. Использование хорошо известных регулярных способов получения случайных чисел приводит к регулярному варианту построения случайного кода.

В § 7.2 приводится оценка вероятности P' для наиболее интересного случая, исключающего лишь случай низкого уровня шумов. Для последнего случая оценка вероятности P' приводится в § 7.3. Там же приводится алгоритм случайного выбора входных слов с последующим удалением части «плохих» входных слов.

В § 7.4 систематизируются полученные результаты и приводятся описания алгоритмов построения оптимальных кодов.

Наконец, в § 7.5. приводятся известные общие результаты теории информации Шеннона, а также ее частные результаты для рассматриваемого канала. Производится также сравнение этих результатов с результатами, приведенными в данной книге.

§ 7.2. Оценка вероятности получения P-различимых входных слов случайным выбором (исключая случай низкого уровня шумов)

До сих пор не приводилось конкретного способа получения P-различимых входных слов $\{x_i\}$. Ниже приводится описание такого способа. Достаточные условия P-различимости формулируются не для самих входных слов x, а для частот $\{M^{\mathbf{m}}_{x_i}\}$ их взаимных матричных расстояний $\mathbf{m} = \mathbf{m}(x_i, x_j) = \| m_{\alpha\alpha'} \|$. Отсюда следует, что P-различимых совокупностей $\{x_i\}$ [(n, M, P)-кодов] много. Нам достаточно получить хотя бы одну из них.

В рассмотренном выше примере (§ 6.5 предыдущей главы) из соотношения (6.31) следует, что числа $M^{\mathbf{m}}_{x_i}$ являются средними значениями «частот» пар входных слов (x_i, x_j) с заданным матричным расстоянием $\mathbf{m} = \mathbf{m}(x_i, x_j)$, когда символы α ($\alpha = \overline{1, a}$) слов выбираются из генеральной совокупности

с вероятностями символов α, равными $\mu_\alpha (\alpha = \overline{1, a})$. Это обстоятельство наводит на мысль, что если осуществить однократную выборку M входных слов $x_i (i = \overline{1, M})$ из указанной генеральной совокупности, то получившаяся при этом конкретная реализация частот $\{M^{\mathbf{m}}_{x_i}\}$, так же как и ее средние значения в рассмотренном примере, будет удовлетворять с некоторой вероятностью $P' \approx 1$ условиям (6.27). Оказывается, что это имеет место на самом деле. Оценку вероятности P' дает следующая теорема.

Теорема 7.1. Если выбрать $M = e^{n(C-\eta)}$ входных слов $x_i (i = \overline{1, M})$ длины n из генеральной совокупности с вероятностями p_α появления символа α, определяемыми соотношением (6.26), и сопоставлять им множества y-ов $\mathfrak{S}^{\mathfrak{M}(\varepsilon)}_{x_i}$, то вероятность P' выполнения для частот $\{M^{\mathbf{m}}_{x_i}\}$ M неравенств (6.27) имеет вид

$$P' \geqslant P'' \approx 1 - e^{-Mk_{\overline{\mathcal{P}}, \overline{D}}(\Delta)},$$

где $\overline{\mathcal{P}} = \left\{C^{\mathbf{m}}_n \prod_{\alpha\alpha'} (p_\alpha p_{\alpha'})^{m_{\alpha\alpha'}}\right\}$; $\overline{D} = \{D_{\mathbf{m}}(\mathbf{m}^{(0)}_1, \mathbf{m}^{(0)}_2\}$; $\Delta > 0$; функция $k_{\overline{\mathcal{P}}\overline{D}}(\Delta)$ определена (1.69).

Эта оценка имеет место для случая

$$C = h(\overline{q}) - \sum_{\alpha=1}^{a} p_\alpha h(\overline{p}_\alpha) < 2 \sum_{\alpha=1}^{a} p_\alpha h(\overline{p}_\alpha) \tag{7.1}$$

и

$$\eta \geqslant \varepsilon^2 + \overline{\varepsilon}' + 2\overline{\varepsilon},$$

где ε, $\overline{\varepsilon}'$ и $\overline{\varepsilon}$ определены соотношением (6.33).

Доказательство. Будем составлять входные слова $x_i (i = \overline{1, M})$ длины n из символов $\alpha (\alpha = \overline{1, a})$, выбираемых из генеральной совокупности $G(\overline{p})$ с вероятностями p_α символов α, где $\overline{p}$ определяются соотношениями (6.26). Тогда числа $\{M^{\mathbf{m}}_{x_i}\}$ оказываются полиномиально распределенными случайными величинами с параметрами $\overline{\mathcal{P}} = \left\{C^{\mathbf{m}}_n \prod_{\alpha\alpha'} (p_\alpha p_{\alpha'})^{m_{\alpha\alpha'}}\right\}$.

Случайная величина

$$Z_i = \sum_{\mathbf{m}} M^{-1} M^{\mathbf{m}}_{x_i} D_{\mathbf{m}}(\mathbf{m}^{(0)}_1, \mathbf{m}^{(0)}_2),$$

оценивающая сверху случайную величину $M^{-1}N(\mathscr{E}'_i)$ (см. теорему 6.3), оказывается линейной функцией с коэффициентами $\overline{D} = \{D_{\mathbf{m}}(\mathbf{m}^{(0)}_1, \mathbf{m}^{(0)}_2)\}$ полиномиально распределенных частот $\{M^{\mathbf{m}}_{x_i}\}$. В этом случае при $M \to \infty$ (что имеет место при $n \to \infty$) к ее функции распределения применима асимптотическая оценка (см. § 6.3 и п. 1.6.2).

В самом деле, используя (6.30) гл. 6, при $\overline{\mu} = \overline{p}$ получим

$$\mathrm{E}\, Z_i = \sum_{\mathbf{m}} C^{\mathbf{m}}_n \prod_{\alpha\alpha'} (p_\alpha p_{\alpha'})^{m_{\alpha\alpha'}} D_{\mathbf{m}}(\mathbf{m}^{(0)}_1, \mathbf{m}^{(0)}_2) \leqslant e^{n\left[-C + \sum_{\alpha=1}^{a} p_\alpha h(\overline{p}_\alpha) + 2\overline{\varepsilon}\right]}, \tag{7.2}$$

где C определено (6.26).

Положим $D = \mathrm{E}\, Z_i + \Delta$, где $\Delta \approx \mathrm{E}\, Z_i > 0$. Тогда согласно соотношению (1.71)

$$P'_i = \mathcal{P}\left(Z_i = \sum_{\mathbf{m}} M^{-1} M^{\mathbf{m}}_{x_i} D_{\mathbf{m}}(\mathbf{m}^{(0)}_1, \mathbf{m}^{(0)}_2) \leqslant D \approx \mathrm{E}\, Z_i\right) \approx 1 - e^{-Mk_{\overline{\mathcal{P}}, \overline{D}}(\Delta)} \tag{7.3}$$

Положим $M = e^{n[C-\eta]}$, где η определено соотношением (6.33) гл. 6. Тогда с учетом соотношения (7.2), с той же вероятностью P'_i, из (7.3) получим:

$$P'_i = \mathcal{P}\left[\sum_{\mathbf{m}} M^{\mathbf{m}}_{x_i} D_{\mathbf{m}}(\mathbf{m}_1^{(0)}, \mathbf{m}_2^{(0)}) \leqslant DM \leqslant e^{n\left[\sum_{\alpha=1}^{a} p_\alpha h(\bar{p}_\alpha) - \eta'\right]}\right] \approx 1 - e^{-Mk_{\overline{\mathcal{P}}, \overline{D}}(\Delta)}, \tag{7.4}$$

где η' определено соотношением (6.24).

Для оценки вероятности P' одновременного выполнения M неравенств (7.4) при $i = \overline{1, M}$ воспользуемся неравенством Буля, так как случайные величины Z_i при различных i являются зависимыми. Имеем

$$P' \geqslant \sum_{i=1}^{M} P'_i - (M-1) = 1 - \sum_{i=1}^{M}(1-P_i) \approx 1 - Me^{-Mk_{\overline{\mathcal{P}}, \overline{D}}(\Delta)} \approx 1 - e^{-Mk_{\overline{\mathcal{P}}, \overline{D}}(\Delta)}. \tag{7.5}$$

Соотношение (7.5) всегда приводит к вероятности P', стремящейся к единице, если $\mathrm{E}Z_i \approx \Delta \to \infty$, при $n \to \infty$. Если $\mathrm{E}Z_i \approx \Delta \to 0$ при $n \to \infty$, то P' стремится к единице, если $M(\mathrm{E}Z_i)^2 \to \infty$. Однако, если $M(\mathrm{E}Z)^2 \to 0$, то $P' \to 0$. Это следует из того, что при $\Delta \to 0$, $k_{\overline{\mathcal{P}}_D}(\Delta) = O(\Delta^2)$ [см. (1.84)].

Приведенные соображения сводятся к ограничению $C < 2\sum_{\alpha} p_\alpha h(\bar{p}_\alpha)$, указанному в условии (7.1) теоремы 7.1. Таким образом, теорема 7.1 доказана полностью.

Заметим, что ограничения $C < 2\sum_{\alpha} p_\alpha h(\bar{p}_\alpha)$ не касаются наиболее интересного случая высокого уровня шумов. В этом случае $\mathbf{p} \to \bar{e}\,\bar{r}$, где $h(\bar{r}) > 0$, здесь рассматривается поэлементное стремление матриц, и неравенство (7.1) эквивалентно требованию $C < 2h(\bar{r})$, которое всегда выполняется из-за того, что $C \to 0$. При низком уровне шумов, когда $\mathbf{p} \to E$, что означает $p^{\alpha}_{\alpha} \to 1$ ($\alpha = \overline{1, a}$), $h(\bar{p}_\alpha) \to 0$ и $C \to \ln a$, условия теоремы 7.1 не выполняются.

§ 7.3. Оценка вероятности получения P-различимых входных слов случайным выбором для случая низкого уровня шумов

Рассмотрим низкий уровень шумов. В этом случае $\mathbf{p} \to E$, что означает $p^{\alpha}_{\alpha} \to 1$ ($\alpha = \overline{1, a}$), откуда

$$C = \max_{\bar{\mu}}\left[h(\bar{\mu}\mathbf{p}) - \sum_{\alpha} \mu_\alpha h(\bar{p}_\alpha)\right] \to \max_{\bar{\mu}} h(\bar{\mu}) = h(1/a, \ldots, 1/a) = \ln a,$$

где максимум достигается при $\bar{\mu} = \bar{p}^{(0)} = (1/a, \ldots, 1/a)$.

Выберем M входных слов x длины n

$$R' = (x_1, \ldots, x_i, \ldots, x_j, \ldots, x_M) \subset R$$

из генеральной совокупности равновероятных символов $\alpha = \overline{1, a}$ с вероятностями $\bar{p}^{(0)} = (p_\alpha) = (1/a, \ldots, 1/a)$. При этом среди $x_i \in R'$ могут быть, вообще говоря, повторяющиеся входные слова.

Выделим из множества R' подмножество $R'' \subset R'$, состоящее из $M' \leqslant M$ входных слов с взаимными расстояниями Хэмминга $d(x_i, x_j)$, удовлетворяющими условию $d(x_i, x_j) > d_0$.

Аналогично множеству $\mathscr{E}_x^{\mathfrak{M}'(\varepsilon)} \subset R^*$ выходных слов y, введенному в § 6.5, определим множество $\mathscr{E}_{x'}^{\mathfrak{M}'(\varepsilon)} \subset R$ входных слов x с расстоянием Хэмминга $d(x', x)$ от x', удовлетворяющим условию $0 \leqslant d(x', x) \leqslant [n\varepsilon] = d_0$.

Будем называть множество $\mathscr{E}_{x'}^{\mathfrak{M}'(\varepsilon)}$ ε-окрестностью входного слова x'. Соотнесем каждому $x_i \in R'$ его ε-окрестность $\mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}$. Ясно, что при этом число всех ε-окрестностей равно числу M всех $x_i \in R'$.

Разобьем совокупность всех ε-окрестностей на два класса. К первому классу отнесем все ε-окрестности $\mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}$, не содержащие, кроме x_i, ни одного входного слова из R' (пустые окрестности). Ко второму классу отнесем все остальные ε-окрестности. Из сказанного выше следует, что всего имеется M' ε-окрестностей первого класса и $M - M'$ ε-окрестностей второго класса.

Лемма 7.1. Пусть $M = e^{nH}$ $(0 \leqslant H \leqslant \ln a)$ входных слов x_i $(i = \overline{1, M})$ выбрано из генеральной совокупности с равновероятными символами $\alpha = \overline{1, a}$. Тогда вероятность P' того, что среди x_i $(i = \overline{1, M})$ имеется $M' \geqslant M(1 - e^{-n\gamma})$ $(\gamma > 0)$ входных слов с взаимными расстояниями $d(x_i, x_j) \geqslant d_0 = [n\varepsilon]$ $\left(0 < \varepsilon < \dfrac{a-1}{a}\right)$, удовлетворяет неравенству

$$P' = \mathscr{P}[M' \geqslant M(1 - e^{-n\gamma})] \geqslant 1 - e^{-n\left[h\left(\varepsilon, \frac{a-1}{a}\right) - h(\varepsilon) - H - \gamma\right] + O(\ln n)}.$$

Доказательство. Зафиксируем x_i и некоторую его ε-окрестность $\mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}$. Тогда вероятность Q попадания в нее одного $x_j \in R'$ $(j \neq i)$ не зависит от x_i и равна при $0 \leqslant \varepsilon < (a-1)/a$

$$Q = \mathscr{P}(\mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}/x_i) = \mathscr{P}(0 \leqslant d(x_i, x_j) \leqslant d = [n\varepsilon]) \approx e^{-n\left[h\left(\varepsilon, \frac{a-1}{a}\right) - h(\varepsilon)\right]}, \quad (7.6)$$

где $h(u, v) = -u \ln v - (1-u)\ln(1-v)$; $h(u) = h(u, u)$.

В самом деле, по определению $d(x_i, x_j) = \sum\limits_{\alpha \neq \alpha'} m_{\alpha\alpha'} = n - \sum\limits_{\alpha=1}^{a} m_{\alpha\alpha}$, где $m_{\alpha\alpha}$ — случайное число пар символов (α, α) в паре (x_i, x_j). В силу независимости x_i и x_j случайное число $m_{\alpha\alpha}$ имеет биномиальное распределение

$$P_{m_\alpha, 1/a^2}(m_{\alpha\alpha}) = C_{m_\alpha}^{m_{\alpha\alpha}} \left(\frac{1}{a^2}\right)^{m_{\alpha\alpha}} \left(1 - \frac{1}{a^2}\right)^{m_\alpha - m_{\alpha\alpha}}.$$

В силу независимости $m_{\alpha\alpha}$ при различных $\alpha = \overline{1, a}$ их сумма $l = \sum\limits_{\alpha-1}^{a} m_{\alpha\alpha}$ имеет биномиальное распределение $P_{n, \frac{1}{a}}(l)$ с параметрами $\sum\limits_{\alpha=1}^{a} m_\alpha = n$ и $\sum\limits_{\alpha=1}^{a} 1/a^2 = \dfrac{1}{a}$.

Поэтому случайное число $d = n - l$ имеет биномиальное распределение $P_{n, \frac{a-1}{a}}(d)$ и при $d_0 < \left[n\,\dfrac{a-1}{a}\right]$

$$Q = \mathscr{P}(\mathscr{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}/x_i) = \mathscr{P}(0 \leqslant d(x_i, x_j) \leqslant d_0 = [n\varepsilon]) =$$
$$= \sum_{d=0}^{d_0} P_{n, \frac{a-1}{a}}(d) \approx e^{-n\left[h\left(\varepsilon, \frac{a-1}{a}\right) - h(\varepsilon)\right]},$$

что и доказывает соотношение (7.6).

Далее, вероятность попадания одного $x_j\,(j \neq i)$ вне ε-окрестности x_i равна $1-Q$. Вероятность того, что все $M-1$ $x_j\,(j \neq 1)$ попадут вне ε-окрестности x_i, равна $(1-Q)^M \geqslant 1-MQ$, при $MQ < 1$, что всегда имеет место при достаточно малом ε. Наконец, вероятность Q' того, что хотя бы одно из $x_j\,(j \neq i)$ попадет в ε-окрестность, x_i равна

$$Q' = 1-(1-Q)^M \leqslant MQ = e^{-n\left[h\left(\varepsilon, \frac{a-1}{a}\right)-h(\varepsilon)-H\right]+O(\ln n)}. \tag{7.7}$$

Из-за возможных пересечений различных ε-окрестностей x_i числа попавших в них x_j являются зависимыми случайными величинами. Поэтому в частности зависимы и случайные величины ξ_i, равные нулю, если ε-окрестность x_i пуста, и равные единице в противном случае. Ясно, что

$$\mathrm{E}\xi_i = 0\,(1-Q') + 1\cdot Q' = Q' \leqslant MQ.$$

Для случайного числа $M - M' = \sum_{i=1}^{M} \xi_i$, несмотря на зависимость величин ξ_i, имеем

$$\mathrm{E}\,(M-M') = M - \mathrm{E}\,M' = \sum_{i=1}^{M} \mathrm{E}\,\xi_i = \sum_{i=1}^{M} Q' = MQ' = M^2Q.$$

Так как $M - M' \geqslant 0$ и $\mathrm{E}\,(M-M') < \infty$, то, используя неравенство Чебышева, имеем для $t > 0$:

$$\begin{aligned}\mathcal{P}\,[M-M' < t\,\mathrm{E}\,(M-M')] &= \mathcal{P}\,[M' \geqslant M - t\,\mathrm{E}\,(M-M')] = \\ &= \mathcal{P}\,[M' \geqslant M\,(1-tQ')] \geqslant 1-t^{-1}.\end{aligned} \tag{7.8}$$

Положим $t = e^{-n\gamma}/Q'$; тогда, используя (7.7), усилим неравенство (7.8)

$$\begin{aligned}\mathcal{P}\,[M-M' < t\,\mathrm{E}\,(M-M')] &= \mathcal{P}\,[M' \geqslant M - t\,\mathrm{E}\,(M-M')] = \\ &= \mathcal{P}\,[M' \geqslant M\,(1-tQ')] \geqslant 1-t^{-1},\end{aligned}$$

что и доказывает лемму.

Используя теорему 6.4 и лемму 7.1, можно оценить вероятность P-различимости в случае низкого уровня шумов.

Теорема 7.2. Для случая низкого уровня шумов выбор $M = e^{nH}$ $(0 \leqslant H < \ln a)$ входных слов $x_i\,(i = \overline{1, M})$ длины n из генеральной совокупности с равновероятными символами, выбрасывание из них x_i с непустыми 2ε-окрестностями $(0 < \varepsilon < (a-1)/2a)$ и сопоставление оставшихся после этого $M' \leqslant M$ входных слов x_i с множествами выходных слов $\mathcal{E}_{x_i}^{\mathcal{M}'(\varepsilon)}$ приводит к вероятности P' получения $M' \geqslant M(1-e^{-n\gamma})$ P-различимых входных слов, имеющей оценку

$$P' = \mathcal{P}\,[M' \geqslant M\,(1-e^{-n\gamma})] \geqslant 1 - e^{-n\left[h\left(2\varepsilon, \frac{a-1}{a}\right)-h(2\varepsilon)-H-\gamma\right]+O(\ln n)}, \tag{7.9}$$

причем

$$P = 1 - e^{-\varepsilon n \ln n + O(n)}. \tag{7.10}$$

Доказательство. В самом деле, выберем $M = e^{nH}$ входных слов длины n из генеральной совокупности с равновероятными символами $\alpha = \overline{1, a}$. Выбросим все входные слова x_i, у которых хотя бы с одним из других выбранных входных слов $x_j (i \neq j)$ расстояние Хэмминга равно $d(x_i, x_j) \leqslant [2\varepsilon n]$.

Тогда, согласно лемме 7.1, вероятность P' того, что при этом останется $M' \geqslant M(1-e^{-n\gamma})$ входных слов, оценивается неравенством (7.9). Соотнесем оставшимся входным словам множества выходных слов $\mathcal{E}_{x_i}^{\mathfrak{M}'(\varepsilon)}$. Тогда, ввиду того что оставшиеся входные слова удовлетворяют условию $d(x_i, x_j) > [2\varepsilon n]$, согласно теореме 6.4 они оказываются P-различимыми, при этом вероятность правильного декодирования согласно теореме 6.4 имеет вид, указанный соотношением (7.10). Таким образом, теорема доказана полностью.

Заметим, что для более быстрого по сравнению с P стремления P' к единице с ростом n достаточно требовать неравенства

$$H < \ln a - \gamma - 2\varepsilon\left[\ln\frac{na}{2\varepsilon} + (1-2\varepsilon)^{-1}\right],$$

из которого следует, что коэффициент при n в соотношении (7.9) больше соответствующего коэффициента в соотношении (7.10).

Если положить $\varepsilon = O((\ln n)^{-1-\varkappa})$, где $\varkappa > 0$, то неравенство примет вид

$$H < \ln a - \gamma + O[(\ln n)^{-\varkappa}]. \tag{7.11}$$

Вместе с тем при достаточно малом γ и ε (большем n) из соотношения (7.11) следует, что H может быть сколь угодно близко к $C = \ln a$. Очевидно, что при $H > C$ теорема 7.2 не имеет места.

§ 7.4. Теоремы о построении оптимального кода и их обсуждение

Результаты теорем 7.1 и 7.2 приводят к процедуре построения оптимального кода. Описание этой процедуры и связанных с ней оценок содержит следующая теорема.

Теорема 7.3. Пусть шумы в дискретном постоянном канале с нулевой памятью заданы $(a \times b)$-матрицей вероятностей переходов символов $\mathbf{p} = \| p_\alpha^\beta \|$.

Выберем из генеральной совокупности $G(\bar{p})$, в которой символ α возникает с вероятностью p_α, $M = e^{nH}$ входных слов длины $n(x_1, \ldots, x_M)$, где $H = C - \eta$,

$$C = h(\bar{q}) - \sum_{\alpha=1}^{a} p_\alpha h(\bar{p}_\alpha) = \max_{\bar{\mu}}(h(\bar{\nu}_1) - \sum_{\alpha} \mu_\alpha h(\bar{p}_\alpha)) < 2\sum_{\alpha} p_\alpha h(\bar{p}_\alpha), \tag{7.12}$$

причем $\bar{\nu}_1 = \bar{\mu}\mathbf{p}$; $\bar{q} = \bar{p}\mathbf{p}$; $\eta \geqslant \varepsilon^2 + \bar{\varepsilon}' + 2\bar{\varepsilon}$; $\bar{\varepsilon}' = \sum_{\alpha} p_\alpha \varepsilon'_\alpha$; $\bar{\varepsilon} = \sum_{\alpha} p_\alpha \varepsilon_\alpha$; $-\varepsilon'_\alpha$ и ε_α — корни уравнения $k_{\bar{p}_\alpha \bar{q}_\alpha}(u) = \varepsilon^2 / p_\alpha$ и ε — произвольное положительное число.

Сопоставим x_i множества $\mathcal{E}_{x_i}^{\mathfrak{M}(\varepsilon)}$, определенные в § 6.3. Тогда, согласно теореме 7.1, вероятность P' того, что таким образом построен (n, M, P)-код, имеет оценку:

$$P' \geqslant P'' \approx 1 - e^{-Mk_{\bar{\mathcal{P}},\bar{D}}(\Delta)}, \quad (\Delta > 0), \tag{7.13}$$

где $\bar{\mathcal{P}}$ и $\bar{D}$ определены в теореме 7.1. При этом вероятность P правильного декодирования (n, M, P)-кода

$$P \approx 1 - e^{-\varepsilon^2 n}$$

и его оптимальность следует из того, что, при $n \to \infty$ и $\varepsilon \to 0$, $M \to M_c = e^{nC}$ и $P \to 1$.

Если $H > C$, то такого кода построить нельзя.

В случае низкого уровня шумов ($p_\alpha^\alpha \to 1$, $\alpha = \overline{1, a}$), когда условия (7.12) не выполняются, может быть осуществлена другая конструкция оптимального кода. В этом случае из генеральной совокупности $G(\bar{p}^{(0)})$, где символы $\alpha = \overline{1, a}$ возникают равновероятно, выбирается $M = e^{nH}$ $(0 < H < \ln a)$ входных слов длины n, далее из них удаляются входные слова с непустыми 2ε-окрестностями $\left(\varepsilon < \frac{a-1}{2a}\right)$ (расстояние измеряется по Хэммингу). Оставшимся после этого входным словам x_l сопоставляются множества $\mathscr{E}_{x_l}^{\mathfrak{M}'(\varepsilon)}$ выходных слов, определенные в § 6.5.

Тогда, согласно теореме 7.2, вероятность P' того, что после этой процедуры будет получен (n, M', P)-код, состоящий из $M' \geqslant M(1 - e^{-n\gamma})$ $(\gamma > 0)$ входных слов, с вероятностью правильного декодирования

$$P = 1 - e^{-\varepsilon n \ln n + O(n)},$$

имеет оценку

$$P' \geqslant 1 - e^{-n\left(h\left(2\varepsilon, \frac{a-1}{a}\right) - h(\varepsilon) - H - \gamma\right)}. \tag{7.14}$$

Его оптимальность следует из того, что при $n \to \infty$ и $\varepsilon = O[(\ln n)^{-\varkappa}] \to 0 (\varkappa > 0)$ $M' \to M \to M_C = e^{n \ln a}$ и $P \to 1$. В случае $H > C = \ln a$ такого кода нельзя построить.

Прокомментируем полученные результаты. Прежде всего укажем на то, что ограничение на $C < 2\sum_\alpha p_\alpha h(\bar{p}_\alpha)$ в теореме 7.1, сохранившееся в теореме 7.2, связано с оценкой вероятности P' (7.13) получения оптимального кода указанной процедурой. Это ограничение не касается наиболее интересного случая высокого уровня шумов в канале. В случае низкого уровня шумов, когда $C > 2\sum_\alpha p_\alpha h(\bar{p}_\alpha)$, во второй части теоремы 7.3 дана оценка аналогичной вероятности P' (7.14), имеющая другую форму.

Важно отметить, что приведенный в теореме 7.3 алгоритм построения оптимального (по Шеннону) кода содержит элемент случайности в выборе входных слов длины n (кодирование). Декодирование проводится регулярными методами.

Таким образом, алгоритм в целом приводит к цели лишь с некоторой вероятностью P'. Однако в рассматриваемом случае с ростом n P' стремится к единице существенно быстрее, чем допустимая вероятность правильного декодирования P [см. (7.13) и (7.11)]. Поэтому в этом случае практически можно принять $P' = 1$.

Противопоставление приведенного случайного алгоритма построения регулярному алгоритму не столь важно. В самом деле, наличие хорошо известных регулярных методов построения псевдослучайных чисел [39] делает такое различие только кажущимся. Эти регулярные методы построения «случайных кодов» составляют регулярный вариант приведенного случайного алгоритма. Практически такие методы уже использовались для случая бинарного симметричного канала [40], где на машинах проводился подсчет эмпирических характеристик кодов, связанных с оценкой P' (для случая низкого уровня шумов). Теоретические оценки P' в этом случае были получены ранее [4] (см. также гл. 8).

В работе [41] показано, что при использовании регулярных методов получения псевдослучайных кодов можно избежать большого объема памяти на входе канала (чего нельзя сделать при регулярной структуре

кода). Учитывая возможности специального вида декодирования [42], понижающего объем памяти на выходе до порядка n^2, можно рассчитывать на получение практических возможностей для реализации оптимального кодирования. Подробнее эти вопросы изложены в гл. 9.

Выше для описания зависимости между параметрами кода использовалась нетабулированная функция $k_{\overline{pd}}(\varepsilon)$. Для получения простых явных связей между параметрами кода можно воспользоваться оценками функции $k_{\overline{pd}}(\varepsilon)$, приведенными в п. 1.6.2. Эти оценки оказываются тем более точными, чем выше уровень шумов в канале. Для случая высокого уровня шумов в канале можно легко сопоставить оптимальные соотношения между параметрами (n, M, P)-кода при статистическом (гл. 5) и комбинаторном (гл. 6) построении декодирования. В этом случае, как и в гл. 5, соотношения гл. 6 заметно упрощаются и будут иметь вид (см. теорему 7.3):

$$M = e^{n[C-\eta]}, \quad P \approx 1 - e^{-\varepsilon^2 n}, \tag{7.15}$$

где

$$\eta \geqslant \varepsilon^2 + \overline{\varepsilon}' + 2\overline{\varepsilon} > 0, \tag{7.16}$$

причем

$$\overline{\varepsilon}' = \sum_\alpha p_\alpha \varepsilon'_\alpha, \quad \overline{\varepsilon} = \sum_\alpha p_\alpha \varepsilon_\alpha,$$

$-\varepsilon'_\alpha < 0 < \varepsilon_\alpha$ — корни уравнения $\varepsilon^2 = p_\alpha k_{\overline{p}_\alpha, \overline{d}_\alpha}(u)$ и $\overline{p} = (\overline{p}_\alpha)$ обращает в максимум выражение

$$\max_{\overline{\mu}} [h(\overline{\mu}\,\mathbf{p}) - \sum_\alpha \mu_\alpha h(\overline{p}_\alpha)] = h(\overline{p}\,\mathbf{p}) - \sum_\alpha p_\alpha h(\overline{p}_\alpha) = C.$$

Исходя из соотношений п. 1.6.2, для случая высокого уровня шумов, и соотношений Бартлетта, имеем при $u \to 0$

$$k_{\overline{p}_\alpha \overline{d}_\alpha}(u) \approx \frac{u^2}{2\gamma''_\alpha(0)},$$

где

$$\gamma''_\alpha(0) \approx 2c_\alpha = 2\sum_\beta p_\alpha^\beta \ln \frac{p_\alpha^\beta}{q_\beta} \left(q_\beta = \sum_\alpha p_\alpha p_\alpha^\beta\right).$$

Поэтому $\varepsilon'_\alpha \approx \varepsilon_\alpha \approx 2\varepsilon \sqrt{c_\alpha / p_\alpha}$,

откуда

$$\overline{\varepsilon}' \approx \overline{\varepsilon} \approx 2\varepsilon \sum_{\alpha=1}^{a} \sqrt{p_\alpha c_\alpha}$$

и условие (7.16) примет вид

$$\eta \geqslant 6\varepsilon \sum_{\alpha=1}^{a} \sqrt{p_\alpha c_\alpha} + \varepsilon^2.$$

Используя неравенство Буняковского, можно показать, что последнее неравенство при $\varepsilon \to 0$, будет выполняться, если положить

$$\eta = 6\varepsilon \sqrt{a} \sqrt{\sum_{\alpha=1}^{a} p_\alpha c_\alpha} = 6\sqrt{a}\sqrt{C}\,\varepsilon. \tag{7.17}$$

Пусть $C-\eta=H$, тогда с учетом (7.17) соотношения (7.15) примут следующий окончательный вид:

$$M=e^{nH},\quad P\approx 1-e^{-\frac{(C-H)^2}{36aC}n}. \tag{7.18}$$

Сравнивая соотношения (7.18) и (5.32) (гл. 5), можно видеть, что они отличаются лишь коэффициентом перед выражением $\frac{(C-H)^2}{C}n$ показателя экспоненты в вероятности P.

В соотношениях (5.32) он равен $^1/_4\,a$, а в соотношениях (7.18) равен $^1/_{36}a$, т. е. в последнем случае имеет место более медленное стремление P к единице с ростом n. Это обстоятельство связано с рядом грубых оценок при выводе условия (7.16) в гл. 6, а также последних оценок, использующих неравенство Буняковского.

Следует отметить, что говоря об оптимальных соотношениях типа (7.18), которым должны удовлетворять соответствующие параметры оптимальных по Шеннону кодов, мы нигде не оговаривали порядка стремления $P\to 1$ с ростом n при $H<C$. В определение оптимальности входит лишь факт такого рода стремления, и всюду это стремление носило экспоненциальный характер. Однако только случай статистического декодирования (гл. 5) ограничивает из статистических соображений величину коэффициента, стоящего перед n в экспоненте P. Вместе с тем величина наивысшего порядка роста M_c с ростом n ограничивается в гл. 6 из чисто комбинаторных соображений. Таким образом, оба этих экспоненциальных выражения для P и M_c могут рассматриваться как «предельно оптимальные».

Перейдем к обзору результатов теории информации и к сравнению полученных результатов с результатами теории информации Шеннона для рассматриваемого канала [10].

§ 7.5. Сравнение с результатами теории информации

7.5.1. *Неконструктивность результатов теории информации для каналов с шумами.* Теория информации в узком смысле (см. введение) охватывает круг теоретических проблем, связанных с двумя упоминавшимися основополагающими теоремами Шеннона об оптимальном кодировании для каналов без шумов и каналов с шумами.

В первоначальной основополагающей работе Шеннона [2] заранее формально вводится ряд величин: энтропия, скорость передачи, пропускная способность и др. Оправданием введения этих величин является то, что с их помощью удается доказать существование оптимального кодирования для канала с шумами[1].

Так, для рассматриваемого дискретного постоянного канала с независимыми шумами Шеннон в [2], отправляясь от дискретного распределения $\overline{p}=(p_1,\ldots,p_\alpha,\ldots,p_a)$ и матрицы переходов символов $\mathbf{p}=\|p_\alpha^\beta\|=(\overline{p_\alpha})$ $(\alpha=\overline{1,a},\ \beta=\overline{1,b})$, вводит заранее формальные величины энтропии

$$H(X)=-\sum_{\alpha=1}^{a}p_\alpha\ln p_\alpha,\quad H(Y)=-\sum_{\beta=1}^{b}q_\beta\ln q_\beta\left(q_\beta=\sum_{\alpha=1}^{a}p_\alpha p_\alpha^\beta\ (\beta=\overline{1,b})\right) \tag{7.19}$$

[1] Для канала без шумов оптимальный по Шеннону код был независисмо от Шеннона построен Фано без использования понятий теории информации.

и условной энтропии

$$H(Y/X) = -\sum_{\alpha=1}^{a} p_\alpha \sum_{\beta=1}^{b} p_\alpha^\beta \ln p_\alpha^\beta = \sum_{\alpha=1}^{a} p_\alpha h(\overline{p_\alpha}). \tag{7.20}$$

Далее он формально вводит так называемую величину скорости передачи

$$R = H(Y) - H(Y/X) \tag{7.21}$$

и пропускную способность канала

$$C = \max_{\overline{p}} R \tag{7.22}$$

и с их помощью дает наброски доказательства существования оптимального кода для канала с шумами, определяемыми в рассматриваемом случае матрицей переходов символов $\mathbf{p} = \| p_\alpha^\beta \|$. Полное строгое доказательство содержится в его работе [10].

Первоначальная работа Шеннона [2] и почти все последовавшие за ней работы по теории информации других авторов пронизаны идеями «разумного» измерения информации по аналогии с измерениями физических величин. Однако уже в работе [3], появившейся вскоре после опубликования работы [2], Шеннон преодолевает формализм первой работы для частного случая гауссовского непрерывного канала в одном пункте. Именно, ему удается вычислить пропускную способность рассматриваемого канала

$$C = W \ln(1 + N_с/WN_ш) \tag{7.23}$$

(обозначения см. п. 4.5.1), исходя из прямых геометрических соображений, не прибегая к формальному определению (7.22).

В работе [10] Шеннон стремится избежать формального рассмотрения декодирования указанием, что последнее может строиться на основе статистической идеи максимума функции правдоподобия. Из других работ по теории информации, избегающих по возможности формального рассмотрения, можно упомянуть работу Бернарда [36] и особенно замечательную монографию Фано [35]. Однако, несмотря на некоторые тенденции неформального развития теории информации, все ее современные положительные результаты связаны в основном с доказательством теорем существования оптимальных кодов для каналов с шумами с помощью формального введения величин типа (7.19)—(7.22).

По-видимому, идеи измерения величины информации имели для Шеннона основное наводящее значение при формулировке теорем существования оптимальных кодов для каналов без шумов и с шумами. Впоследствии эти теоремы сравнительно просто доказывались им и другими авторами методами, не выходящими за пределы теории вероятностей и математического анализа[1]. Однако открытие Шенноном существования оптимальных кодов для каналов с шумами и возможность доказательства их существования с помощью известных математических методов были куплены дорогой ценой: неконструктивностью его теории в ее важнейшей части — в оптимальном кодировании для каналов с шумами.

Последующее развитие теории информации не изменило положения, поскольку использовавшийся в основном непрерывный аппарат математического анализа оказался пригодным лишь для доказательства теорем существования оптимальных кодов для каналов с шумами. Вместе с

[1] Следует заметить, что строгое доказательство теорем существования оптимальных кодов для общих случаев каналов с шумами довольно сложно.

тем нельзя переоценить значение теории информации Шеннона для стимулирования развития конструктивных методов оптимального кодирования, изложенных в этой книге, не говоря уже об общенаучном интересе этой теории. В самом деле, теоремы существования нацеливали конструктивную теорию на построение кодов с определенными оптимальными свойствами. Кроме того, ряд идей теории информации, используемых для доказательства теорем существования, был применен в конструктивной теории для прямого построения оптимальных кодов. Это относится в первую очередь к идее случайного кодирования.

7.5.2. *Сравнение предельного случая идеального приемника В. А. Котельникова с результатом Шеннона.* Начнем сравнение результатов теории информации и развитой конструктивной теории с первоначального примера асимптотического случая идеального приемника В. А. Котельникова (см. § 4.5), связанного с непрерывным Гауссовским каналом, Теорема Шеннона об оптимальном кодировании для рассматриваемого канала [2,3] утверждает, что существует такой способ кодирования и декодирования, при котором вероятность правильного декодирования P с ростом числа измерений $n=T/2\ W$ «дискретов» сигналов

$$P\begin{cases}\to 1 \text{ при } H<C,\\ \to 0 \text{ при } H>C,\end{cases} \tag{7.24}$$

где H — энтропия источника, а C — пропускная способость канала, определяемая в рассматриваемом случае соотношением (7.23).

Рассмотрим случай высокого уровня шумов, что эквивалентно в рассматриваемом случае требованию стремления отношения сигнал/шум $c=N_{\text{с}}/N_{\text{ш}}\to 0$. Тогда, согласно соотношению (7.23), $C\to c$ и сравнение соотношений (4.4) и (7.24) указывает на их эквивалентность. Заметим, что в ряде работ, в том числе и самого Шеннона [10], был установлен экспоненциальный характер стремления P к единице и к нулю с ростом числа измерений n «дискретов» сигналов.

7. 5. 3. *Сравнение предельных соотношений для параметроз (n, M, P)-кода.* В работе [10] Шеннон специально рассмотрел случай дискретного постоянного канала с независимыми шумами. Используя результаты [20], в [10] он получил оценки для асимптотического поведения вероятности $1-P$ ошибки (дополняет до единицы вероятность правильного декодирования). Эти оценки того же экспоненциального типа, что и оценки (5.24) гл. 5. Для случая высокого уровня шумов оценка вероятности правильного декодирования, приведенная в [10], совпадает с аналогичной оценкой (5.32).

7.5.4. *Характерные особенности конструктивного варианта теории информации.* Конструктивное решение проблемы, связанной с построением оптимального кода для канала без шумов, сравнительно элементарно (код Фано — Шеннона). Решение проблемы построения оптимального кода для канала с шумами существенно более сложно и является в некотором смысле «пробным камнем» различных современных математических методов.

Основной результат, полученный выше с помощью комбинаторных методов, состоит в построении оптимальных в смысле Шеннона кодов для общего случая дискретного постоянного канала с независимыми шумами. В связи с этим развитая теория может рассматриваться как первые шаги конструктивной теории информации, в которой в отличие от существующей теории приводится конструкция оптимальных кодов для каналов с шумами.

В конструктивную же теорию должны входить известные конструктивные результаты оптимального кодирования для каналов без шумов,

также имеющие в асимптотическом случае глубокий комбинаторный смысл (см. § 2.5). Кроме указанного принципиального отличия конструктивного варианта теории информации от теории информации Шеннона, первая отличается от второй и методологически. В самом деле, в развитой в настоящей книге теории заранее формально не вводятся величины энтропии, скорости передачи, пропускной способности и др. Последние возникают сами собой как некоторые «характерные» константы в ходе построения оптимальных кодов.

Отметим, что в конструктивной теории отчетливо выделяется та ее часть, которая требует развития нового математического аппарата. Именно, для построения входных слов оптимального кода (см. гл. 7) требуется развитие новых комбинаторных методов (основная лемма о пересечениях). Оптимальная процедура декодирования может быть полностью построена на основе известных статистических принципов выбора между гипотезами (гл. 5).

На основе тех же статистических методов могут быть доказаны теоремы существования оптимальных кодов для рассматриваемого нами канала без предварительного введения упоминавшихся формальных понятий теории информации (см. гл. 5). Таким образом, первоначальное формальное введение этих понятий в рассмотренной ситуации вряд ли оправдано и результаты гл. 5 полностью выясняют их статистический смысл.

Глава 8

ЧАСТНЫЙ СЛУЧАЙ БИНАРНОГО СИММЕТРИЧНОГО КАНАЛА

§ 8.1. Частные случаи постоянного дискретного канала с независимыми шумами

8.1.1 *Вводные замечания.* Эта глава посвящена оптимальному кодированию для простейшего случая постоянного дискретного канала с независимыми шумами. Речь идет о бинарном симметричном канале [2], задаваемом одним параметром — вероятностью p искажения бинарных символов 0 и 1. Помимо иллюстрации общих алгоритмов и соотношений, в этом частном случае можно получить ряд более сильных результатов, чем в общем случае. В идейном плане изложение здесь соответствует изложению для общего случая и сопровождается подробными доказательствами. Поэтому данную главу с учетом определений гл. 6 можно читать независимо от других глав.

Бинарный симметричный канал является простейшим и удобным для расчетов, с его помощью можно описать ряд практически важных реальных каналов с шумами. Однако, помимо него, имеется ряд других важных частных случаев дискретного канала, но для них не происходит сколько-нибудь заметных упрощений общих алгоритмов и соотношений. Поэтому в этом параграфе мы лишь кратко остановимся на важнейших частных случаях дискретного канала, а последующее изложение будет посвящено исключительно бинарному симметричному каналу.

8.1.2. *Дискретный канал с дважды стохастической матрицей переходов.* Рассмотрим постоянный дискретный канал с независимыми шумами, определяемыми матрицей переходов символов $\mathbf{p} = \| p_\alpha^\beta \|$, которая, помимо обязательных условий $\sum_{\beta=1}^{b} p_\alpha^\beta = 1 \; (\alpha = \overline{1,a})$, удовлетворяет еще дополнительным условиям $\sum_{\alpha=1}^{a} p_\alpha^\beta = 1 \; (\beta = \overline{1,b})$. Такую матрицу переходов символов будем называть дважды стохастической. Другими словами, такая матрица, кроме обязательного свойства равенства единице сумм элементов по строкам, обладает аналогичным свойством и для сумм элементов по столбцам.

Одной из вероятностных схем, моделирующих такой канал при $a = b$, может служить следующая схема.

Рассмотрим генеральную совокупность, содержащую подстановки $S_l = \begin{pmatrix} 1 \ldots & \alpha \ldots & a \\ \beta_l(1) \ldots & \beta_l(\alpha) \ldots & \beta_l(a) \end{pmatrix} (l = \overline{1,L})$ с долями, соответствующими вероятностям $p_l \; (l = \overline{1,L}, \sum_{l=1}^{L} p_l = 1)$.

Рассматриваемый канал определяется случайной подстановкой

$$S = \begin{pmatrix} S_1, \ldots, S_l, \ldots, S_L \\ p_1, \ldots\ p_l, \ldots, p_L \end{pmatrix}.$$

Пусть в каждый момент передачи t происходит выбор из генеральной совокупности одной из подстановок $S_l = (\beta_l(\alpha))$ и в соответствии с ней входной символ α переходит в выходной символ $\beta_l(\alpha)$. Тогда в рассматриваемом случае

$$\mathbf{p} = \| p_\alpha^\beta \| = \sum_{l=1}^{L} p_l S_l = \mathrm{E}S, \tag{8.1}$$

где подстановка $S_l = \begin{pmatrix} \alpha \\ \beta_l(\alpha) \end{pmatrix}$ рассматривается как $(a \times a)$-матрица со всеми элементами, равными нулю, кроме элементов, стоящих на пересечении α-й строки и $\beta_l(\alpha)$-го столбца, равных единице. Ясно, что матрицы S_l являются вырожденным случаем дважды стохастических матриц и этим же свойством обладает их взвешенная сумма (8.1) (имеет место и обратное утверждение Биркгофа, фон Неймана [14]). Заметим, что стохастическая матрица $\mathbf{p} = \| p_\alpha^\beta \| = \| p_{\beta-\alpha} \| \left(\sum_{\alpha=1}^{a} p_\alpha = 1, \beta - \alpha = \gamma \bmod a, \alpha, \beta = \overline{1,a} \right)$, приведенная в доп. I, также является дважды стохастической.

В этом случае, если $\bar{p} = \left(\overbrace{\frac{1}{a}, \ldots, \frac{1}{a}}^{a} \right)$, то, как легко видеть,

$$\bar{q} = \bar{p}\mathbf{p} = \left(\overbrace{\frac{1}{b}, \ldots, \frac{1}{b}}^{b} \right).$$

8.1.3. *Эквиэнтропийный канал.* Пусть матрица $\mathbf{p} = \| p_\alpha^\beta \| = (\bar{p}_\alpha)$ переходов символов обладает свойством $h(\bar{p}_\alpha) = h = \mathrm{const}$. Будем называть канал с такой матрицей переходов символов эквиэнтропийным. Другими словами, матрица переходов символов для такого канала имеет постоянную h-функцию от строк.

В этом случае фундаментальная константа канала

$$C = \max_{\bar{\mu}} [h(\bar{\mu}\mathbf{p}) - \sum_{\alpha=1}^{a} \mu_\alpha h(\bar{p}_\alpha)] = \max_{\bar{\mu}} h(\bar{\mu}\mathbf{p}) - h. \tag{8.2}$$

Пусть эквиэнтропийный канал обладает, кроме того, еще и дважды стохастической матрицей; тогда, исходя из замечания в конце предыдущего пункта, имеем $\left(\overbrace{\frac{1}{a}, \ldots, \frac{1}{a}}^{a} \right) \mathbf{p} = \left(\overbrace{\frac{1}{b}, \ldots, \frac{1}{b}}^{b} \right)$ и, как легко видеть, максимум в соотношении (8.2) достигается при $\bar{\mu} = \left(\underbrace{\frac{1}{a}, \ldots, \frac{1}{a}}_{a} \right)$ и равен $C = \ln b - h$.

8.1.4. *Стирающий канал.* Рассмотрим частный случай дискретного канала с (2×3)-матрицей вероятностей перехода символов вида:

$$\mathbf{p} = \| p_\alpha^\beta \| = \begin{pmatrix} q & 0 & p \\ 0 & q & p \end{pmatrix} (\alpha = 1,2;\ \beta = 1,2,3;\ p + q = 1). \tag{8.3}$$

Определенный канал называется стирающим каналом [38]. Это название оправдано следующей вероятностной схемой, приводящей к рассматриваемому каналу. Пусть на вход канала поступают два символа 1 и 2. На выходе канала в результате действия шумов в каждый момент дискретного времени передаваемый символ с вероятностью q переходит сам в себя и с вероятностью $p = 1 - p$ делается «неразборчивым» (стирается), вследствие чего не выносится никакого решения о его происхождении. Этот неопределенный символ формально можно обозначить 3.

Стирающий канал является, как легко видеть, эквиэнтропийным, однако матрица переходов (8.3) не является дважды стохастической, поэтому для определения фундаментальной константы C воспользуемся соотношением (8.2), которое в нашем случае имеет вид

$$C = \max_{\bar{\mu}} h(\bar{\mu}\mathbf{p}) - h,$$

где p определяется из (8.3) и $h = h(\bar{p}) = h(p) = -p\ln p - q\ln q$.

Далее, вводя $h(\bar{\mu}) = h(\mu)$ и $\bar{\mu}\mathbf{p} = (\mu, 1-\mu)\begin{pmatrix} q & 0 & p \\ 0 & q & p \end{pmatrix} = (\mu q,\ (1-\mu)q,\ p)$, получим $h(\bar{\mu}\mathbf{p}) = qh(\mu) + h(p) - h(p) = qh(\mu)$, т. е. максимум $\max_{\bar{\mu}} h(\bar{\mu}\mathbf{p}) = q\max_{\bar{\mu}} h(\mu) = qh(1/2) = q\ln 2$ достигается при $\bar{\mu} = \bar{p}^{(0)} = (1/2,\ 1/2)$.

Итак, имеем окончательно для стирающего канала [38]

$$C = q\ln 2,\ \bar{p}^{(0)} = (1/2,\ 1/2).$$

Далее излагаются результаты [6], касающиеся исключительно бинарного симметричного канала.

В отличие от общего случая дискретного канала здесь можно обойтись без использования аппарата производящих функций и без k-функции, ограничиваясь более элементарными средствами. Вместе с тем для бинарного симметричного канала можно провести более глубокие, чем в общем случае, исследования пересечений «комбинаторных» сфер. При этом в асимптотическом случае появляются новые, важные, связанные между собой функции $f_{\theta_1}(\theta)$ и $g_{\theta_1}(\theta)$ (см. § 8.4). С их помощью легко формулируются достаточные условия P-различимости, а также устанавливается их прозрачный геометрический смысл (см. § 8.5).

§ 8.2. Постановка задачи и определения

Рассмотрим передачу бинарных символов 0 и 1 по постоянному, симметричному каналу с независимыми шумами. Это означает, что в дискретные моменты времени с номерами $1,2,\ldots,t,\ldots,n$ бинарные символы $\alpha_t = 0,1$ на входе канала переходят в соответствующие символы $\beta_t = 0,1$ на выходе канала с условными вероятностями

$$\mathbf{p} = \| p_\alpha^\beta \| = \begin{pmatrix} p_0^0 p_0^1 \\ p_1^0 p_1^1 \end{pmatrix}, \tag{8.4}$$

где

$$p_\alpha^\beta = \mathcal{P}(\beta_t = \beta / \alpha_t = \alpha)\ (\alpha, \beta = 0,1),$$

не зависящими как от номера момента времени t (постоянство канала), так и от значений $\alpha_{t'}$ и $\beta_{t'}$ при $t' < t$ (независимые шумы).

Кроме того, рассмотрим случай $p_0^1 = p_1^0 = p$ и $p_0^0 = p_1^1 = q = 1 - p$ (симметричность канала).

Не нарушая общности, можно считать, что

$$0 \leqslant p \leqslant 0{,}5 \leqslant q \leqslant 1. \tag{8.5}$$

Рассмотрим входные слова длины n на входе канала $x = (\alpha_1, \alpha_2, \ldots, \alpha_t, \ldots, \alpha_n)$ и соответствующие им выходные слова длины n, $y = (\beta_1, \beta_2, \ldots, \beta_t, \ldots, \beta_n)$ на выходе канала.

Совокупность всевозможных входных слов x длины n будем обозначать R и называть пространством входных слов длины n. Аналогично совокупность всевозможных выходных слов y длины n обозначим R^* и будем называть пространством выходных слов длины n.

Задание вероятностей (8.4) в нашем случае дает возможность вычислить условные вероятности $p_x(y) = \mathcal{P}(y/x)$ переходов входных слов x в выходные слова y.

Можно показать, что в нашем случае

$$p_x(y) = p^{d(x,y)} q^{n-d(x,y)}, \tag{8.6}$$

где $d(x, y) = d$ означает число разнящихся друг от друга символов α_t и β_t, входящих в x и y соответственно. Будем называть $d = d(x, y)$ расстоянием Хэмминга между x и y (см. гл. 6)

Рассмотрим произвольное множество $\mathscr{E}$ выходных слов $y \in \mathscr{E} \subset R^*$. Используя вероятность (8.6), легко подсчитать условную вероятность $p_x(\mathscr{E}) = \mathcal{P}(y \in \mathscr{E}/x)$ появления какого-либо из $y \in \mathscr{E}$ на выходе, если x появляется на входе канала. Эту вероятность мы будем коротко обозначать $p_x(\mathscr{E})$ и называть вероятностью множества $\mathscr{E}$ на выходе при фиксации x на входе. Она равна

$$p_x(\mathscr{E}) = \sum_{y \in \mathscr{E}} p_x(y). \tag{8.7}$$

Ясно, что при $\mathscr{E} = R^*$ для любого x $p_x(R^*) = 1$.

В дальнейшем изложении мы будем пользоваться законом аддитивности вероятности (8.7) в следующей общей форме:

$$p_x(\mathscr{E}_1 \cup \mathscr{E}_2) = p_x(\mathscr{E}_1) + p_x(\mathscr{E}_2) - p_x(\mathscr{E}_1 \cap \mathscr{E}_2). \tag{8.8}$$

Кроме того, если $\mathscr{E}_1 \subseteq \mathscr{E}_2$, т. е. $\mathscr{E}_1$ входит или совпадает с $\mathscr{E}_2$, то

$$p_x(\mathscr{E}_1) \leqslant p_x(\mathscr{E}_2).$$

Легко видеть, что теми же свойствами обладает число элементов множества $\mathscr{E}$, обозначаемое $N(\mathscr{E})$,

$$N(\mathscr{E}_1 \cup \mathscr{E}_2) = N(\mathscr{E}_1) + N(\mathscr{E}_2) - N(\mathscr{E}_1 \cap \mathscr{E}_2), \tag{8.9}$$
$$N(\mathscr{E}_1) \leqslant N(\mathscr{E}_2).$$

Если $\mathscr{E}_1$ и $\mathscr{E}_2$ не пересекаются (что мы будем обозначать в виде $\mathscr{E}_1 \cap \mathscr{E}_2 = \varnothing$, где $\varnothing$ — символ пустого множества), то соотношения (8.8) и (8.9) упростятся, так как будут содержать лишь аддитивные члены.

Если $\mathscr{E}_1 \subseteq \mathscr{E}_2$, то определим $\mathscr{E}_3$ как такое множество, которое при объединении с $\mathscr{E}_1$ дает $\mathscr{E}_2$. Символически будем записывать $\mathscr{E}_3$ в виде $\mathscr{E}_3 = \mathscr{E}_2 - \mathscr{E}_1$.

Из (8.8) и (8.9) следует

$$p_x(\mathscr{E}_3) = p_x(\mathscr{E}_2) - p_x(\mathscr{E}_1), \tag{8.10}$$
$$N(\mathscr{E}_3) = N(\mathscr{E}_2) - N(\mathscr{E}_1).$$

Зафиксируем произвольное входное слово $x \in R$ и соотнесем ему множество $\mathscr{E}$ выходных слов.

Согласно определению 1 гл. 6, множество $\mathscr{E} \subset R^*$ P-представляет $x \in R$, если $p_x(\mathscr{E}) = P \underset{n\to\infty}{\to} 1$. Если $\lim\limits_{n\to\infty} p_x(\mathscr{E}) = 0$, то мы скажем, что $\mathscr{E}$ не представляет x. Тривиальным примером множества, представляющего любое $x \in R$, является R^*.

В § 8.3 будут рассмотрены подмножества $\mathscr{E} \subset R^*$, представляющие x.

Используя соотношение (8.8), легко показать, что если у $\mathscr{E}'$ дополнение $\overline{\mathscr{E}}'$ P-представляет x и $\mathscr{E}' \subset \mathscr{E}$, где $\mathscr{E}$ P-представляет x, то $\mathscr{E}'' = \mathscr{E} - \mathscr{E}'$ по-прежнему P-представляет x.

Таким образом, если выбросить из $\mathscr{E}$ P-представляющего x, его часть $\mathscr{E}'$ с дополнением $\overline{\mathscr{E}}'$ P-представляющим x, то оставшаяся его часть $\mathscr{E}''$ по-прежнему будет P-представлять x.

Ясно, что если $\mathscr{E}$ и $\mathscr{E}' \subset \mathscr{E}$ P-представляют x, то $\mathscr{E}'' = \mathscr{E} - \mathscr{E}'$ не представляет x.

Можно показать, что если $\mathscr{E}_1$ и $\mathscr{E}_2$ P-представляют x, то и их пересечение $\mathscr{E}_3 = \mathscr{E}_1 \cap \mathscr{E}_2$ P-представляет x.

Нас будет интересовать тот случай, когда множества $\mathscr{E}_i$, P-представляющие различные x_i ($i = 1, 1, 2, \ldots, M$), таковы, что

$$\bigcup_{i=1}^{M} \mathscr{E}_i = R^*,$$

и их части вида

$$\mathscr{E}_i'' = \mathscr{E}_i \cap \bigcup_{j\neq i} \mathscr{E}_j$$

не представляют x_i, а дополнения $\overline{\mathscr{E}}_i''$ P-представляют x.

Этот случай важен тем, что выбрасывая $\mathscr{E}_i''$ из $\mathscr{E}_i$, мы получим множества $\mathscr{E}_i' = \mathscr{E}_i - \mathscr{E}_i''$, по-прежнему P-представляющие x_i и такие, что $\mathscr{E}_i' \cap \mathscr{E}_j' = \varnothing$ $(i \neq j)$.

Поэтому при появлении $y \in \mathscr{E}_i'$ на выходе канала при достаточно большом n можно с вероятностью, сколь угодно близкой к единице, утверждать, что на его входе имеет место x_i. Заметим, что для того, чтобы избежать неоднозначности в выборе соответствующего x_i при появлении $y \in \mathscr{E}_i''$ ($p_{x_i}(\mathscr{E}_i'') \underset{n\to\infty}{\to} 0$), можно произвольным образом распределить $y \in \mathscr{E}_i''$ по множествам $\mathscr{E}_j$.

Легко видеть, что получившиеся таким образом множества $\widetilde{\mathscr{E}}_i$ не пересекаются, по-прежнему P-представляют x_i и заполняют все пространство выходных слов R^*.

Изложенное оправдывает следующий вариант приведенного в гл. 6 определения 2. Совокупность M входных слов длины n

$$x_1 x_2 \ldots x_i \ldots x_j \ldots x_M \tag{8.11}$$

называется P-различимой, а входящие в нее входные слова x_i — P-различимыми, если существуют такие P-представляющие их множества выходных слов длины n соответственно

$$\mathscr{E}_1 \mathscr{E}_2 \ldots \mathscr{E}_i \ldots \mathscr{E}_j \ldots \mathscr{E}_M,$$

что $\mathscr{E}_i'' = \mathscr{E}_i \cap \bigcup\limits_{j\neq i} \mathscr{E}_j$ $(i = \overline{1, M})$ не представляют x_i, а дополнения $\overline{\mathscr{E}}_i''$ P-представляют x_i.

Так как из определения множеств $\mathscr{E}_i'$, $\mathscr{E}_i''$, $\mathscr{E}_i$ и $\widetilde{\mathscr{E}}_i$ следует, что

$$\mathscr{E}_i' = \mathscr{E}_i - \mathscr{E}_i'' \subseteq \widetilde{\mathscr{E}}_i \subseteq \mathscr{E}_i,$$

то имеем

$$p_{x_i}(\mathscr{E}_i') = p_{x_i}(\mathscr{E}_i) - p_{x_i}(\mathscr{E}_i'') \leqslant p_{x_i}(\widetilde{\mathscr{E}}_i) \leqslant p_{x_i}(\mathscr{E}_i).$$

Если использовать P-различимые входные слова длины n (8.11) и соответствующие им множества $\widetilde{\mathscr{E}}_i$ выходных слов длины n для того, чтобы с их помощью закодировать L высоковероятностных [2] сообщений источника, то при $L < M$ можно осуществить при достаточно большом n почти безошибочную передачу с вероятностью правильного декодирования

$$P \geqslant \min_{1 \leqslant i \leqslant M} p_{x_i}(\widetilde{\mathscr{E}}_i) \to 1$$

Известно, что оптимальный в смысле Шеннона [2] код содержит предельно большое число $M = 2^{nH}$ P-различимых входных слов длины n, где H — энтропия источника. Наша задача сводится к построению такого кода.

Как уже отмечалось в гл. 6, нетрудно дать развернутое толкование определений 1 и 2 подчеркиванием того, что $x_i = x_i^{(n)}$ и $\mathscr{E}_i = \mathscr{E}_i^{(n)}$ является n-ми членами бесконечных при $n \to \infty$ последовательностей. Однако во избежание громоздких формулировок мы этого не будем делать, тем более что в последующем изложении конкретная структура x_i и $\mathscr{E}_i$ делает такое развернутое толкование не обязательным.

§ 8.3. Необходимые условия для того, чтобы $\mathscr{E}$ P-представляло x

Приступим к более детальному изучению множеств $\mathscr{E}$, P-представляющих x.

Зафиксируем $x \in R$ и рассмотрим все те $y \in R^*$, для которых $d(x, y) = d = \text{const}$. Совокупность всех таких y обозначим через $\mathscr{E}_x^d \subset R^*$.

Для изучения множеств $\mathscr{E}_x^d$ рассмотрим некоторые свойства расстояния $d(x, y)$ между x и y. Прежде всего придадим данному в предыдущем параграфе определению числа $d = d(x, y)$ математическую форму.

В самом деле, если рассматривать x и y как упорядоченные множества n элементов (нулей и единиц), то

$$d = d(x, y) = N(x - (x \cap y)) = N(y - (x \cap y)) = n - N(x \cap y), \quad (8.12)$$

где $x \cap y$ означает общую часть x и y, а $x - (x \cap y)$ и $y - (x \cap y)$ — разнящиеся части x и y.

Пользуясь определением (8.12), легко вывести следующие свойства $d = d(x, y)$:

1. $d(x, y)$ — неотрицательное целое число, заключенное в переделах $0 \leqslant d(x, y) \leqslant n$, причем

$$d(x, y) = \begin{cases} 0, \text{ при } x + y = \overline{0} = \underbrace{(0, 0, \ldots, 0)}_{n} \bmod 2 \\ n, \text{ при } x + y = \overline{1} = \underbrace{(1, 1, \ldots, 1)}_{n} \bmod 2 \end{cases}$$

(здесь $x+y=z$; mod 2 означает покомпонентное их сложение по правилу: $0+0=1+1=0$ и $0+1=1+0=1$);

2. $d(x, y)=d(y, x)$;

3. $d(x+z, y+z)=d(x, y)$; (8.13)

4. $d(x, y)+d(x+\overline{1}, y)=n$; (8.14)

5. $d(x, y)=d(x', y')+d(x-x', y-y')$, (8.15)

где x' и y' подмножества соответственного x и y, определенные на одних и тех же моментах времени.

Приступим к изучению свойств множества $\mathscr{E}_x^d$.

Прежде всего заметим, что свойство (8.13) позволяет установить взаимно-однозначное соответствие между элементами множеств $\mathscr{E}_x^d$ и $\mathscr{E}_{\overline{0}}^d$ в форме

$$\mathscr{E}_x^d=\mathscr{E}_{\overline{0}}^d+x, \tag{8.16}$$

понимая под правой частью этого соотношения совокупность y-ов, полученных в результате сложения x по mod 2 со всеми элементами $\mathscr{E}_{\overline{0}}^d$.

Из (8.16) следует равночисленность множеств $\mathscr{E}_x^d$ и $\mathscr{E}_{\overline{0}}^d$. Последнее множество, как легко видеть, имеет число элементов $N(\mathscr{E}_{\overline{0}}^d)=C_n^d$, отсюда

$$N(\mathscr{E}_x^d)=N(\mathscr{E}_{\overline{0}}^d)=C_n^d=\frac{n!}{d!\,(n-d)!}. \tag{8.17}$$

Используя (8.6), (8.17) и определение $\mathscr{E}_x^d$, получим

$$p_x(\mathscr{E}_x^d)=C_n^d p^d q^{n-d}. \tag{8.18}$$

Введем в рассмотрение множество

$$\mathscr{E}_x^{d_1, d_2}=\bigcup_{d=d_1}^{d_2}\mathscr{E}_x^d, \tag{8.19}$$

где $d_1\leqslant d_2$.

Заметим, что при $d_1=0$ и $d_2=n$

$$\bigcup_{d=0}^{n}\mathscr{E}_x^d=\mathscr{E}_x^{0,n}=R^*.$$

Кроме того,

$$\mathscr{E}_x^{d_1}\cap\mathscr{E}_x^{d_2}=\begin{cases}\mathscr{E}_x^{d_1} & \text{при } d_1=d_2,\\ \varnothing, & \text{при } d_1\neq d_2.\end{cases} \tag{8.20}$$

Из (8.18), (8.19) и (8.20) следует, что

$$p_x(\mathscr{E}_x^{d_1, d_2})=\sum_{d=d_1}^{d_2}C_n^d p^d q^{n-d}=1-\sum_{d=0}^{d_1-1}C_n^d p^d q \quad -\sum_{d=d_2+1}^{n}C_n^d p\, q^{n-d}. \tag{8.21}$$

В дальнейшем изложении часто будут встречаться множества $\mathscr{E}_x^{[n(p-\varepsilon)],\,[n(p+\varepsilon)]}$, где $0<\varepsilon<p$ и $[a]$ означает целую часть числа a. Поэтому для них мы введем специальное обозначение

$$\mathscr{E}_x^{[n(p-\varepsilon)],\,[n(p+\varepsilon)]}=\mathscr{E}_x^{[n(p\pm\varepsilon)]}.$$

Нашей дальнейшей целью является изучение асимптотического поведения вероятности (8.21) при $n \to \infty$. Заметим, что с помощью формулы Стирлинга можно получить соотношение *

$$C_n^{[n\theta]} \approx 2^{nh(\theta)} \quad (0 < \theta < 1), \tag{8.22}$$

где

$$h(\theta) = -\theta \log \theta - (1-\theta)\log(1-\theta),$$

причем, $0 \leqslant h(\theta) = h(1-\theta) \leqslant 1$ функция, определенная для $0 \leqslant \theta \leqslant 1$ и имеющая при $\theta = 0{,}5$ максимум $h(0{,}5) = 1$.

Производные $h(\theta)$ имеют вид:

$$h'(\theta) = \log\frac{1-\theta}{\theta}; \quad h''(\theta) = -\frac{1}{\ln 2}\,\frac{1}{\theta(1-\theta)};$$

$$h'''(\theta) = \frac{1}{\ln 2}\,\frac{1-2\theta}{\theta^2(1-\theta)^2}.$$

При этом на интервале $0 < \theta \leqslant 0{,}5$ с ростом θ, $h(\theta)$ и $h''(\theta)$ монотонно возрастают, а $h'(\theta)$ и $h'''(\theta)$ монотонно убывают. Из этого факта, используя разложение Тейлора, можно получить важные оценки.

Пусть $0 \leqslant \varepsilon \leqslant p\frac{q}{q-p}$, тогда при любом $0 \leqslant p \leqslant 0{,}5$ и $0 \leqslant p - \varepsilon \leqslant p + \varepsilon \leqslant 0{,}5$ имеют место следующие оценки:

$$\frac{2}{3}\varepsilon^2 h''(p) \leqslant \frac{\varepsilon^2}{2}h''(p) - \frac{\varepsilon^3}{6}h'''(p) \leqslant h(p-\varepsilon) - h(p) + \varepsilon h'(p) =$$

$$= \frac{\varepsilon^2}{2}h''(p) + O[(0{,}5-p)^4] \leqslant \frac{\varepsilon^2}{2}h''(p), \tag{8.23}$$

$$\frac{\varepsilon^2}{2}h''(p) \leqslant h(p+\varepsilon) - h(p) - \varepsilon h'(p) = \frac{\varepsilon^2}{2}h''(p) +$$

$$+ O[(0{,}5-p)^4] \leqslant \frac{\varepsilon^2}{2}h''(p) + \frac{\varepsilon^3}{6}h'''(p) \leqslant \frac{\varepsilon^2}{3}h''(p), \tag{8.24}$$

откуда

$$0 \leqslant h(p+\varepsilon) - h(p-\varepsilon) - 2\varepsilon h'(p) \leqslant \frac{\varepsilon^2}{3}h'''(p).$$

Приступим теперь к асимптотической оценке вероятности (8.18). Положим $d = [n\theta]\,(0 \leqslant \theta \leqslant 1)$, тогда, используя (8.22), будем иметь

$$C_n^d p^d q^{n-d} \approx 2^{n[h(\theta) - h(p) - (\theta-p)h'(\theta)]}. \tag{8.25}$$

Так как единственный максимум (8.25) имеет место при $\theta = p$, то с помощью (8.21), (8.23), (8.24) и (8.25) получим

$$P = p_x(\mathcal{E}_x^{0,\,[n(p+\varepsilon)]}) \approx p_x(\mathcal{E}_x^{[n(p\pm\varepsilon)]}) > 1 - e^{\left[-n\frac{\varepsilon^2}{3pq} + O(\log n)\right]}. \tag{8.26}$$

Таким образом, при любом $0 < \varepsilon < \min\left(p\frac{q}{q-p},\ 0{,}5-p\right)$ с ростом $n \to \infty$,

$$p_x(\mathcal{E}_x^{0,[n(p+\varepsilon)]}) \text{ и } p_x(\mathcal{E}_x^{[n(p\pm\varepsilon)]}) \to 1,$$

---

* В дальнейшем изложении этой главы символ log используется для обозначения логарифма по основанию 2 и приближение $K(n) \approx 2^{nk}$ понимается в смысле $K(n) = 2^{nk + O(\log n)}$.

т. е. множества $\mathscr{E}_x^{0,[n(p+\varepsilon)]}$ и $\mathscr{E}_x^{[n(p\pm\varepsilon)]}$ при любом $0 \leqslant \varepsilon \leqslant \min\left(p\frac{q}{q-p}, 0{,}5-p\right)$ P-представляют x. С другой стороны, используя (8.22), видим

$$N(\mathscr{E}_x^{0,[n(p+\varepsilon)]}) \approx N(\mathscr{E}_x^{[n(p\pm\varepsilon)]}) \approx 2^{nh(p+\varepsilon)}.$$

Можно показать, что x не может представлять никакое множество $\mathscr{E} \subset R^*$ с числом элементов $N(\mathscr{E}) \leqslant 2^{nh(p-\varepsilon)}$ и в остальном произвольное. Иначе говоря, имеет место следующая теорема.

Теорема 8.1. Каково бы ни было $\mathscr{E} \subset R^*$, если $N(\mathscr{E}) \leqslant 2^{nh(p-\varepsilon)}$, то $\mathscr{E}$ не представляет x. Другими словами, необходимым условием того, чтобы $\mathscr{E}$ P-представляло x, является $N(\mathscr{E}) > 2^{nh(p+\varepsilon)}$, где $0 \leqslant \varepsilon \leqslant \min\left(p\frac{q}{q-p}, 0{,}5-p\right)$.

Доказательство. Пусть задано произвольное $\mathscr{E}$ с $N(\mathscr{E}) \leqslant 2^{nh(p-\varepsilon)}$, где $0 \leqslant \varepsilon \leqslant \min\left(p\frac{q}{q-p}, 0{,}5-p\right)$, соотнесенное данному x. Соотнесем тому же x $\mathscr{E}_x^{[n(p\pm\varepsilon)]}$, представляющее его. Тогда

$$\mathscr{E} = \mathscr{E}' \cup \mathscr{E}'' (\mathscr{E}' \cap \mathscr{E}'' = \varnothing), \tag{8.27}$$

где

$$\mathscr{E}' = \mathscr{E} \cap \mathscr{E}_x^{[n(p\pm\varepsilon)]} \subset \mathscr{E}_x^{[n(p\pm\varepsilon)]}$$

и

$$\mathscr{E}'' = \mathscr{E} \cap (R^* - \mathscr{E}_x^{[n(p\pm\varepsilon)]}) \subset R^* - \mathscr{E}_x^{[n(p\pm\varepsilon)]}.$$

Из (8.27) и (8.26) следует, что

$$p_x(\mathscr{E}) = p_x(\mathscr{E}') + p_x(\mathscr{E}'') \tag{8.28}$$

и

$$p_x(\mathscr{E}'') \leqslant p_x(R^* - \mathscr{E}_x^{[n(p\pm\varepsilon)]}) \leqslant e^{\left[-n\frac{\varepsilon^2}{3pq}+O(\log n)\right]}. \tag{8.29}$$

Произведем теперь оценку сверху вероятности $p_x(\mathscr{E}')$. На основании (8.7) будем иметь, учитывая, что $p \leqslant 0{,}5 \leqslant q$,

$$p_x(\mathscr{E}') = \sum_{y\in\mathscr{E}'} p^{d(x,y)} q^{(n-d(x,y)} \leqslant \sum_{y\in\mathscr{E}'} p^{\min\limits_{y\in\mathscr{E}'} d(x,y)} q^{n-\min\limits_{y\in\mathscr{E}'} d(x,y)} =$$

$$= p^{\min\limits_{y\in\mathscr{E}'} d(x,y)} q^{n-\min\limits_{y\in\mathscr{E}'} d(x,y)} \sum_{y\in\mathscr{E}'} 1 \leqslant p^{\min\limits_{y\in\mathscr{E}_x^{[n\pm\varepsilon)]}} d(x,y)} q^{n-\min\limits_{y\in\mathscr{E}_x^{[n(p\pm\varepsilon)]}} d(x,y)} N(\mathscr{E}') \leqslant$$

$$\leqslant p^{[n(p-\varepsilon)]} q^{n-[n(p-\varepsilon)]} \cdot 2^{nh(p-\varepsilon)} = 2^{n[h(p-\varepsilon)-h(p)+\varepsilon h'(p)]}.$$

Таким образом, используя неравенство (8.23), будем иметь

$$p_x(\mathscr{E}') < e^{-n\frac{\varepsilon^2}{2pq}+O(\log n)}. \tag{8.30}$$

Из (8.28) и оценок (8.29) и (8.30) получим

$$p_x(\mathscr{E}) < e^{-n\frac{\varepsilon^2}{3pq}+O(\log n)}.$$

Отсюда заключаем, что с ростом $n \to \infty$ $p_x(\mathscr{E}) \to 0$, т. е. $\mathscr{E}$ не представляет x, что и требовалось доказать.

Из теоремы 1 можно вывести ряд важных следствий.

Следствие 8.1.1. Если $\mathscr{E}$ P-представляет x, то

$$N(\mathscr{E}) \geqslant 2^{nh(p)}.$$

Доказательство. Предположим, что $\mathscr{E} \subset R^*$ P-представляет x и $N(\mathscr{E}) < 2^{nh(p)}$; в этом случае можно найти такое $\varepsilon > 0$, при котором $N(\mathscr{E}) \leqslant \leqslant 2^{nh(p-\varepsilon)} < 2^{nh(p)}$, что противоречит условию теоремы 8.1. Поэтому остается возможность лишь $N(\mathscr{E}) \geqslant 2^{nh(p)}$, что и доказывает следствие.

Следствие 8.1.2. Число M P-различимых входных слов длины n не может превзойти числа $M_p = 2^{n[1-h(p)]}$, т. е.

$$M \leqslant M_p.$$

Доказательство. Так как по определению множества $\tilde{\mathscr{E}}_1, \tilde{\mathscr{E}}_2, \ldots, \tilde{\mathscr{E}}_M$ P-представляющие M входных P-различимых слов длины n не пересекаются, то

$$R^* \supseteq \bigcup_{i=1}^{M} \tilde{\mathscr{E}}_i \text{ и } \tilde{\mathscr{E}}_i \cap \tilde{\mathscr{E}}_j = \varnothing, \text{ при } i \neq j.$$

Поэтому $2^n = N(R^*) \geqslant \sum_{i=1}^{M} N(\tilde{\mathscr{E}}_i)$, но согласно следствию 8.1.1 $N(\tilde{\mathscr{E}}_i) \geqslant$ $\geqslant 2^{nh(p)}$, откуда

$$2^n \geqslant \sum_{i=1}^{M} N(\tilde{\mathscr{E}}_i) \geqslant \sum_{i=1}^{M} 2^{nh(p)} = 2^{nh(p)} M.$$

Последнее соотношение приводит к утверждению следствия

$$M \leqslant 2^{n[1-h(p)]}.$$

Следствие 8.2.3. Если удалить из множества $\mathscr{E}$, P-представляющего x, произвольную часть $\mathscr{E}'$ с числом элементов $N(\mathscr{E}') \leqslant 2^{nh(p-\varepsilon)}$, то оставшаяся часть $\mathscr{E}'' = \mathscr{E} - \mathscr{E}'$ множества $\mathscr{E}$ имеет вероятность

$$p_x(\mathscr{E}'') > p_x(\mathscr{E}) - e^{-n\frac{\varepsilon^2}{3pq} + O(\log n)}$$

и по-прежнему P-представляет x.

Доказательство. Из условия имеем

$$p_x(\mathscr{E}) = p_x(\mathscr{E}') + p_x(\mathscr{E}''),$$

откуда, используя (8.30), получим

$$p_x(\mathscr{E}'') = p_x(\mathscr{E}) - p_x(\mathscr{E}') > p_x(\mathscr{E}) - e^{-n\frac{\varepsilon^2}{3pq} + O(\log n)}.$$

Так как $\mathscr{E}$ P-представляет x, то $p_x(\mathscr{E}) \to 1$ при $n \to \infty$, а $p_x(\mathscr{E}') \to 1$, что и доказывает следствие.

§ 8.4. Основная лемма о пересечениях

В § 8.3 было показано, что $\mathscr{E}_x^{0,[(p+\varepsilon)n]}$ P-представляет $x (P > 1 - e^{-\frac{\varepsilon^2}{3pq} + O(\log n)})$ и имеет число элементов $N(\mathscr{E}_x^{0,[n(p+\varepsilon)]}) \approx 2^{nh(p+\varepsilon)}$. Поэтому для того чтобы M входных слов длины n

$$x_1, x_2, \ldots, x_i, \ldots, x_M$$

были P-различимыми, достаточно, чтобы пересечения

$$\mathscr{E}_i' = \mathscr{E}_i^{0,[n(p+\varepsilon)]} \cap \bigcup_{j\neq i} \mathscr{E}_j^{0,[n(p+\varepsilon)]}$$

не представляли, а дополнения $\mathscr{E}_i'$ P-представляли их, соответственно.

С другой стороны, из-за возможных пересечений $\mathscr{E}_i^{0,[n(p+\varepsilon)]}$ при заполнении ими всего R^* следует, что

$$\sum_{i=1}^{M} N(\mathscr{E}_i^{0,[n(p+\varepsilon)]}) \geqslant N(R^*) = 2^n,$$

откуда

$$\sum_{i=1}^{M} N(\mathscr{E}_{x_i}^{0,[n(p+\varepsilon)]}) = \sum_{i=1}^{M} 2^{nh(p+\varepsilon)+O(\log n)} =$$

$$= M \cdot 2^{nh(p+\varepsilon)+O(\log n)} \geqslant 2^n \text{ и } 2^{n[1-h(p-\varepsilon)]-O(\log n)} \leqslant M \leqslant M_p.$$

Таким образом, при достаточно малом ε мы можем сколь угодно приблизиться к максимально возможному числу длинных входных P-различимых слов длины n, необходимому для создания оптимального кода.

Приступим к изучению пересечений $\mathscr{E}_x^{0,[n(p+\varepsilon)]}$ при различных x, но предварительно изучим пересечения $\mathscr{E}_x^d$ при различных x и d.

Воспользуемся соотношением

$$\mathscr{E}_x^d = \bigcup_f \mathscr{E}_{x'}^f \times \mathscr{E}_{x-x'}^{d-f}, \tag{8.31}$$

где x' — произвольное подмножество x, а символ $\times$ означает взятие всевозможных пар элементов множеств $\mathscr{E}_{x'}^f$ и $\mathscr{E}_{x-x'}^{d-f}$ и образование из них новых «составных» элементов с сохранением принятого упорядочения.

Соотношение (8.31) легко выводится [7] на основе (8.15).

Заметим, что до сих пор мы определяли расстояние $d = d(x, y)$ только между $x \in R$ и $y \in R^*$. Однако ничто не мешает нам рассматривать расстояния $d = d(x_1, x_2)$ между x_1 и $x_2 \in R$, а также определять в этом случае множества $\mathscr{E}_x^d \subset R$. Не вводя специальных обозначений для отличия $\mathscr{E}_x^d \subset R$ и $\mathscr{E}_x^d \subset R^*$, мы при необходимости будем различать эти две возможности.

Основная Лемма 8.1 [7]. Пересечение $\mathscr{E}_x^{d_1}$ и $\mathscr{E}_{x_2}^{d_2}$ с R^* может быть представлено в форме

$$\mathscr{E}_{x_1}^{d_1} \cap \mathscr{E}_{x_2}^{d_2} = \begin{cases} \mathscr{E}_{x_1-(x_1\cap x_2)}^{\frac{d_1-d_2+d}{2}} \times \mathscr{E}_{x_1\cap x_2}^{\frac{d_1+d_2-d}{2}}, & \text{если } d_1+d_2+d \equiv 0 \bmod 2 \\ \varnothing, & \text{если } d_1+d_2+d \equiv 1 \bmod 2, \end{cases} \tag{8.32}$$

где $d = d(x_1, x_2)$.

Доказательство. Заметим, что $x_2 - x_1 \cap x_2 = (x_1 - x_1 \cap x_2) + \overline{1} \bmod 2$, поэтому, используя (8.31) и (8.14), будем иметь:

$$\mathscr{E}_{x_1}^{d_1} = \bigcup_{f_1} \mathscr{E}_{x_1\cap x_2}^{f_1} \times \mathscr{E}_{x_1-x_1\cap x_2}^{d_2-f_1};$$

$$\mathscr{E}_{x_2}^{d_2} = \bigcup_{f_2} \mathscr{E}_{x_1\cap x_2}^{f_2} \times \mathscr{E}_{x_2-x_1\cap x_2}^{d_2-f_2} = \bigcup_{f_2} \mathscr{E}_{x_1\cap x_2}^{f_2} \times \mathscr{E}_{x_1-x_1\cap x_2}^{d-(d_2-f_2)}.$$

Используя дистрибутивность операций $\bigcup$, $\bigcap$ и $\times$, а также (8.20) и предыдущие представления $\mathscr{E}_{x_1}^{d_1}$ и $\mathscr{E}_{x_2}^{d_2}$, будем иметь

$$\mathscr{E}_{x_1}^{d_1} \bigcap \mathscr{E}_{x_2}^{d_2} = \bigcup_{f_1, f_2} \mathscr{E}_{x_1 \cap x_2}^{f_1} \bigcap \mathscr{E}_{x_1 \cap x_2}^{f_2} \times \mathscr{E}_{x_1 - x_1 \cap x_2}^{d_1 - f_1} \bigcap \mathscr{E}_{x_1 - x_1 \cap x_2}^{d - (d_2 - f_2)} =$$

$$= \begin{cases} \mathscr{E}_{x_1 \cap x_2}^{f_1} \times \mathscr{E}_{x_1 - x_1 \cap x_2}^{d_1 - f_1} & \text{при } f_1 = f_2 \text{ и } d_1 - f_1 = d - (d_2 - f_2), \\ \varnothing & \text{в остальных случаях.} \end{cases}$$

Итак, непустое пересечение имеет место лишь в случае, когда $f_1 = f_2$ и $d_1 - f_1 = d - (d_2 - f_2)$, откуда $d_1 - d_2 + d = 2(d_1 - f_1)$, т. е. $d_1 + d_2 + d \equiv 0 \bmod 2$, при этом $f_1 = \frac{d_1 + d_2 - d}{2}$ и $d_1 - f_1 = \frac{d_1 - d_2 + d}{2}$, что и доказывает лемму.

Соотношение (8.32) позволяет вывести свойство $d = d(x, y)$, аналогичное неравенству треугольника

$$d(x_1, x_2) \leqslant d(x_1, y) + d(y, x_2) \leqslant 2n - d(x_1, x_2).$$

Из отношений (8.17) и (8.32) и определения символа $\times$ следует, что

$$N(\mathscr{E}_{x_1}^{d_1} \bigcap \mathscr{E}_{x_2}^{d_2}) = \begin{cases} C_d^{\frac{d_1 - d_2 + d}{2}} C_{n-d}^{\frac{d_1 + d_2 - d}{2}} & \text{если } d_1 + d_2 + d \equiv 0 \bmod 2, \\ 0, & \text{если } d_1 + d_2 + d \equiv 1 \bmod 2, \end{cases} \tag{8.33}$$

где $d = d(x_1, x_2)$. Соотношение (8.33) было получено в п. 1.7.4 с помощью производящих.

Используя (8.19) и (8.32), легко показать, что

$$\mathscr{E}_{x_1}^{d_1', d_1''} \bigcap \mathscr{E}_{x_2}^{d_2', d_2''} = \bigcup \mathscr{E}_{x_1 - x_1 \cap x_2}^{\frac{d_1 - d_2 + d}{2}} \times \mathscr{E}_{x_1 \cap x_2}^{\frac{d_1 + d_2 - d}{2}}, \tag{8.34}$$

$$d_1' \leqslant d_1 \leqslant d_1'',$$

$$d_2' \leqslant d_2 \leqslant d_2'',$$

$$d_1 + d_2 + d \equiv 0 \bmod 2.$$

Из (8.33) и (8.34) следует

$$N(\mathscr{E}_{x_1}^{d_1', d_1''} \bigcap \mathscr{E}_{x_2}^{d_2', d_2''}) = \sum C_d^{\frac{d_1 - d_2 + d}{2}} C_{n-d}^{\frac{d_1 + d_2 - d}{2}},$$

$$d_1' \leqslant d_1 \leqslant d_2'', \tag{8.35}$$

$$d_2' \leqslant d_2 \leqslant d_2'',$$

$$d_1 + d_2 + d \equiv 0 \bmod 2.$$

Оценим величину (8.35) для случая $d_1'' = d_2'' < [0{,}5n]$.

В этом случае легко показать, что

$$C_d^{\frac{d}{2}} C_{n-d}^{\frac{2d_1'' - d}{2}} \leqslant N(\mathscr{E}_{x_1}^{d_1', d_1''} \bigcap \mathscr{E}_{x_2}^{d_2'', d_2''}) \leqslant$$

$$\leqslant (d_1'' - d_1' + 1)(d_1'' - d_2' + 1) C_d^{\frac{d}{2}} C_{n-d}^{\frac{2d_1'' - d}{2}}. \tag{8.36}$$

В самом деле, первый из сомножителей в (8.35) достигает максимума для всех значений $d_1 = d_2$, второй в нашем случае при $d_1 = d_2 = d_1''$, что и оправдывает неравенство (8.36).

В частном случае при $d_1' = d_2' = 0$, $d_1'' = d_2'' = [n(p-\varepsilon)]$ и $d = [n\theta]$, где $0 < \theta < 2(p+\varepsilon)$, имеем

$$C_{[n\theta]}^{\left[n\frac{\theta}{2}\right]} \cdot C_{[n(1-\theta)]}^{\left[n\frac{2(p+\varepsilon)-\theta}{2}\right]} \leqslant N\left(\mathscr{C}_{x_1}^{0,[n(p+\varepsilon)]} \cap \mathscr{C}_{x_2}^{0,[n(p+\varepsilon)]}\right) \leqslant$$

$$\leqslant \{[2n(p+\varepsilon)]+1\}^2 C_{[n\theta]}^{\left[n\frac{\theta}{2}\right]} C_{[n(1-\theta)]}^{\left[n\frac{2(p+\varepsilon)-\theta}{2}\right]}. \tag{8.37}$$

Откуда при $n \to \infty$, используя (8.22), будем иметь

$$N\left(\mathscr{C}_{x_1}^{0,[n(p+\varepsilon)]} \cap \mathscr{C}_{x_2}^{0,[n(p+\varepsilon)]}\right) \approx 2^{n\left(\theta+(1-\theta)h\left(\frac{2(p+\varepsilon)-\theta}{2(1-\theta)}\right)\right)} = 2^{ng_{p+\varepsilon}(\theta)}. \tag{8.38}$$

Проведем предварительный анализ асимптотического проведения (8.38), определяемого видом функции $g_{\theta_1}(\theta)$. Для этого воспользуемся легко проверяемым тождеством (см. (8.17))

$$\frac{C_n^d}{C_n^{d_1}} C_d^{\frac{d}{2}} C_{n-d}^{d_1-\frac{d}{2}} \equiv C_{d_1}^{\frac{d}{2}} C_{n-d_1}^{d-\frac{d}{2}} \equiv C_{d_1}^{\frac{d}{2}} C_{n-d_1}^{\frac{d}{2}}. \tag{8.39}$$

При $d = [n\theta]$, $d_1 = [n\theta_1]$, $0 < \theta < 2\theta_1$ и $n \to \infty$ из (8.39) имеем

$$h(\theta) - h(\theta_1) + \theta + (1-\theta)h\left(\frac{2\theta_1-\theta}{2(1-\theta)}\right) \equiv \theta_1 h\left(\frac{\theta}{2\theta_1}\right) + (1-\theta)h\left(\frac{\theta}{2(1-\theta_1)}\right). \tag{8.40}$$

Отсюда, вводя обозначение

$$f_{\theta_1}(\theta) = h(\theta) - \theta_1 h\left(\frac{\theta}{2\theta_1}\right) - (1-\theta_1)h\left(\frac{\theta}{2(1-\theta_1)}\right),$$

используя тождество (8.40), получим

$$g_{\theta_1}(\theta) = \theta + (1-\theta)h\left(\frac{2\theta_1-\theta}{2(1-\theta)}\right) = h(\theta_1) - f_{\theta_1}(\theta). \tag{8.41}$$

Так как при $0 \leqslant \theta \leqslant 2\theta_1 \leqslant 1$

$$f_{\theta_1}'(\theta) = -\frac{1}{2}\log\left[1 - \left(\frac{1-2\theta_1}{1-\theta}\right)\right]^2 \geqslant 0,$$

то, полагая $\theta_1 = p + \varepsilon < 0{,}5$, видим, что $g_{p+\varepsilon}(\theta)$ является монотонно убывающей функцией с ростом θ, с максимумом при $\theta = 0$, где $g_{p+\varepsilon}(0) = h(p+\varepsilon)$, и минимумом при $\theta = 2(p+\varepsilon)$, где $g_{p+\varepsilon}(2(p+\varepsilon)) = 2(p+\varepsilon)$.

Функция $f_{\theta_1}(\theta)$ имеет важное значение для последующего изложения. На рис. 8.1 приведены ее графики.

Из (8.32) следует, что при $d(x_1, x_2) \geqslant 2[n(p+\varepsilon)]$

$$\mathscr{C}_{x_1}^{0,[n,(p+\varepsilon)]} \cap \mathscr{C}_{x_2}^{0,[n(p+\varepsilon)]} = \varnothing.$$

Таким образом, множества $\mathscr{C}_{x_i}^{0,[n(p+\varepsilon)]}$ $(i = \overline{1, M})$ пересекаются лишь при $d(x_i, x_j) < 2[n(p+\varepsilon)]$, и если бы мы решили в качестве M P-различимых

входных слов использовать $x_1, x_2, \ldots, x_i, \ldots, x_j, \ldots, x_M$ с условием $d(x_i, x_j) \geqslant 2[n(p+\varepsilon)]$, то, как легко видеть, их число оценивалось бы сверху [38] так:

$$M \leqslant \begin{cases} \dfrac{2^n}{\sum\limits_{d=0}^{2[n(p+\varepsilon)]} C_n^d} \approx 2^{n[1-h(2(p+\varepsilon))]}, & \text{при } 0 \leqslant p+\varepsilon < 0{,}25, \\ 0, & \text{при } 0{,}25 \leqslant p+\varepsilon \leqslant 0{,}5, \end{cases} \tag{8.42}$$

что существенно меньше максимального числа M_p P-различимых входных слов длины n (см. следствие 8.1.2). Поэтому для приближения к M нужно отказаться от требования

$$d(x_i, x_j) \geqslant 2[n(p+\varepsilon)],$$

т. е. допускать пересечения множеств $\mathscr{E}_{x_i}^{0,[n(p+\varepsilon)]}$.

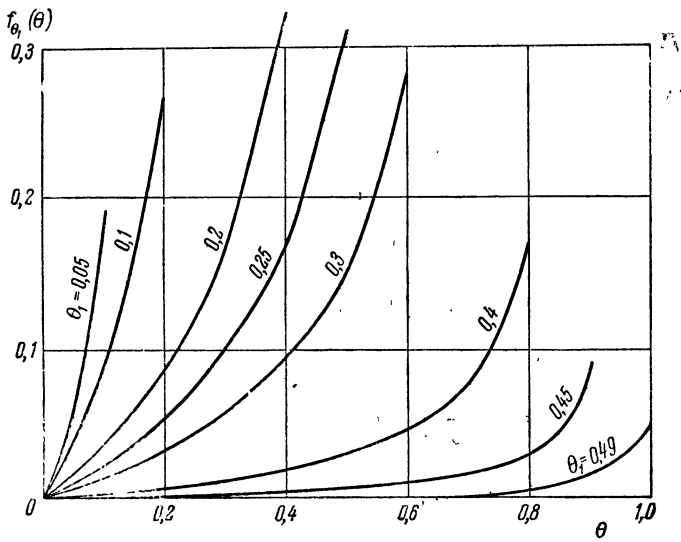


Рис. 8.1. Графики функций $f_{\theta_1}(\theta)$

В § 8.5 будет показано, что такого рода пересечения при определенных условиях не мешают осуществлению оптимального кодирования.

Отметим попутно, что отказ от пересечений можно оправдать лишь в случае очень малых вероятностей p, когда можно считать

$$p = \frac{\lambda}{n} \to 0,$$

где $\lambda < 1$.

Этот случай в отличие от рассматриваемого, где p может быть произвольно мало, но не должно зависеть от n, можно назвать случаем Пуассона. В этом случае

$$p_x(\mathscr{E}_x^d) \approx \frac{\lambda^d}{d!} e^{-\lambda} = P_\lambda(d),$$

и множества, представляющие x, как легко видеть, имеют вид

$$\mathscr{E}_x^{0,[n\varepsilon]} = \bigcup_{d=0}^{[n\varepsilon]} \mathscr{E}_x^d,$$

где ε — произвольное положительное число, заключенное в интервале $e/n < \varepsilon < 1$.

Используя (8.21) и соотношение (1.86), можно показать, что

$$p_x(\mathscr{E}_x^{0,[n\varepsilon]}) \sim 1 - 2^{-\varepsilon n \log n + O(n)}, \tag{8.43}$$

при $n \to \infty$ вероятность (8.43) стремится к 1.

С другой стороны, при $d(x_i, x_j) > 2[n\varepsilon]$ $\mathscr{E}_{x_i}^{0,[n\varepsilon]}$ и $\mathscr{E}_{x_j}^{0,[n\varepsilon]}$, согласно (8.32), не пересекаются.

Как уже отмечалось, число M входных слов в этом случае имеет оценку (8.42)

$$M \leqslant 2^{n(1-h(2\varepsilon))+O(\log n)},$$

а максимальное их число здесь равно $M_0 = 2^{n(1-h(0))} = 2^n$.

Таким образом, при достаточно малом ε мы можем сколь угодно близко подойти к оптимальному коду.

§ 8.5. Достаточные условия P-различимости

Во всех дальнейших построениях существенно используется ряд свойств функции $g_{\theta_1}(\theta)$, определяющей асимптотическое поведение числа элементов пересечения множеств $\mathscr{E}_x^{0,[n(p+\varepsilon)]}$ [см. (8.38)].

Изучим подробнее функцию $g_{\theta_1}(\theta)$ и некоторые связанные с ней функции. В предыдущем параграфе мы определили функцию $f_{\theta_1}(\theta)$ соотношением (8.41), которое дает простую связь ее с $g_{\theta_1}(\theta)$. Введем в рассмотрение функцию

$$h_{\theta_1}(\theta) = h(\theta) - h(\theta_1).$$

Кроме того, согласно (8.41), имеем

$$f_{\theta_1}(\theta) = h(\theta) - \theta_1 h\left(\frac{\theta}{2\theta_1}\right) - (1-\theta_1) h\left(\frac{\theta}{2(1-\theta_1)}\right) = h(\theta_1) - \theta -$$

$$-(1-\theta) h\left(\frac{2\theta_1 - \theta}{2(1-\theta)}\right) = h(\theta_1) - g_{\theta_1}(\theta).$$

Далее, ограничиваясь интервалом $0 < \theta < 0{,}5$, где

$$h_{\theta_1}(\theta) = h'(\theta) > 0 \text{ и } h''_{\theta_1}(\theta) = h''(\theta) \leqslant 0, \tag{8.44}$$

имеем

$$f'_{\theta_1}(\theta) = \frac{1}{2} \log \frac{(1-\theta)^2}{(1-\theta)^2 - (1-2\theta_1)^2} \geqslant 0, \tag{8.45}$$

$$f''_{\theta_1}(\theta) = \frac{1}{\ln 2} \frac{(1-2\theta_1)^2}{(1-\theta)[(1-\theta)^2 - (1-2\theta_1)^2]} \geqslant 0. \tag{8.46}$$

Последние две производные, так же как и сама функция, определены для θ в пределах $0 \leqslant \theta \leqslant 2\theta_1$.

Из соотношений (8.44), (8.45) и (8.46) следует, что на указанных интервалах изменения θ с ростом функции $h_{\theta_1}(\theta)$ и $f_{\theta_1}(\theta)$ монотонно возрастают. Функция $h_{\theta_1}(\theta)$ обращена выпуклостью вверх и имеет максимум в точке $\theta = 0{,}5$, а функция $f_{\theta_1}(\theta)$ обращена выпуклостью вниз и достигает наибольшего значения при $\theta = 2\theta_1$.

Заметим, что в точке $\theta = 2\theta_1(1-\theta_1)$

$$h_{\theta_1}(2\theta_1(1-\theta_1)) = f_{\theta_1}(2\theta_1(1-\theta_1)) = h(2\theta_1(1-\theta_1) - h(\theta_1)$$

и

$$h'_{\theta_1}(2\theta_1(1-\theta_1)) = f'_{\theta_1}(2\theta_1(1-\theta_1)) = h'(2\theta_1(1-\theta_1)) = \log\frac{1-2\theta_1(1-\theta_1)}{2\theta_1(1-\theta_1)}.$$

Отсюда следует, что

$$h_{\theta_1}(\theta) \leqslant f_{\theta_1}(\theta), \tag{8.47}$$

и знак равенства в (8.47) имеет место лишь в точке $\theta = 2\theta_1(1-\theta_1)$, где кривые $h_{\theta_1}(\theta)$ и $f_{\theta_1}(\theta)$ касаются друг друга.

Проведем через общую точку наших кривых касательную, имеющую уравнение

$$l_{\theta_1}(\theta) = h(2\theta_1(1-\theta_1)) - h(\theta_1) + (\theta - 2\theta_1(1-\theta_1))h'(2\theta_1(1-\theta_1)). \tag{8.48}$$

Ясно, что

$$h_{\theta_1}(\theta) \leqslant l_{\theta_1}(\theta) \leqslant f_{\theta_1}(\theta).$$

Для дальнейшего изложения важно оценить сверху абсциссу $\theta' < 0{,}5$ точки пересечения кривой $f_{\theta_1}(\theta)$ с прямой $l(\theta) = 1 - h(\theta_1) = \text{const}$, параллельной оси абсцисс.

Ясно, что $\theta' < \theta'' < 0{,}5$, где θ'' — абсцисса точки пересечения прямой $l_{\theta_1}(\theta)$ с прямой $l(\theta) = 1 - h(\theta_1)$. Используя (8.48), получаем

$$\theta''' = 2\theta_1(1-\theta_1) + \frac{1 - h(2\theta_1(1-\theta_1))}{h'(2\theta_1(1-\theta_1))}.$$

Представляя $2\theta_1(1-\theta_1) = 0{,}5 - 0{,}5(1-2\theta_1)^2$ и учитывая, что $h'(\theta) \leqslant \leqslant \frac{1-2\theta}{\theta \ln 2}$, будем иметь

$$\theta'' = 2\theta_1(1-\theta_1) + \frac{h(0{,}5) - h(2\theta_1(1-\theta_1))}{h'(2\theta_1(1-\theta_1))} \leqslant 2\theta_1(1-\theta_1) +$$

$$+ \frac{(0{,}5 - 2\theta_1(1-\theta_1))h'(2\theta_1(1-\theta_1)) + (1-2\theta_1)^4 h''(0{,}5)/2\cdot 5}{h'(2\theta_1(1-\theta_1))} \leqslant$$

$$\leqslant 0{,}5 - \theta_1(1-\theta_1)(1-2\theta_1)^2;$$

итак, имеем окончательную оценку

$$\theta' \leqslant \theta'' \leqslant 0{,}5 - \Delta, \tag{8.49}$$

где $\Delta = \theta_1(1-\theta_1)(1-2\theta_1)^2$.

Остановимся, наконец, на одном свойстве функции $h(\theta)$, непосредственно используемом в этом параграфе для оценок разностей функций $g_{\theta_1}(\theta)$. Это свойство описывается следующим неравенством:

$$h(\theta+\delta) - h(\theta) \leqslant h(\tilde{\theta}+\delta) - h(\tilde{\theta}) \leqslant h(\delta), \tag{8.50}$$

где $0 \leqslant \theta \leqslant \tilde{\theta} \leqslant \tilde{\theta} + \delta \leqslant 0{,}5$.

Соотношение (8.50) следует из общей теории выпуклых функций, к которым относится и функция $h(\theta)$. Однако это и целый ряд других полез-

ных соотношений для $h(\theta)$ можно получить непосредственно из соотношения (8.2.2), пользуясь «комбинаторным» происхождением $h(\theta)$.

В самом деле, из соотношения (8.31) между $\mathscr{E}_x^d$ следует аналогичное соотношение для C_n^d при любом фиксированном ν $(0 \leqslant \nu \leqslant n)$:

$$C_n^d = \sum_f C_\nu^f C_{n-\nu}^{d-f}. \tag{8.51}$$

Поэтому $C_\nu^f C_{n-\nu}^{d-f} \leqslant C_n^d$,

откуда, используя асимптотическое выражение (8.2.2), получим соотношение

$$\alpha h(\theta_1) + (1-\alpha) h(\theta_3) \leqslant h[\alpha\theta_1 + (1-\alpha)\theta_3] \tag{8.52}$$

для любых $0 \leqslant \alpha, \theta_1, \theta_3 \leqslant 1$.

Далее при $\theta_1 \leqslant \alpha\theta_1 + (1-\alpha)\theta_3 = \theta_2 \leqslant \theta_3 \leqslant 0{,}5$, получим из (8.52)

$$\frac{h(\theta_2) - h(\theta_1)}{\theta_2 - \theta_1} \geqslant \frac{h(\theta_3) - h(\theta_2)}{\theta_3 - \theta_2}. \tag{8.53}$$

Если $0 < \theta_1 \leqslant \theta_2 \leqslant \theta_3 \leqslant \theta_4 \leqslant 0{,}5$, то из (8.53) следует

$$\frac{h(\theta_2) - h(\theta_1)}{\theta_2 - \theta_1} \geqslant \frac{h(\theta_4) - h(\theta_3)}{\theta_4 - \theta_3}. \tag{8.54}$$

Полагая в (8.54) $\theta_1 = \theta, \theta_2 = \theta + \delta, \theta_3 = \tilde{\theta}$ и $\theta_4 = \tilde{\theta} + \delta$, получим неравенства (8.50).

После сделанных предварительных замечаний перейдем к систематическому изучению числовых характеристик взаимных расстояний совокупности входных слов длины n. Важность такого изучения вытекает из результатов предыдущего параграфа, где установлена связь расстояния между входными словами длины n и числом элементов пересечения соответственно P-представляющих их множеств $\mathscr{E}_{x_i}^{0,[n(p+\varepsilon)]}$ выходных слов (8.38).

Рассмотрим совокупность R' входных слов длины n

$$(x_1, x_2, \ldots, x_i, \ldots, x_j, \ldots, x_M) = R' \subset R.$$

Выберем некоторое x_i и рассмотрим разбиение R на непересекающиеся множества $\mathscr{E}_{x_i}^d$

$$R = \bigcup_{d=0}^{n} \mathscr{E}_{x_i}^d. \tag{8.55}$$

Беря пересечения обеих частей (8.55) с R' и, обозначая $\mathscr{E}_{x_i}^d \cap R' = \mathscr{E}_{x_i}^{\prime d}$, будем иметь

$$R' = \bigcup_{d=0}^{n} \mathscr{E}_{x_i}^{\prime d}. \tag{8.56}$$

Соотношение (8.56) является разбиением R' на непересекающиеся множества $\mathscr{E}_{x_i}^{\prime d}$ входных слов длины n, равноотстоящих от $x_i \in R'$.

Введем для $N(\mathscr{E}_{x_i}^{\prime d})$ специальное обозначение

$$N(\mathscr{E}_{x_i}^{\prime d}) = M_{x_i}^d. \tag{8.57}$$

Тогда $\sum_{d=0}^{n} M_{x_i}^d = M - 1$ и набор чисел $\{M_{x_i}^d\}$ $(1 = \overline{1,M},\ d = \overline{0,n})$ полностью характеризует совокупность R' с точки зрения взаимных расстояний $d(x_i, x_j)$, входящих в R' входных слов длины n.

Зафиксируем некоторое $\varepsilon \left(0 < \varepsilon < \min\left(p\frac{p}{q-p},\ 0{,}5 - p\right)\right)$ и соотнесем каждому входному слову $x_i \in R'$ P-представляющее его множество $\mathscr{E}_{x_i}^{0,[n(p+\varepsilon)]} \subset R^*$ $(i = \overline{1\ M})$. Выберем некоторое x_i и составим пересечение вида

$$\mathscr{E}_{x_i}^{\prime 0,\ [n\ (p+\varepsilon)]} = \mathscr{E}_{x_i}^{0,\ [n\ (p+\varepsilon)]} \cap \bigcup_{i \neq j} \mathscr{E}_{x_j}^{0,\ [n\ (p+\varepsilon)]}.$$

Так как $\mathscr{E}_{x_j}^{0,\ [n\ (p-\varepsilon)]}$ $(j \neq i)$ могут иметь пересечения между собой, то имеет место соотношение

$$\mathscr{E}_{x_i}^{\prime 0,\ [n\ (p+\varepsilon)]} = \mathscr{E}_{x_i}^{0,\ [n\ (p+\varepsilon)]} \cap \bigcup_{j \neq i} \mathscr{E}_{x_j}^{0,\ [n\ (p+\varepsilon)]} \subseteq \bigcup_{j \neq i} \mathscr{E}_{x_i}^{0,\ [n\ (p+\varepsilon)]} \cap \mathscr{E}_{x_j}^{0,\ [n\ (p+\varepsilon)]},$$

откуда

$$N(\mathscr{E}_{x_i}^{\prime 0,\ [n\ (p+\varepsilon)]}) \leqslant \sum_{j \neq i} N(\mathscr{E}_{x_i}^{0,\ [n\ (p+\varepsilon)]} \cap \mathscr{E}_{x_j}^{0,\ [n\ (p+\varepsilon)]}) =$$

$$= \sum_{d=0}^{n} \sum_{x_j \in \mathscr{E}_{x_i}^{\prime d}} N(\mathscr{E}_{x_i}^{0,\ [n\ (p+\varepsilon)]} \cap \mathscr{E}_{x_j}^{0,\ [n\ (p+\varepsilon)]}).$$

Используя соотношение (8.37), будем иметь

$$N(\mathscr{E}_{x_i}^{\prime 0,\ [n\ (p+\varepsilon)]}) \leqslant \sum_{d=0}^{2\ [n\ (p+\varepsilon)]} \sum_{x_j \in \mathscr{E}_{x_i}^{\prime d}} ([2n\ (p+\varepsilon)] + 1)^2 C_{n-d}^{[n\ (p+\varepsilon)] - \frac{d}{2}} =$$

$$= \sum_{d=0}^{2\ [n\ (p+\varepsilon)]} ([2n\ (p+\varepsilon)] + 1)^2 C_d^{\frac{d}{2}}\ C_{n-d}^{[n\ (p+\varepsilon)] - \frac{d}{2}} N(\mathscr{E}_{x_i}^{\prime d}).$$

Вводя обозначение (8.57), получим окончательно

$$N(\mathscr{E}_{x_i}^{\prime 0,\ [n\ (p+\varepsilon)]}) \leqslant \sum_{d=0}^{2\ [n\ (p+\varepsilon)]} M_{x_i}^d C_n^{\frac{d}{2}} C_{n-d}^{n\ (p+\varepsilon) - \frac{d}{2}} ([2n\ (p+\varepsilon)] + 1)^2. \quad (8.58)$$

В асимптотическом случае при $d = [n\theta]$ и $n \to \infty$ согласно (8.38)

$$N(\mathscr{E}_{x_i}^{\prime 0,\ [n\ (p+\varepsilon)]}) \leqslant \sum_{d=0}^{2\ [n\ (p+\varepsilon)]} M_{x_i}^d \cdot 2^{ng_{p+\varepsilon}\left(\frac{d}{n}\right) + O(\log n)}. \quad (8.59)$$

Так как, согласно (8.26), множества $\mathscr{E}_{x_i}^{0,\ [n\ (p+\varepsilon)]}$ P-представляют x_i, то для того, чтобы последние были P-различимыми, нужно, чтобы множества $\mathscr{E}_{x_i}^{\prime 0,\ [n\ (p+\varepsilon)]}$ их P-представляли, а для этого, согласно теореме 8.1, достаточно, чтобы

$$N(\mathscr{E}_{x_i}^{\prime 0,\ [n\ (p+\varepsilon)]}) \leqslant 2^{nh\ (p-\varepsilon)}.$$

Последнее требование, согласно (8.59), будет выполняться, если потребовать от $\{M_{x_i}^d\}$, чтобы для всех $(i = \overline{1,M})$

$$\sum_{d=0}^{2[n(p+\varepsilon)]} M_{x_i}^d \, 2^{ng_{p+\varepsilon}\left(\frac{d}{n}\right) + O(\log n)} \leqslant 2^{nh(p-\varepsilon)}. \tag{8.60}$$

Произведем теперь оценку непосредственно отдельных значений $M_{x_i}^d$, при которых соблюдается условие (8.60).

Теорема 8.2. Достаточным условием P-различимости совокупности $M = 2^{n(1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon))}$ входных слов длины n является выполнение следующих соотношений для всех $i = \overline{1,M}$:

$$M_{x_i}^d \begin{cases} \equiv 0 & \text{для } 0 \leqslant d < [nf_{p+\varepsilon}^{-1}(h_{p-\varepsilon}(p+\varepsilon))], \\ \leqslant 2^{n\left(f_{p+\varepsilon}\left(\frac{d}{n}\right) - h_{p-\varepsilon}(p+\varepsilon)\right)} & \\ \leqslant 2 & \text{для } [nf_{p+\varepsilon}^{-1}(h_{p-\varepsilon}(p+\varepsilon))] \leqslant d \leqslant 2[n(p+\varepsilon)]. \end{cases} \tag{8.61}$$

Доказательство. Выше было показано, что для различимости совокупности M входных слов длины n достаточно соблюдения следующих неравенств при любых $(i = \overline{1,M})$:

$$S = \sum_{d=0}^{2[n(p+\varepsilon)]} M_{x_i}^d \, 2^{n\left(h(p+\varepsilon) - f_{p+\varepsilon}\left(\frac{d}{n}\right)\right)} \leqslant 2^{nh(p+\varepsilon)}.$$

Усилим неравенство еще больше, потребовав, чтобы выполнялось условие (8.61). В самом деле,

$$S \leqslant \sum_{d=[nf_{p+\varepsilon}^{-1}(h_{p-\varepsilon}(p+\varepsilon))]}^{2[n(p+\varepsilon)]} 2^{n\left(f_{p+\varepsilon}\left(\frac{d}{n}\right) - h_{p-\varepsilon}(p+\varepsilon)\right)} \cdot 2^{n\left(h(p+\varepsilon) - f_{p+\varepsilon}\left(\frac{d}{n}\right)\right) + O(\log n)} =$$

$$= 2^{nh(p-\varepsilon) + O(\log n)},$$

что и требовалось доказать.

Неравенство (8.60) заведомо будет иметь место, если для всех $(i = \overline{1,M})$:

$$M_{x_i}^d \begin{cases} \equiv 0 & \text{для } 0 \leqslant d < [nf_p^{-1}(h(3\varepsilon))], \\ \leqslant 2^{n\left(f_p\left(\frac{d}{n}\right) - h(3\varepsilon)\right)} & \text{для } [nf_p^{-1}(h(3\varepsilon))] \leqslant d \leqslant 2[n(p+\varepsilon)], \end{cases}$$

так как для всех $\theta\,(0 < \theta < 2(p+\varepsilon))$ имеем $f_p(\theta) - h(3\varepsilon) \leqslant f_{p+\varepsilon}(\theta) - h_{p-\varepsilon}(p+\varepsilon)$, что эквивалентно

$$f_p(\theta) - h(3\varepsilon) + g_{p+\varepsilon}(\theta) = h(p) - h(3\varepsilon) + g_{p+\varepsilon}(\theta) - g_p(\theta) \leqslant h(p-\varepsilon).$$

$$\text{Но } l = h(p) - h(3\varepsilon) + \underset{p+\varepsilon}{g}(\theta) - g_p(\theta) = h(p) - h(3\varepsilon) +$$

$$+ (1-\theta)\left[h\left(p - \frac{\theta}{1-\theta}(0{,}5-p) + \frac{\varepsilon}{1-\theta}\right) - h\left(p - \frac{\theta}{1-\theta}(0{,}5-p)\right)\right].$$

Исходя из свойства $h(\theta)$ (8.50), получим

$$l \leqslant h(p) - h(3\varepsilon) + \left[h\left(\frac{\varepsilon}{0{,}5}\right) - h(0)\right] = h(p) - h(3\varepsilon) + h(2\varepsilon) \leqslant$$

$$\leqslant h(p-\varepsilon) - h(2\varepsilon) + h(2\varepsilon) = h(p-\varepsilon),$$

что и требовалось доказать.

В заключение остановимся на одном интересном примере так называемого идеального кода. Если просуммировать по всем $d/2$ первую и вторую часть соотношения (8.39), то, используя соотношение (8.51), будем иметь

$$\sum_{\frac{d}{2}} \frac{C_n^d}{C_n^{d_1}} C_d^{\frac{d}{2}} C_{n-d}^{d_1-\frac{d}{2}} \equiv \sum_{\frac{d}{2}} C_{d_1}^{\frac{d}{2}} C_{n-d_1}^{n-d_1} \equiv C_n^{n-d_1} \equiv C_n^{d_1}.$$

Кроме того ясно, что

$$\sum_{\frac{d}{2}} \frac{C_n^d}{C_n^{d_1}} = \frac{2^n}{C_n^{d_1}}.$$

Если положить $d = [n\theta]$, $d_1 = [n(p+\varepsilon)]$ и $n \to \infty$, то будем иметь

$$\sum_d 2^{n\left[h\left(\frac{d}{n}\right) - h(p+\varepsilon)\right] + O(\log n)} 2^{ng_{p+\varepsilon}\left(\frac{d}{n}\right) + O(\log n)} \equiv 2^{nh(p+\varepsilon) + O(\log n)} \quad (8.62)$$

и

$$\sum_d 2^{n\left[h\left(\frac{d}{n}\right) - h(p+\varepsilon)\right] + O(\log n)} \equiv 2^{n(1-h(p+\varepsilon)) + O(\log n)}. \quad (8.63)$$

Умножив обе части равенств (8.62) и (8.63) на $2^{-n[h(p+\varepsilon) - h(p-\varepsilon)]}$, будем иметь

$$\sum_d 2^{n\left[h\left(\frac{d}{n}\right) - 2h(p+\varepsilon) + h(p-\varepsilon)\right] + O(\log n)} \cdot 2^{ng_{p+\varepsilon}\left(\frac{d}{n}\right) + O(\log n)} =$$

$$= 2^{nh(p-\varepsilon) + O(\log n)} \quad (8.64)$$

и

$$\sum_d 2^{n\left[h\left(\frac{d}{n}\right) - 2h(p+\varepsilon) + h(p-\varepsilon)\right] + O(\log n)} = 2^{n\left[1 - 2h(p+\varepsilon) + h(p+\varepsilon)\right] + O(\log n)}. \quad (8.65)$$

Если положить в (8.64) и (8.65)

$$M_{x_i}^d = 2^{n\left[h\left(\frac{d}{n}\right) - 2h(p+\varepsilon) + h(p-\varepsilon)\right] + O(\log n)} \leqslant 2^{n\left[f_{p+\varepsilon}\left(\frac{d}{n}\right) - h_{p-\varepsilon}(p+\varepsilon)\right]},$$

что означает (из-за целочисленности $M^d_{x_i}$) в частности

$$M^d_{x_i} \equiv 0 \text{ для } 0 \leqslant d \leqslant [n(p+\varepsilon)], \tag{8.66}$$

то на основании того, что выбранные $M^d_{x_i}$ удовлетворяют условиям теоремы 8.2 (см. (8.61)) следует, что x_i с такими $M^d_{x_i}$ будут P-различимыми. По аналогии с идеальным кодом теории корректирующих кодов удовлетворяющий условию (8.66) код будем называть идеальным.

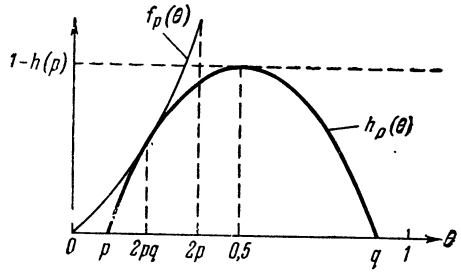


Рис. 8.2. Геометрическая интерпретация достаточных условий оптимальности

В следующем параграфе будет показано, что числа $M^d_{x_i}$ идеального когда являются математическими ожиданиями чисел M^d_{xi} случайно выбранного кода и теорема 8.2 по сути дела оценивает допустимые отклонения чисел $M^d_{x_i}$ случайно выбранного кода от их математических ожиданий в том смысле, что они не приведут к неразличимым входным словам длины n.

В предельном случае при $n \to \infty$ и $\varepsilon \to 0$, когда распределения $\mathcal{P}(\varepsilon_i/x_i)$ вырождаются в δ-функции, а случайные величины $M^d_{x_i}$ — в средние значения, рассмотренная ситуация допускает простую геометрическую интерпретацию, изображенную на рис. 8.2.

Здесь характеристикой идеального случайно выбранного кода является функция $h_p(\theta) = 1/n \log M^{[n\theta]}_{x_i}$, а достаточные условия оптимальности кода определяются кривой $f_p(\theta)$, ниже которой будут располагаться указанные характеристики кода. Интересно отметить (о чем указывалось ранее) наличие единственной точки касания этих кривых при $\theta = 2pq$.

Заметим, что до сих пор понятие идеального кода относилось лишь к случаю кодов с пустыми $2(p+\varepsilon)$-окрестностями (см. § 8.6) для всех входных слов длины n. Мы показали, что для идеального кода в нашем смысле существенны все взаимные расстояния входящих в него входных слов длины n.

§ 8.6. Получение P-различимых входных слов случайным выбором

Выше построение оптимального кода было сведено к построению максимально большой совокупности P-различимых входных слов длины n и найдены достаточные условия (8.61) для того, чтобы данная совокупность входных слов была различимой.

Условия (8.61) указывают на то, что различимых совокупностей входных слов может быть много. Нам достаточно иметь хотя бы одну из них.

Нахождение регулярного способа построения P-различимой совокупности входных слов с заранее заданными числами $\{M^d_{x_i}\}$, удовлетворяющими условиям (8.61), представляет, по-видимому, чрезвычайно сложную задачу. Но задача облегчается, если от способа построения не требовать того, чтобы он приводил к совокупности входных слов с заранее назначенными числами $\{M^d_{x_i}\}$, удовлетворяющими условиям (8.61).

Практически вполне приемлем способ построения совокупности входных слов с заранее неизвестными числами $\{M_{x_i}^d\}$, однако, гарантирующий, что эти числа со сколь угодно близкой к единице вероятностью удовлетворяют условиям (8.61). Другими словами, этот способ должен гарантировать получение одной из P-различимых совокупностей входных слов с вероятностью, сколь угодно близкой к единице. Таким, как мы увидим ниже, является случайный выбор входных слов из оптимальной генеральной совокупности с вероятностями символов 0 и 1, равными 0,5.

Произведем выборку с повторением M последовательностей, состоящих из n нулей и единиц, из генеральной совокупности нулей и единиц, содержащей тех и других в равных долях. Обозначим полученные таким образом последовательности $(x_1x_2 \ldots x_i \ldots x_j \ldots x_M) = R'$; их совокупность R' назовем выборкой. Все члены выборки будем называть выборочными значениями и рассматривать как элементы R (среди них могут быть, вообще говоря, и повторяющиеся).

Если отождествить R с R^*, то множества $\mathscr{E}_x^d$ будут иметь смысл всех $x' \in R$ с $d(x, x') = \text{const}$. Соответственно определяются и множества $\mathscr{E}_x^{d_1,d_2}$.

Сопоставим каждому x_i множество $\mathscr{E}_{x_i}^{0,[n\eta]}$ и будем называть его η-окрестностью x_i. Ясно, что общее число η-окрестностей x_i равно числу всех x_i, т. е. M.

Разобьем совокупность всех η-окрестностей x_i на два класса. К первому классу отнесем все η-окрестности, содержащие, кроме x_i, еще хотя бы один элемент $x_j \in R'$, а ко второму — все остальные пустые η-окрестности, т. е. все η-окрестности, не содержащие, кроме x_i, ни одного элемента R'. Пусть число η-окрестностей первого класса равно k, а второго — k_0 ($k + k_0 = M$). Тогда будет иметь место следующая теорема.

Теорема 8.3. Пусть $M = 2^{nH}$ $(0 < H < 1)$; тогда

$$P_0 = \mathscr{P}(k_0 > M(1-2^{-n\gamma})) > 1-2^{-n(1-H-h(\eta)-\gamma)+O(\log n)}, \qquad (8.67)$$

где $\gamma > 0$.

Доказательство. Зафиксируем некоторую η-окрестность $\mathscr{E}_{x_i}^{0,[n\eta]} x_i$. Тогда вероятность попадания в нее одного выборочного значения x_j $(i \neq j)$ равна $\sum_{d=0}^{[n\eta]} C_n^d \frac{1}{2^n} \approx 2^{-n(1-h(\eta))}$, а вероятность попадания одного выборочного значения вне ее равна $1-2^{-n(1-h(\eta))+O(\log n)}$. Вероятность того, что все $M-1$ выборочных значений попадут вне $\mathscr{E}_{x_i}^{0,[n\eta]}$, будет равна

$$(1-2^{-n(1-h(\eta))+O(\log n)})^{M-1} \geqslant 1-M\cdot 2^{-n(1-h(\eta))+O(\log n)},$$

а вероятность того, что хотя бы одно из выборочных значений, кроме x_i, попадет в $\mathscr{E}_{x_i}^{0,[n\eta]}$, равна

$$r = 1-[1-2^{-n(1-h(\eta))+O(\log n)}]^{M-1} \leqslant M\cdot 2^{-n(1-h(\eta))+O(\log n)} = 2^{-n(1-H-h(\eta))+O(\log n)}.$$

Заметим, что из-за возможных пересечений различных η-окрестностей числа попавших в них выборочных значений являются зависимыми случайными величинами. Поэтому, в частности, являются зависимыми и случайные величины ξ_i, равные нулю, если в η-окрестности x_i нет, кроме x_i, ни одного выборочного значения из R_1 или, в противном случае, равные единице. Ясно, что

$$\mathrm{E}\,\xi_i = 0\cdot(1-r)+1\cdot r = r = 2^{-n(1-H-h(\eta))+O\log n} \quad \text{и} \quad k = \sum_{i=1}^{M} \xi_i.$$

Заметим, что математическое ожидание суммы случайных величин равно сумме их математических ожиданий в случае произвольной зависимости последних, откуда

$$\mathrm{E}\,k = \sum_{i=1}^{M} \mathrm{E}\,\xi_i = \sum_{i=1}^{M} r = M \cdot r \leqslant 2^{-n\,(1-H-h\,(\eta))+O\log n)}\,M.$$

Так как $k > 0$ и $\mathrm{E}\,k < \infty$, то, используя неравенство Чебышева, будем иметь для $t > 0$;

$$\mathcal{P}\,(k < t\,\mathrm{E}\,k) = \mathcal{P}\,(k_0 \gg M - t\,\mathrm{E}\,k) > 1 - \frac{1}{t}. \tag{8.68}$$

Полагая в (8.68) $t = 2^{n\,(1-H-h\,(\eta))+\gamma+O\,(\log n)}$ и используя приведенное выше выражение для $\mathrm{E}\,k$, будем иметь

$$P = \mathcal{P}\,(k_0 \geqslant M - t\,\mathrm{E}\,k) = \mathcal{P}\,(k_0 \geqslant M\,(1-2^{-n\gamma})) > \\ > 1-2^{-n\,(1-H-h\,(\eta)-\gamma)+O\,(\log n)},$$

что и требовалось доказать.

Приступим теперь к анализу совместного распределения чисел $\{M_{x_i}^d\}$ $(i = \overline{1,M};\ d = \overline{0,n})$, являющихся в нашем случае случайными величинами (будем называть их частотами).

Если зафиксировать x_i, то $\{M_{x_i}^d\}$ являются неотрицательными целыми числами, удовлетворяющими соотношению

$$\sum_{d=0}^{n} M_{x_i}^d = M - 1.$$

Легко показать, что совместное распределение их при фиксированном x является полиномиальным

$$\mathcal{P}\,\{M_{x_i}^d = k_d\} = \frac{(M-1)!}{\prod\limits_{d=0}^{n} k_d!} \prod_{d=0}^{n} p_d^{k_d},$$

с параметрами

$$p_d = C_n^d \frac{1}{2^n} = 2^{-n\left(1-h\left(\frac{d}{n}\right)\right)+O\,(\log n)}, \quad \left(\sum_{d=0}^{n} p_d = 1\right). \tag{8.69}$$

Произведем группировку частот

$$M_{x_i}^{*} = \sum_{d=[nh^{-1}\,(1-H-\varepsilon'/2)]}^{[nh^{-1}\,(1-H+\varepsilon'/4)]} M_{x_i}^d; \quad M_{x_i}^{**} = \sum_{d=[n\,(0{,}5-\Delta)]}^{n} M_{x_i}^d, \tag{8.70}$$

при условии, что:

$$0 < \Delta < 0{,}5 - h^{-1}\,(1 - H + \varepsilon'/4) \text{ и } \varepsilon \gg \varepsilon' > 0,$$

оставив все частоты, не вошедшие в суммы без изменения. Полученные таким образом частоты обозначим через $\{M_{x_i}^d\}'$. Они снова будут иметь

полиномиальное распределение с параметрами (8.69) для негруппированных частот и с параметрами

$$p_* = \sum_{d=[nh^{-1}(1-H-\varepsilon'/2)}^{[nh^{-1}(1-H+\varepsilon'/4)]} p_d = 2^{-n(H-\varepsilon'/4)+O(\log n)} \quad \text{и} \quad p_{**} =$$

$$= \sum_{d=[n(0.5-\Delta)]}^{n} p_d > 1 - 2^{-n\Delta^2} \tag{8.71}$$

для группированных частот (8.70). Заметим, что при $n \to \infty$, $p_{**} \to 1$, и поэтому все остальные параметры стремятся к нулю.

Легко показать (см. п. 1.6.3), что в этом случае совместное распределение частот $\{M^d_{x_i}\}''$, получающееся исключением из частот $\{M^d_{x_i}\}'$ частоты $M^{**}_{x_i}$, распадается в произведение распределений Пуассона для отдельных частот с параметрами:

$\lambda_d = (M-1)p_d$ для негруппированных частот,

и

$\lambda_* = (M-1)p_*$ для группированной частоты $M^*_{x_i}$.

Положим

$$M = 2^{nH}; \tag{8.72}$$

тогда, учитывая (8.71) и (8.72), будем иметь:

$\lambda_d = 2^{n\left(h\left(\frac{d}{n}\right)-(1-H)\right)+O(\log n)}$ для негруппированных частот

и

$\lambda_* = 2^{n\frac{\varepsilon'}{4}+O(\log n)}$ для группированной частоты $M^*_{x_i}$.

Предыдущие замечания будут использованы в дальнейшем изложении. Докажем следующую теорему.

Теорема 8.4. При случайном выборе $M = 2^{nH}$ входных слов, каковы бы ни были значения $\{M^d_{x_i}\}$ для $0 \leqslant d \leqslant [n\eta]$ и $[n(0{,}5-\Delta)] \leqslant d \leqslant n$ $(i = \overline{1,M})$, вероятность P_1 одновременного осуществления для всех $\{M^d_{x_i}\}$ $([n\eta] < d < [n(0{,}5-\Delta)]$, $i = \overline{1,M})$ неравенств

$$M^d_{x_i} < K_d = \begin{cases} 2^{n\varepsilon'/2} & \text{для } [n\eta] < d\,[nh^{-1}(1-H-\varepsilon'/2)], \\ 2^{n\left(h\left(\frac{d}{n}\right)-(1-H-\varepsilon')\right)} & \text{для } [nh^{-1}(1-H+\varepsilon'/4) \leqslant d \leqslant [n(0{,}5-\Delta)] \end{cases} \tag{8.73}$$

$$M^*_{x_i} < K_* = 2^{n\varepsilon'/2}$$

имеет оценку

$$P_1 > 1 - \exp\left[-2^{n\varepsilon'/2}\ln 2 + O(\log n)\right]. \tag{8.74}$$

Если ввести ограничения для $\{M_{x_i}^d\}^*$ и в интервале $0 \leqslant d \leqslant [n\eta]$, потребовав для всех $i = \overline{1,M}$ выполнение неравенств:

$$M_{x_i}^d < K_d = \begin{cases} \dfrac{H+\delta}{1-H-h\left(\frac{d}{n}\right)} > 0 & \text{для } 0 \leqslant d \leqslant [n\eta], \\ 2^{n\varepsilon'/2} & \text{для } [n\eta] < d < [nh^{-1}(1-H-\varepsilon'/2)], \\ 2^{n\left[h\left(\frac{d}{n}\right)-(1-H-\varepsilon')\right]} & \text{для } [nh^{-1}(1-H-\varepsilon'/2)] < d < [n(0{,}5-\Delta)], \end{cases} \tag{8.75}$$

$M_{x_i}^* < K_* = 2^{n\varepsilon'/2}$,

то вероятность P_2 их совместного осуществления будет иметь оценку

$$P_2 > 1 - 2^{-n\delta + O(\log n)}.$$

Доказательство. Произведем случайную выборку $M = 2^{nH}$ входных слов и определим по ней частоты $\{M_{x_i}^d\}$. Далее зафиксируем некоторое x_i и будем различать пять последовательных интервалов изменения $d\,(0 \leqslant d \leqslant n)$:

$$I_1 = (0 \leqslant d \leqslant [n\eta]);\ I_2 = ([n\eta] < d < [nh^{-1}(1-H-\varepsilon'/2)]);$$
$$I_3 = ([nh^{-1}(1-H-\varepsilon'/2)] \leqslant d < [nh^{-1}(1-H+\varepsilon'/4)]);$$
$$I_4 = ([nh^{-1}(1-H+\varepsilon'/4)] \leqslant d < [nh^{-1}(0{,}5-\Delta)]);$$
$$I_5 = ([nh^{-1}(0{,}5-\Delta)] \leqslant d \leqslant n).$$

Соответственно будем различать пять наборов частот $\{M_{x_i}^d\}$, из которых третий и пятый сгруппированы следующим образом:

$$\{M_{x_i}^d\}_{d \in I_1};\ \{M_{x_i}^d\}_{d \in I_2};\ M_{x_i}^{**} = \sum_{d \in I_3} M_{x_i}^d;\ \{M_{x_i}^d\}_{d \in I_4};\ M_{x_i}^{**} = \sum_{d \in I_5} M_{x_i}^d.$$

Соответствующие параметры полиномиального распределения будут иметь вид:

$$p_d = 2^{-n\left(1-h\left(\frac{d}{n}\right)\right)+O(\log n)}\ (d \in I_1,\ I_2,\ I_4);\ p_* = 2^{-n(H-\varepsilon'/4)+O(\log n)};$$
$$p_{**} > 1 - 2^{-n\Delta^2+O(\log n)}.$$

Так как $p_{**} \to 1$ при $n \to \infty$, то частное распределение первых четырех наборов частот распадается в произведение функций распределений Пуассона с параметрами, соответственно:

$$\left.\begin{aligned} \lambda_d &= (M-1)\,p_d = 2^{n\left(h\left(\frac{d}{n}\right)-(1-H)\right)+O(\log n)} \\ \lambda_* &= (M-1)\,p_* = 2^{n\varepsilon'/4+O(\log n)}. \end{aligned}\right\} \quad \text{для } d \in I_1,\ I_2,\ I_4, \tag{8.76}$$

Из отношений (8.76) следует, что при $n \to \infty$

$$\lambda_d \to 0 \quad \text{для } d \in I_1,\ I_2;$$
$$\lambda_d \to \infty \quad \text{для } d \in I_4 \text{ и } \lambda_* \to \infty.$$

Учитывая, что функция распределения Пуассона имеет асимптотические оценки (см. п. 1.6.3)

$$\mathcal{P}(m < K) = 1 - \sum_{m=K}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} > \begin{cases} 1 - \lambda^{-K}, & \text{при } \lambda \to 0 \\ 1 - \exp\left[-\dfrac{K^2}{2\lambda}\right], & \text{при } \lambda \to \infty,\ K/\lambda \to \infty, \end{cases}$$

получим для вероятности P_{1,x_i} одновременного выполнения неравенств (8.73) при фиксированном x_i

$$P_{1,x_i} > \prod_{d \in I_2} (1 - \lambda_d^{-K_d}) \left(1 - \exp\left[-\frac{K_*^2}{2\lambda_*}\right]\right) \prod_{d \in I_4} \left(1 - \exp\left[-\frac{K_d^2}{2\lambda_d}\right]\right).$$

Подставляя в предыдущее соотношение явные выражения для λ_d, λ_*, K_d и K_*, после несложных преобразований получим оценку

$$P_{1,x_i} > 1 - 2^{-2^{n\varepsilon'/2 + O(\log n)}}. \tag{8.77}$$

Далее, для вероятности P_{2x_i} одновременного выполнения неравенств (8.75) при фиксированном x_i аналогично будем иметь

$$P_{2,x_i} > \prod_{d \in I_1} (1 - \lambda_d^{-K_d}) P_{1,x_i},$$

откуда подстановкой явных выражений для λ_{vd} и K получим

$$P_{2,x_i} > 1 - 2^{-n(H+\delta) + O(\log n)}. \tag{8.78}$$

Все предыдущие оценки касались частных группированных распределений $\{M_{x_i}^d\}$ при фиксированном x_i. Эти же оценки (8.77) и (8.78) имеют место и для любого другого x_i. Для оценки совместного выполнения неравенств (8.77) и (8.78), одновременного для всех x_i $(i = \overline{1,M})$, воспользуемся неравенством Буля (см. (2.12))

$$p_{1,2,\ldots,M} \geqslant \sum_{i=1}^{M} p_i - (M-1), \tag{8.79}$$

где $p_{1,2,\ldots,M}$ — вероятность одновременного наступления M зависимых событий $A_1, A_2, \ldots, A_M$; $p_1, p_2, \ldots, p_M$ — вероятности их появления в отдельности.

Если $p_1 = p_2 = \ldots = p_M$, то (8.79), как легко видеть, переходит в неравенство

$$p_{1,2,\ldots,M} \geqslant 1 - M(1 - p_1). \tag{8.80}$$

Полагая в (8.80) для нашего случая $M = 2^{nH}$, $p_i = P_{1,x_i}$, $p_{1,2,\ldots,M} = P_1$, и используя (8.58), получим

$$P_1 > 1 - 2^{-2^{n\varepsilon'/2 + O(\log n)}}.$$

Далее, полагая $M = 2^{nH}$; $p_i = P_{2,x_i}$; $p_{1,2,\ldots,M} = P_2$; используя (8.80) и (8.76), получим: $P_2 > 1 - 2^{-n\delta + O(\log n)}$, что и требовалось доказать.

Из теорем 8.3 и 8.4 можно получить ряд важных следствий, открывающих способ получения P-различимых совокупностей входных слов.

Следствие 8.4.1. Произведем случайный выбор $M = 2^{nH} = 2^{n(1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon))}$ входных слов

$$x_1 x_2 \ldots x_i \ldots x_M, \tag{8.81}$$

где

$$h_{p-\varepsilon}(p+\varepsilon) < f_{p+\varepsilon}[h^{-1}(h(p+\varepsilon)+h_{p-\varepsilon}(p+\varepsilon)-\gamma)] - \varepsilon' < 1 - h(p+\varepsilon).$$

Удалим те из них, которые имеют непустые $f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')$-окрестности. Тогда вероятность $P_{0.1}$ того, что:

1) после этого останется $M' = k_0 > M(1-2^{-n\gamma})$ входных слов

$$x_{i_1} x_{i_2} \ldots x_{i_k} \ldots x_{i_{M'}}; \tag{8.82}$$

2) они окажутся P-различимыми, оценку имеет

$$P_{0.1} > 1-2^{-n[h(p+\varepsilon)+h_{p-\varepsilon}(p+\varepsilon)-h\left(f^{-1}_{p+\varepsilon}\left(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon'\right)\right)-\gamma]+O(\log n)}. \tag{8.83}$$

Доказательство. Между частотами $\{M^d_{x_i}\}$ и $\{M'^d_{x_{i_k}}\}$ входных слов (8.81) и (8.82) имеет место следующее очевидное соотношение:

$$0 \leqslant M'^d_{x_{i_k}} \leqslant M^d_{x_{i_k}},$$

поэтому выполнение для частот $\{M^d_{x_{i_k}}\}$ неравенств (8.73) с вероятностью P_1 влечет за собой выполнение аналогичных неравенств для частот $\{M'^d_{x_{i_k}}\}$ с неменьшей вероятностью $P'_1 \geqslant P_1$.

С другой стороны, удаление из выборки (8.81) входных слов с непустыми $f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')$-окрестностями приводит к тому, что

$$M'^d_{x_{i_k}} \equiv 0, \text{ для } 0 \leqslant d \leqslant [nf^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')]. \tag{8.84}$$

При этом с вероятностью P_0 число M' входных слов (8.82) оказывается не меньше, чем $M(1-2^{-n\gamma})$ (см. теорему 8.3). Но одновременное выполнение для $\{M'^d_{x_{i_k}}\}$ условий (8.84) и (8.73) влечет за собой выполнение для них и условий (8.61), что является достаточным условием P-различимости (8.82). Здесь: $H = 1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon)$; $\eta = f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')$; $\Delta = (p+\varepsilon)(q-\varepsilon)(1-2(p+\varepsilon))^2$ [см. (8.49)].

Вместе с тем вероятность $P_{0.1}$ одновременного выполнения для (8.82) неравенств (8.73) и того, чтобы M' было не меньше, чем $M(1-2^{-n\gamma})$, оценивается неравенством Буля

$$P_{0.1} \geqslant P_0 + P'_1 - 1 \geqslant P_0 + P_1 - 1. \tag{8.85}$$

Используя оценки (8.67) и (8.74) при $\eta = f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')$, из (8.85) получим (8.83), что и требовалось доказать.

Следствие 8.4.2. Если $h^{-1}(0{,}5) \simeq 0{,}111 < p+\varepsilon < 0{,}5$ и $h_{p-\varepsilon}(p+\varepsilon) \leqslant f_{p+\varepsilon}[h^{-1}(2h(p+\varepsilon)-1+2h_{p-\varepsilon}(p+\varepsilon)-\delta)] - \varepsilon'$, то вероятность P_2 получения P-различимой совокупности

$$x_1, x_2, \ldots, x_i, \ldots, x_M \tag{8.86}$$

при $M = 2^{nH} = 2^{n(1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon))}$ входных слов непосредственно, случайным выбором, без последующего удаления входных слов с непустыми $f^{-1}_{p+\varepsilon}(h(p+\varepsilon)+\varepsilon')$-окрестностями, имеет оценку

$$P_2 > 1-2^{-n\delta+O(\log n)}.$$

Доказательство. Из условий следствия имеем $h(f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')) > 2h(p+\varepsilon)-1+2h_{p-\varepsilon}(p+\varepsilon)-\delta$, откуда $1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon)+\delta < h(p+\varepsilon)+h_{p-\varepsilon}(p+\varepsilon)-h(f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon'))$.

Разделив обе части неравенства на его правую часть, получим

$$\frac{1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon)+\delta}{h(p+\varepsilon)+h_{p-\varepsilon}(p+\varepsilon)-h(f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon'))} \leqslant 1. \qquad (8.87)$$

Воспользуемся теперь для анализа частот $\{M^d_{x_i}\}$ выборки (8.86) второй частью теоремы 8.4, полагая $H = 1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon)$; $\eta = f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')$; $\Delta = (p+\varepsilon)(q-\varepsilon)(1-2(p+\varepsilon))^2$.

Тогда для $\{M^d_{x_i}\}$ с вероятностью P_2 выполняются неравенства (8.75) и, в частности, используя (8.87), получим:

$$M^d_{x_i} < K_d = \frac{1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon)+\delta}{h(p+\varepsilon)+h_{p-\varepsilon}(p+\varepsilon)-h\left(\frac{d}{n}\right)} \leqslant$$

$$\leqslant \frac{1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon)+\delta}{h(p+\varepsilon)+h_{p-\varepsilon}(p+\varepsilon)-h\{f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')\}} \leqslant 1$$

для $0 \leqslant d \leqslant [nf^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')]$.

Но так как $M^d_{x_i}$ имеют распределение Пуассона, то утверждение $M^d_{x_i} < 1$ эквивалентно утверждению $M^d_{x_i} \equiv 0$.

Таким образом, в нашем случае вероятность P_2 теоремы 8.4 оценивает соблюдение для частот $\{M^d_{x_i}\}$ неравенств (8.75), в которых первые неравенства для $M^d_{x_i}$ в интервале $0 \leqslant d \leqslant [nf^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')]$ заменены условием $M^d_{x_i} \equiv 0$ на том же интервале.

Легко видеть, что соблюдение для $\{M^d_{x_i}\}$ преобразованных условий (8.75) эквивалентно соблюдению для $\{M^d_{x_i}\}$ условий (8.61). Поэтому выборка (8.66) является P-различимой с вероятностью

$$P_2 > 1-2^{-n\delta+O(\log n)},$$

что и доказывает следствие.

Рассмотрим некоторые частные случаи доказанных следствий. Если необходимо иметь большие по сравнению с вероятностью правильного декодирования вероятности получения P-различимых входных слов случайным выбором, то, полагая $\delta = 2\gamma = 2h_{p-\varepsilon}(p+\varepsilon)$, будем иметь при

$$h_{p-\varepsilon}(p+\varepsilon) < f_{p+\varepsilon}(p+\varepsilon)-\varepsilon'$$

и

$$h_{p-\varepsilon}(p+\varepsilon) \leqslant f_{p+\varepsilon}(h^{-1}(2h(p+\varepsilon)-1)),$$

соответственно

$$P_{0,1} > 1-2^{-n\left(h(p+\varepsilon)-h\left(f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')\right)\right)+O(\log n)} \tag{8.88}$$

и

$$P_2 > 1-2^{-n2h_{p-\varepsilon}(p+\varepsilon)+O(\log n)}. \tag{8.89}$$

В заключение рассмотрим вырожденный случай малых H, когда $0 \leqslant H = 1-h(f^{-1}(h(p+\varepsilon)+\varepsilon')) - \beta < h_2(p+\varepsilon)(q-\varepsilon)(0,5)$, чему соответствуют большие значения

$$h_{p+\varepsilon}(0,5) \leqslant h_{p-\varepsilon}(p+\varepsilon) \leqslant \begin{cases} f_{p+\varepsilon}(2(p+\varepsilon)) & (0,18 \leqslant p+\varepsilon < 0,25)^*, \\ f_{p+\varepsilon}(0,5) & (0,25 \leqslant p+\varepsilon < 0,5). \end{cases}$$

В этом случае можно показать, что

$$\left.\begin{aligned} P_{0,1} &> 1-2^{-n(\beta-\gamma)} \\ \text{и} \\ P_2 &> 1-2^{-n(\beta-H)}. \end{aligned}\right\}$$

Если положить $\delta = 2\gamma = 2(h_{p-\varepsilon}(p+\varepsilon) - h_{p+\varepsilon}(0,5))$, то получим

$$P_{0,1} > 1-2^{-n\left(1-h\left(f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')\right)\right)+O(\log n)}$$

и

$$P_2 > 1-2^{-n2(h_{p-\varepsilon}(p+\varepsilon)-h_{p+\varepsilon}(0,5))+O(\log n)}.$$

§ 8.7. Основная теорема и ее практическое использование

Окончательный результат всех предыдущих рассмотрений формулирует следующая основная теорема о конструкции оптимального кода в случае бинарного симметричного канала.

Основная теорема [6].

Для любого $p(0<p<0,5)$ и $\varepsilon(h_{p-\varepsilon}(p-\varepsilon) < f_{p+\varepsilon}[h^{-1}(h(p+\varepsilon)+ + h_{p-\varepsilon}(p+\varepsilon)-\gamma)]-\varepsilon')$ случайный выбор $M = 2^{n(1-h(p+\varepsilon)-h_{p-\varepsilon}(p+\varepsilon))}$ выходных слов длины n

$$x_1 x_2 \ldots x_i \ldots x_M \tag{8.90}$$

с последующим удалением входных слов с непустыми $f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')$-окрестностями обеспечивает с вероятностью

$$P_{0,1} > 1-2^{-n\left(h(p+\varepsilon)+h_{p-\varepsilon}(p+\varepsilon)-h\left(f^{-1}_{p+\varepsilon}(h_{p-\varepsilon'}(p+\varepsilon)+\varepsilon')\right)-\gamma\right)+O(\log n)} \tag{8.91}$$

получение $M' \geqslant M(1-2^{-n\gamma})$ входных слов

$$x_{i_1} x_{i_2} \ldots x_{i_k} \ldots x_{i_{M'}},$$

таких, что при соотнесении каждому из них соответственно совокупности выходных слов $\mathcal{S}^{0,[n(p+\varepsilon)]}_{x_{i_k}}$ $(k = \overline{1,M'})$ они становятся P-различимыми.

Вероятность P правильного декодирования

$$P > 1-2^{-n\frac{\varepsilon^2}{3pq\ln 2}+O(\log n)}. \tag{8.92}$$

* Здесь $\theta = 0,18$ является решением уравнения $h_\theta(0,5) = f_\theta(2\theta)$.

В случае, когда $h^{-1}(0{,}5) \simeq 0{,}111 < p + \varepsilon < 0{,}5$ и $h_{p-\varepsilon}(p+\varepsilon) < f_{p+\varepsilon}[h^{-1}(2h(p+\varepsilon) - 1 + 2h_{p-\varepsilon}(p+\varepsilon) - \delta)] - \varepsilon'$, можно не проводить удаления из (8.90) входных слов с непустыми $f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')$-окрестностями.

При этом получение непосредственно случайной выборки (8.90) в качестве P-различимой совокупности входных слов с той же вероятностью правильного декодирования (8.92) имеет вероятность

$$P_2 > 1 - 2^{-n\delta + O(\log n)},$$

где $h(\theta) = -\theta\log\theta - (1-\theta)\log(1-\theta)$ и $f_p(\theta) = h(\theta) - ph\left(\frac{\theta}{2p}\right) - (1-p)h\left(\frac{\theta}{2(1-p)}\right)$, а $h^{-1}(\ )$ и $f_p^{-1}(\ )$ означают обратные им функции.

Заметим, что при $\gamma = {}^1\!/_2\delta = h_{p-\varepsilon}(p+\varepsilon)$ [см. (8.88), (8.89)] во всех случаях, рассмотренных в теореме, вероятность получения оптимального кода существенно больше, чем вероятность правильного декодирования. Поэтому требование большой вероятности правильного декодирования при нашей конструкции автоматически приводит к большой вероятности получения оптимального кода.

Вторая часть теоремы указывает на то, что лишь для случая сравнительно больших вероятностей $p\,(0{,}1 < p < 0{,}5)$ можно строить оптимальный код, используя чисто случайную выборку без последующего выбрасывания из нее n-цепочек с непустыми $f^{-1}_{p+\varepsilon}(h_{p-\varepsilon}(p+\varepsilon)+\varepsilon')$-окрестностями.

Заметим, что условия, ограничивающие выбор ε, в теореме таковы, что при $p \to 0{,}5$ и $p \to 0$ ε стремится к нулю, вследствие чего для обеспечения большей вероятности правильного декодирования согласно (8.92) требуется брать большие длины входных слов. Последнее обстоятельство делает такого рода коды трудно осуществимыми.

Если при $p \to 0{,}5$ это обстоятельство является оправданным, так как это стремление указывает на тяжелый режим работы канала с шумами, то при $p \to 0$ это обстоятельство целиком связано с несовершенством конструкции описанного кода, использующего случайный выбор входных слов. Поэтому при $p \to 0$, когда $p \ll 0{,}1$ (и можно считать, что $p = \lambda/n$, где $\lambda < 1$), целесообразно идти по пути, отвергнутому в § 8.3, и пользоваться в качестве множеств выходных слов пуассоновскими множествами $\mathscr{C}^{0,[n\varepsilon]}_{x_i}$. Они, как было указано в § 8.3, не пересекаются, если x_i таковы, что $d(x_i, x_j) \geqslant 2[n\varepsilon]$ $(i \neq j)$.

Получение таких заведомо различимых x_i в количестве $M' = M(1-2^{-n\gamma})$ штук осуществляется, согласно теореме 8.3 [при $H + \gamma < 1 - h(2\varepsilon)$ и $\eta = 2\varepsilon$] случайным выбором $M = 2^{nH}$ входных слов, с последующим удалением из них входных слов с непустыми 2ε-окрестностями, с вероятностью

$$P_0 > 1 - 2^{-n(1-h(2\varepsilon)-H-\gamma) + O(\log n)}.$$

Вероятность правильного декодирования в этом случае, согласно (8.43), имеет оценку

$$P > 1 - 2^{-\varepsilon n\log n + O(n)},$$

причем ε может иметь любое значение между e/n и 0,25.

Следует также указать на то, что можно не проводить выбрасывания непустых η-окрестностей при любых $0 < p < 0{,}5$.

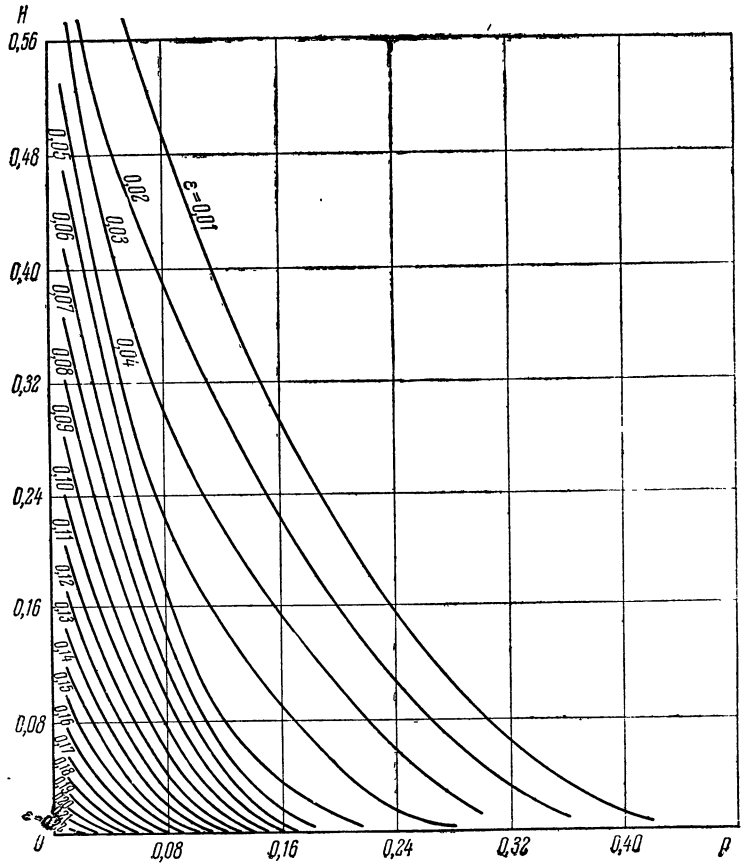


Рис. 8.3. Зависимости между параметрами p, H, ε

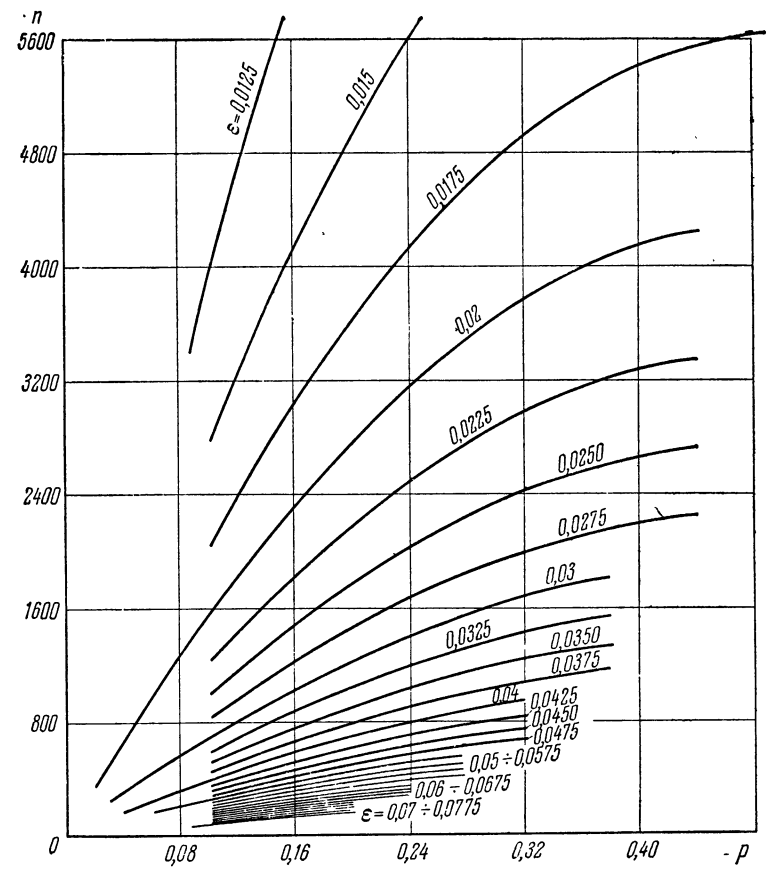


Рис. 8.4. Зависимости между параметрами p, n, ε ($P=0{,}9$)

В самом деле, так как этих окрестностей согласно теореме с вероятностью (8.91) меньше, чем $M2^{-n\gamma}$, то, кодируя ими, мы придем к вероятности меньше, чем вероятность правильного декодирования (8.92) лишь для $2^{-n\gamma}$ доли кодируемых высоковероятных сообщений источника, что составляет при больших n и при $\gamma \approx \varepsilon$, вполне допустимую их долю.

§ 8.8. Расчетные кривые и численные примеры

На основании соотношений, приведенных в основной теореме, могут быть построены графики, облегчающие получение численных связей между основными параметрами p, n, $M = 2^{nH}$ и P и вспомогательным параметром ε, используемым для построения множеств выходных слов P-представляющих соответствующие входные слова.

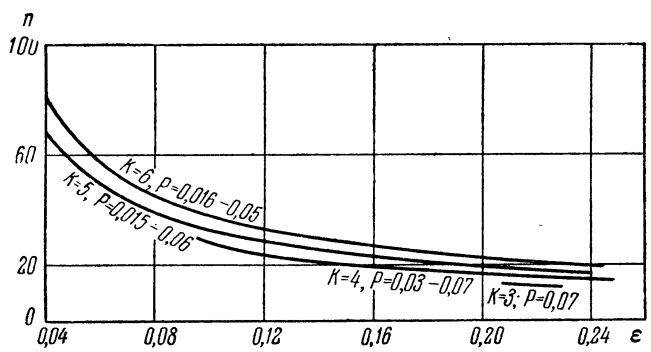


Рис. 8.5. Зависимости между параметрами p, n, ε и $P = 1 - 10^{-k}$ (для малых $p \ll 0{,}1$)

На рис. 8.3 приводятся графики зависимости между параметрами p, H и ε; на рис. 8.4 — графики зависимости между параметрами p, n и ε при $P = P_1 = 1 - 10^{-1} = 0{,}9$; на рис. 8.5 для случая малых $p = \lambda/n \to 0$ приводятся графики зависимости между p, n и ε при $P = P_k = 1 - 10^{-k}$.

Рассмотрим примеры использования указанных графиков.

Пример 8.1. Пусть уровень шумов в бинарном симметричном канале определяется параметром $p = 0{,}28$. Необходимо построить оптимальный код для источника с энтропией $H = 0{,}06$ так, чтобы вероятность правильного декодирования была равна $P = P_2 = 1 - 10^{-2} = 0{,}99$ ($k = 2$). Кроме того, необходимо определить длину n входных слов и параметр ε для конструкции множеств выходных слов $\mathscr{E}_{x_i}^{0,\,[n(p+\varepsilon)]}$.

Решение. По семейству кривых рис. 8.3 для заданных $H = 0{,}06$ и $p = 0{,}28$ определяется $\varepsilon = 0{,}02$, а затем для $\varepsilon = 0{,}02$ и $p = 0{,}28$ по семейству кривых рис. 8.4 находится $n = n_1 = 3\,500$, приводящее к вероятности $P = P_1 = 1 - 10^{-1} = 0{,}9$. Чтобы получить $n = n_k$, соответствующее $P = P_k = 1 - 10^{k}$, нужно умножить значение $n = n_1$ на k ($n_k = kn_1$). В рассматриваемом случае $k = 2$ и отсюда $n = n_2 = 2n_1 = 2 \cdot 3500 = 7000$.

Пример 8.2. Рассмотрим случай малого $p = 0{,}02$ и $H = 0{,}36$. Требуется определить параметр ε и значение n, необходимое для достижения вероятности правильного декодирования $P = P_5 = 1 - 10^{-5}$ ($k = 5$).

Решение. На рис. 8.3 по $p = 0{,}02$ и $H = 0{,}36$ находим величину $\varepsilon = 0{,}06$, а затем на рис. 8.5 по $\varepsilon = 0{,}06$ и $k = 5$ находим $n = 50$.

Глава 9

ОЦЕНКА ТЕХНИЧЕСКИХ ВОЗМОЖНОСТЕЙ РЕАЛИЗАЦИИ ОПТИМАЛЬНОГО КОДИРОВАНИЯ

§ 9.1. Вводные замечания

Основное преимущество оптимальных методов кодирования перед другими его видами заключается в том, что они обеспечивают за счет передачи длинных сигналов в целом надежную передачу предельно большого их числа M_c. При этом, если число высоковероятных сообщений источника $M<M_c$, то задержка в передаче будет связана только с ожиданием прихода первого длинного сигнала (в случае быстрого декодирования) и приводит лишь к сдвигу приема во времени на величину длительности первого сигнала без накопления такой задержки во времени. В то же время другие виды кодирования (например, посимвольное) приводят к накоплению задержки во времени при аналогичной передаче.

Следует иметь в виду, что оптимальное кодирование требует большие длины входных слов и огромное (экспоненциальное) их число $M=e^{nH}$, что в свою очередь предъявляет высокие требования к быстродействию и объему памяти кодирующей и декодирующей аппаратуры. Связанное с этим усложнение аппаратуры ставит остро вопрос о ее надежности. Надо всегда иметь в виду, что недостаточность быстродействия, объема памяти и надежности соответствующей аппаратуры может свести на нет указанные выше преимущества оптимального кодирования.

Найденные свойства оптимальных кодов отражают универсальные предельные закономерности. Вопрос же об их технической реализации, связанный с экономическим фактором, является преходящим, учитывая бурный прогресс современной техники. Вместе с тем уже сейчас целесообразно рассмотрение этого вопроса с учетом возможностей современного уровня техники. Этому вопросу посвящен обстоятельный доклад [43], касающийся в основном случая низкого уровня шумов, а также неоптимальных корректирующих кодов.

В этой главе будут рассмотрены принципиальные технические возможности осуществления оптимального кодирования с учетом быстродействия и объема памяти соответствующей аппаратуры. Общая постановка и решение такого рода задач с учетом надежности аппаратуры приведены в дополнении II.

§ 9.2. Статистика источника и первое заполнение

Рассмотрим вопрос об особенностях статистики длинных сообщений источника, влияющих на характер аппаратуры на входе канала. В самом деле, согласно [2], широкий класс источников вырабатывает практически лишь высоковероятную группу равновероятных сообще-

ний в числе $M=e^{nH}$, где n — число символов дискретных (или подвергнутых дискретизации) сообщений, а H — константа, связанная с вероятностной природой источника, называемая в теории информации его энтропией ($0\leqslant H\leqslant \ln a$, где a — число символов источника).

Таким образом, все высоковероятные сообщения имеют одну и ту же вероятность $\bar{e}^{nH}$. Естественно, что для изучения вероятностных свойств частот сообщений в s моментах времени можно использовать распределение Мизеса $P_{M,s}(\bar{k})$ [см. (1.32)], принимая M за число предметов.

Интересно выяснить, как скоро появятся все длинные сообщения источника в предположении их независимости (чем сообщения длиннее, тем оправданнее такое предположение в случае конечных во времени связях между их символами).

Другими словами, вычислим распределение первого момента s появления последнего из M высоковероятных сообщений источника.

Для этого рассмотрим в $s-1$ момент одномерное распределение Мизеса для координаты k_0, означающей в нашем случае число непоявившихся высоковероятных сообщений

$$P_{M,\,s-1}(k_0)\approx \frac{\Lambda_{M,\,s-1}^{k_0}}{k_0!}\,e^{-\Lambda_{M,\,s-1}},$$

где $\Lambda_{M,\,s-1}=Me^{-s/M}$ [см. (1.87)].

Легко видеть, что искомая вероятность будет равна

$$P_M(s)=\frac{1}{M}P_{M,\,s-1}(1)\approx\frac{1}{M}\,Me^{-s/M}e^{-Me^{-s/M}}. \tag{9.1}$$

В нашем случае удобнее перейти к функции распределения, для чего нужно проинтегрировать $P_M(s)$ по s, считая последнюю непрерывной величиной. Такое интегрирование выражения (9.1) по s в пределах от M до y приводит к вероятности

$$\mathcal{P}(s<y)=F(y)\approx e^{-Me^{-y/M}}. \tag{9.2}$$

Отсюда p-квантиль полученного распределения, которое можно назвать распределением первого заполнения, будет иметь вид

$$y_p\approx M\ln M-M\ln\ln 1/p=e^{nH}[nH-\ln\ln 1/p]. \tag{9.3}$$

Из соотношения (9.3) следует, что медиана распределения первого заполнения имеет вид

$$y_{0,5}\approx M\ln M-M\ln\ln 2=e^{nH}[nH-\ln\ln 2].$$

Соотношения (9.2) и (9.3) позволяют определить время (пропорциональное s), при котором с большой вероятностью $p\approx 1$ произойдет появление всех высоковероятных сообщений источника. Из (9.3) следует, что при $p\to 1$

$$y_p\approx M\ln M+M\ln\frac{1}{1-p}=M\ln\frac{M}{1-p}.$$

Если исходить из того, что на входе канала в долговременной памяти хранятся все образцы высоковероятных сообщений источника, то приведенные оценки позволят судить о распределении времени первого использовавсех ячеек указанной памяти.

§ 9.3. Варианты устройств кодирования

Простейшее устройство кодирования мыслится как жесткий коммутатор (таблица), коммутирующий входные ячейки памяти, различающие сообщения с входными словами оптимального кода (рис. 9.1). В этом случае объем памяти на входе должен составлять, $ne^{nH} \approx e^{nH}$ ячеек.

При этом совершенно не важно, по какому закону образованы входные слова $x_i\,(i = \overline{1,M})$, так как в любом случае их надо (в каком-то устройстве) соотносить с инородными им сообщениями источника. Это обстоятельство

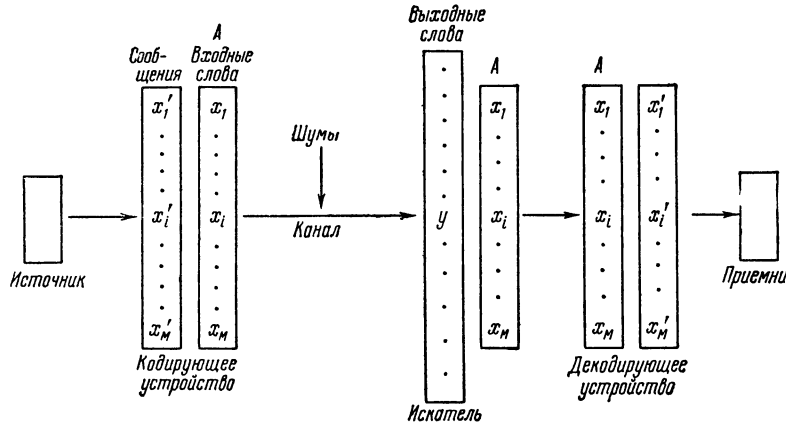


Рис. 9.1. Принципиальная схема реализации кодирования и декодирования

часто не учитывают, противопоставляя регулярные способы получения входных слов $F(i) = x_i$ $(i = \overline{1, M})$ случайным способам (см. ниже).

Рассмотрим другие варианты осуществления кодирования на входе канала. Пусть задано некоторое посимвольное преобразование $f_\tau(x') = x$ сообщений источника $x_i'\,(i = \overline{1, M})$ во входные слова $x_i\,(i = \overline{1,M})$, зависящее и от символов сообщения $x' = (\alpha_1', \ldots, \alpha_n')$, стоящих на местах $\tau = (i_1, \ldots, i_t)$ $(0 \leqslant t \leqslant n)$.

Ясно, что для этого преобразования объем таблицы, который необходимо хранить в памяти, имеет порядок $e^{Ht}\,(0 \leqslant t \leqslant n)$. Кроме того, можно показать, что чем сложнее преобразование $f_\tau(x') = x$, тем «качественнее» оказываются входные слова x_i $(i = \overline{1,M})$.

После сделанных замечаний рассмотрим два крайних случая преобразования $f_\tau(x') = x$. Пусть $t = 0$, тогда $\tau = \oslash$ и преобразование оказывается предельно простым, не требует памяти на входе, и является одинаковым для всех сообщений источника. Из-за этого не происходит улучшение качества входных слов по сравнению с сообщениями источника.

В случае $t = n$ мы возвращаемся к первоначальному простейшему способу качественного кодирования с объемом памяти порядка e^{nH}. Компромиссным решением может оказаться выбор некоторого промежуточного значения $t\,(0 \leqslant t \leqslant n)$, при котором небольшой объем памяти сочетался бы с достаточно хорошим качеством входных слов.

Опишем теперь наиболее перспективный с точки зрения необходимого объема памяти вариант устройства кодирования. Пусть задано постоянное, не посимвольное преобразование сообщений в целом $g(x') = x$, определенное на местах их символов. Например, это преобразование может пред-

ставлять собой последовательность k циклических сдвигов x' на определенные участки в λ_r символов ($r = \overline{1,k}$) (при этом последний символ следует за первым) и посимвольное сложение всех образовавшихся таким образом k последовательностей длины n по модулю a (см. дополнение I).

Ясно, что в случае указанного преобразования не требуется экспоненциального объема памяти на входе канала и качество входных слов будет тем лучше, чем сложнее указанное преобразование. Для разобранного примера $g(x') = x$ в дополнении I показано, что с ростом k (для однозначного декодирования k не должно превосходить n) входные слова будут все лучше имитировать случайные бернуллиевские последовательности.

Однако второй вариант преобразования $g(x') = x$ уступает первому варианту $f_\tau(x') = x$ по времени осуществления. В самом деле, первый вариант может не содержать арифметических операций, в то время как второй, по-видимому, без них не может обойтись. Но поскольку при современном быстродействующем машинном осуществлении операций кодирования основное время тратится именно на арифметические операции, то в указанном смысле преобразование $g(x') = x$ хуже преобразования $f_\tau(x') = x$. Однако основным техническим препятствием осуществления оптимального кодирования является экспоненциальный рост объема памяти [43], и поэтому окончательно преобразование $g(x') = x$ предпочтительнее преобразования $f_\tau(x') = x$.

§ 9.4. Варианты устройств декодирования

Искаженное шумами выходное слово y поступает в решающее устройство типа искателя (см. рис. 9.1), где происходит идентификация y с одним из M входных слов x_i ($i = \overline{1,M}$). Вариант с M устройствами идентификации, связанными с каждым из x_i ($i = \overline{1,M}$), позволяющий проводить одновременную идентификацию сразу по всем x_i ($i = \overline{1,M}$), был отвергнут в § 5.4 из-за необходимости иметь экспоненциальное число идентичных приборов, что, по-видимому, технически не оправдано и в настоящее время не осуществимо. Вместе с тем отказ от этого требует последовательного во времени сравнения y с экспоненциальным числом x_i ($i = \overline{1,M}$) и наличия общего экспоненциального времени. В связи с этим необходимо большое быстродействие искателя.

В главах 5 и 6 были рассмотрены два варианта декодирования, основанные на статистических и комбинаторных соображениях. Технические преимущества первого варианта состоят в том, что аппаратура решающего устройства должна состоять из сравнительно несложного быстродействующего однопорогового (в классическом случае) или двухпорогового (в последовательном случае) анализатора [34]. При втором варианте требуется аппаратура, которая должна хранить в памяти структуру множеств $\mathscr{E}^i$ выходных слов, соответствующих входным словам x_i ($i = \overline{1,M}$), что приводит к необходимости большого объема памяти на выходе. Кроме того, для установления принадлежности выходного слова y к множеству $\mathscr{E}_i$ ($i = \overline{1,M}$) необходим перебор экспоненциального числа элементов последнего, что затягивает уже элементарный акт идентификации (необходимо произвести M таких актов при декодировании одного выходного слова y).

Рассмотрим третий вариант декодирования, основанный на максимуме функции правдоподобия. Этот вариант, использованный в работах [10, 35, 36], сводится к принятию решения о том, что передавалось входное слово x_i, обращающее в максимум выражение условной вероятности (функции

правдоподобия)

$$p_{x_i}(y) = \max_{x_j (j=\overline{1,M})} p_{x_j}(y),$$

где y — принятое выходное слово [если максимум $p_x(y)$ достигается при нескольких x_i, то принимается решение о передаче произвольного из них].

В простейшем случае бинарного симметричного канала [36] отыскание x_i, обращающего в максимум $p_x(y)$, эквивалентно отысканию x_i, обращающего в минимум хэмминговское расстояние

$$d(x_i, y) = \min_{x_j (j=\overline{1,M})} d(x_j, y).$$

Таким образом, в этом случае принимается решение о передаче x_i, имеющего минимальное расстояние $d(x_i, y)$ с принятым выходным словом y (если таких x_i несколько, то принимается решение о передаче произвольного из них).

Как указано в работе [41], третий вариант декодирования требует для своего осуществления в среднем в два раза больше времени, чем первые два варианта. В самом деле, окончание процедуры декодирования в первых двух вариантах при поочередном переборе x_i ($i = 1,2,\ldots, M$) связано с первым положительным статистическим решением в первом варианте или первым попаданием y в множество $\mathcal{E}_i$ ($i = 1,2,\ldots, M$) во втором варианте. Учитывая большую вероятность P правильного декодирования и равномерное распределение передаваемых высоковероятных сообщений источника, легко показать, что в среднем решения в первых двух вариантах выносятся после половины ($M/2$) переборов вариантов x_i ($i = 1,2,\ldots, M$). В третьем же варианте схемы декодирования необходим перебор всех M вариантов для отыскания варианта, приводящего к

$$\min_{x_i (i=\overline{1,M})} d(x_i, y).$$

§ 5.4 упоминалось о выигрыше в темпе передачи при использовании последовательного декодирования вместо классического. Однако для осуществления последовательного декодирования требуется наличие надежного канала обратной связи. Поэтому окончательное решение о целесообразности такого декодирования с технической точки зрения не является очевидным.

Рассмотрим вопрос о том, можно ли избежать экспоненциального объема памяти на выходе канала при декодировании, так же как это имело место в случае использования преобразования $g(x') = x$ при кодировании на входе канала.

Необходимость перебора всех вариантов x_i ($i = \overline{1,M}$) для идентификации их с принятым словом y, на первый взгляд, не создает каких-либо возможностей избежать экспоненциального объема памяти на выходе канала. Однако на самом деле это не так.

Допустим, что существует некоторое регулярное правило $G(i) = x_i'(i = \overline{1,M})$, однозначно вырабатывающее для каждого значения i соответствующее сообщение источника x_i'. Другими словами, пусть существует регулярный способ построения сообщений источника. Эти сообщения с достаточно большой скоростью воспроизводятся по указанному правилу $G(i) = x_i'$, затем преобразуются, согласно правилу $g(x_i') = g[G(i)] = x_i$, во входные слова и последовательно идентифицируются с принятым выходным словом y.

После решения о том, что передавалось x_i обратным преобразованием $x'_i = g^{-1}(x_i)$ на выходе канала, восстанавливается входное сообщение.

Как видно, описанная выше процедура не требует экспоненциального объема памяти на выходе канала и предъявляет лишь повышенные требования к скорости осуществления воспроизведения сообщений x_i источника по регулярному закону $G(i) = x'_i$.

Интересно отметить противоположность ситуации на входе и выходе канала. В самом деле, наличие регулярного правила образования входных слов $F(i) = x_i$ $(i = \overline{1,M})$ не избавляет от необходимости иметь экспоненциальный объем памяти на входе канала. Избавляет от этого лишь регулярное преобразование $g(x'_i) = x_i$ сообщений источника в псевдослучайные входные слова. Наоборот, на выходе канала от экспоненциального объема памяти избавляет наличие регулярного правила образования сообщений источника $G(i) = x'_i$ $(i = \overline{1,M})$.

Таким образом, окончательно избавиться от экспоненциального объема памяти как на входе, так и на выходе можно только при наличии описанных однозначного преобразования $g(x'_i) = x_i$ и регулярного правила $G(i) = x'_i$, что формально позволяет получать x_i с помощью регулярного правила $F(i) = g[G(i)] = x_i$ $(i = \overline{1,M})$.

Как отмечалось выше, получение преобразования $g(x'_i) = x_i$ связано с известными широко применяемыми в настоящее время регулярными методами получения случайных чисел.

Итак, решение всей проблемы технического осуществления оптимального кодирования сводится к построению регулярного правила $G(i) = x'_i$ $(i = \overline{1,M})$ воспроизведения сообщений источника. Насколько нам известно, такого правила в настоящее время не существует. Вместе с тем при составлении искусственных языков типа «алгол» важно было бы учесть необходимость образования последних по регулярному правилу $G(i) = x'_i$ для целей оптимального кодирования.

Практически вполне удовлетворительным было бы решение несколько ослабленной проблемы. Достаточно было бы получить регулярное правило $\widetilde{G}(i) = \widetilde{x}'_i$ $(i = \overline{1, M})$ воспроизведения некоторой последовательности $\widetilde{x}'_i$ $(i = \overline{1, M})$, которая достаточно хорошо воспроизводит совокупность x'_i $(i = \overline{1, M})$ сообщений источника. Действительно, не обязательно требовать соответствия друг другу $\widetilde{x}'_i = x'_i$ при одних и тех же i, они могут соответствовать друг другу при разных индексах, например в случае $x'_i = x'_j$ $(i \neq j)$. Далее, из-за асимптотического с ростом n характера всех рассмотрений допустимо, чтобы соответствующие друг другу $\widetilde{x}'_i$ и x'_j разнились в некотором числе t символов, имеющем порядок $o(n)$. Кроме того, можно было бы допустить, чтобы последовательности $\widetilde{x}'_i$ и $\widetilde{x}_i$ $(i = \overline{1, M})$ существенно разнились в числе L членов, удовлетворяющем условию

$$L/M < 1 - P, \tag{9.4}$$

где P — вероятность правильного декодирования.

В самом деле, вследствие равновероятности высоковероятных сообщений источника, слева в соотношении (9.4) находится вероятность неправильной идентификации y из-за неточного воспроизведения входного слова. Поэтому общая вероятность ошибки декодирования в рассматриваемом случае не должна превосходить величины $1-2(1-P)$, что в силу экспоненциаль-

ного характера $(1-P)$ приводит к сохранению прежней вероятности правильного декодирования $1-2(1-P) \approx 1-(1-P)=P$.

Таким образом, если положить $L=e^{nH'}$, $M=e^{nH}$ и $P=1-e^{-\varepsilon^2 n}$, то допустимый коэффициент H' в экспоненте L должен, согласно (9.4), удовлетворять неравенству $H' \leqslant H-\varepsilon^2$.

Итак, без построения последовательности $\widetilde{G}(i)=\widetilde{x}'_i$, хорошо воспроизводящей последовательность $G(i)=x'_i$, проблема технической реализации оптимального кодирования остается нерешенной и требует дальнейших исследований.

В настоящей главе не были рассмотрены вопросы надежности аппаратуры, осуществляющей оптимальное кодирование и декодирование, которые могут быть решающими для всей проблемы.

Учет наряду с быстродействием и объемом памяти еще и надежности аппаратуры, осуществляющей статистические процедуры более общего типа, чем рассмотренные выше, содержится в дополнении II.

Дополнение I

СТОХАСТИЧЕСКИЕ СУММЫ

§ I.1. Определение и постановка задачи

Теория сумм независимых случайных слагаемых при условии, когда значения этих слагаемых берутся из коммутативной группы, рассмотрена в ряде работ [44]. Ниже будет рассмотрен случай конечной коммутативной группы, при этом кроме определения не будут использованы никакие факты из их теории.

Конечная коммутативная группа порядка a $G_a = (0, 1, \ldots, \alpha, \ldots, a-1)$ состоит из a элементов. Она полностью определяется $(a \times a)$-таблицей сложения (таблица Келли), в которой паре элементов $(\alpha, \alpha') \in G_a$ сопоставляется элемент $\beta \in G_a$, что символически записывается как $\alpha + \alpha' = \beta$.

В силу коммутативности группы $G_a(\alpha + \alpha' = \alpha' + \alpha = \beta)$ $(a \times a)$-таблица сложения симметрична относительно главной диагонали. Элемент $\alpha' = \beta - \alpha$ определяется из условия, что он в сумме с элементом α дает элемент β. Простейшим примером конечной коммутативной группы G_a порядка a являются вычеты по модулю a, определяемые неотрицательными целыми числами $0, 1, \ldots, \alpha, \ldots, a-1$. Их сумма $\alpha + \alpha' = \beta \pmod a$ определяется обычным сложением по модулю a.

Близкие по смыслу стохастическим функциям стохастические суммы дискретных случайных слагаемых рассматриваемого типа (не обязательна независимость слагаемых, значения последних берутся из конечной коммутативной группы) являются обобщением классического случая сумм независимых случайных слагаемых.

Обобщение состоит в том, что ищется не выражение вероятности *отдельного* значения суммы по вероятности *отдельных* значений слагаемых (композиция законов распределения), а выражение вероятности *совокупности* частот значений суммы в s моментов времени по аналогичным вероятностям для не обязательно независимых слагаемых. При изучении стохастических сумм оказывается важным использование понятия равномерно расположенных случайных переменных, эквивалентное понятию симметричного СА(§ 2.4).

С помощью комбинаторных приемов можно найти связь между распределениями частот значений стохастической суммы и аналогичными распределениями случайных слагаемых в s моментов времени, а также простую связь между моментами суммы и слагаемых. Помимо этого, можно установить частный случай, когда стохастические функции и сумма эквивалентны друг другу с точки зрения теории вероятностей, так как имеют одно и то же распределение. В общем же случае ни одна из этих схем не может быть рассмотрена как частный случай другой.

Приступим к определению стохастической суммы H случайных переменных $\Xi^{(r)}$ $(r = \overline{1, k})$. Для этого соотнесем моментам времени $\sigma = (1, .., t, \ldots, s)$ соответствующие случайные переменные $\xi_t^{(r)}$ $(r = \overline{1, k})$ по следующей

схеме:

$$\sigma = (1, \ldots, t, \ldots, s) \qquad (I.1)$$
$$\begin{matrix} \xi_1^{(1)} \ldots & \xi_t^{(1)} \ldots & \xi_s^{(1)} \\ \vdots & \vdots & \vdots \\ \xi_1^{(r)} \ldots & \xi_t^{(r)} \ldots & \xi_s^{(r)} \\ \vdots & \vdots & \vdots \\ \xi_1^{(k)} \ldots & \xi_t^{(k)} \ldots & \xi_s^{(k)} \end{matrix}$$

Каждое из случайных переменных $\xi_t^{(r)}$ принимает конечное число значений $\xi_t^{(r)} = 0, 1, 2, \ldots, \alpha_t^{(r)}, \ldots, a-1$ элементов конечной, порядка a коммутативной группы $G_a = \{0, 1, \ldots, \alpha, \ldots, a-1\}$.

Определим в каждый момент времени t случайную переменную η_t, принимающую значения из той же группы $\eta_t = 0, 1, 2, \ldots, \beta_t, \ldots, a-1 \in G_a$, причем $\eta_t = \beta_t$, если в тот же момент времени t

$$\xi_t^{(1)} = \alpha_t^{(1)}, \ldots, \xi_t^{(r)} = \alpha_t^{(r)}, \ldots, \xi_t^{(k)} = \alpha_t^{(k)} \quad \text{и} \quad \sum_{r=1}^{k} \alpha_t^{(r)} = \beta_t.$$

Если рассмотреть отрезок в s моментов времени $\sigma = (1, \ldots, t, \ldots, s)$, то на нем каждая из последовательностей $\xi_t^{(r)}$ $(t = \overline{1, s})$ примет s значений, частоты которых соответственно $m^{(r)} = (m_0^{(r)}, m_1^{(r)}, \ldots, m_{a-1}^{(r)})$ можно рассматривать как значения a-мерных случайных величин $\Xi^{(r)}$ $(r = \overline{1, k})$.

В эти же самые моменты времени последовательность η_t будет принимать s значений, частоты которых $\overline{m} = (m_0, \ldots, m_{a-1})$ можно рассматривать как значения a-мерной случайной величины H. Последнюю будем называть стохастической суммой a-мерных случайных слагаемых $\Xi^{(r)}$ $(r = \overline{1, k})$ и обозначать

$$\mathrm{H} = \Xi^{(1)} \dot{+} \ldots \dot{+} \Xi^{(r)} \dot{+} \ldots \dot{+} \Xi^{(t)} = \sum_{r=1}^{k} \cdot \Xi^{(r)}.$$

Основная задача состоит в определении распределения $P^*_{(\overline{m})} = \mathcal{P}(\mathrm{H} = \overline{m})$ стохастической суммы H по совместному распределению слагаемых $P(\overline{m}^{(1)}, \ldots, \overline{m}^{(k)}) = \mathcal{P}(\Xi^{(1)} = \overline{m}^{(1)}, \ldots, \Xi^{(k)} = \overline{m}^{(k)})$. Способ, с помощью которого по $P(\overline{m}^{(1)}, \ldots, \overline{m}^{(r)}, \ldots, \overline{m}^{(k)})$ получается $P^*(\overline{m})$, мы будем называть многомерной композицией распределения $P(\overline{m}^{(1)}, \ldots, \overline{m}^{(r)}, \ldots, \overline{m}^{(k)})$.

Введем следующее определение, эквивалентное определению симметричного СА (гл. 2).

Будем называть случайные переменные ξ_t, соотнесенные дискретным моментам времени $\sigma = (1, \ldots, t, \ldots, s)$, *равномерно расположенными*, если вероятность появления значений $\xi_t = 0, 1, 2, \ldots, a-1$ в эти моменты, соответственно $p(x) = p(\alpha_1, \ldots, \alpha_t, \ldots, \alpha_s) = \mathcal{P}(\xi_1 = \alpha_1 \ldots \xi_t = \alpha_t \ldots, \xi_s = \alpha_s)$, зависит только от частот этих значений $m = (m_0, m_1, \ldots, m_{a-1})$ и не зависит от расположения конкретных значений в конкретные моменты времени, т. е. $p(\alpha_1, \ldots, \alpha_t, \ldots, \alpha_s) = f(\overline{m})$, где $x = (\alpha_1, \ldots, \alpha_t, \ldots, \alpha_s) \in \mathcal{E}_\sigma^{\overline{m}}$.

Отметим, что число возможных расположений значений $0, 1, 2, \ldots, a-1$ в σ моментов времени, при фиксированном значении их частот $\overline{m} = (m_0, m_1, \ldots, m_{a-1})$, равно $C_s^{m_1,\ldots,m_\alpha,\ldots,m_{a-1}} = C_s^{\overline{m}} = \dfrac{s!}{m_0!\ldots m_\alpha!\ldots m_{a-1}!}$. При этом, если вероятность появления частот $\overline{m} = (m_0, \ldots, m_\alpha, \ldots, m_{a-1})$ равна $P(\overline{m})$, то вероятность появления любого расположения значений $0, 1, \ldots, \alpha, \ldots, a-1$ с фиксированными частотами $\overline{m}$ равна $P(\overline{m})/C_s^{\overline{m}}$.

В дальнейших рассмотрениях мы не будем требовать независимости $\Xi^{(r)}$ $(r = 1, 2, \ldots, k)$, однако будем предполагать, что хотя бы $k-1$ из них равномерно расположены.

Решение поставленной задачи состоит из двух этапов. На первом этапе при допущении равномерной расположенности t — 1-го слагаемого вычисляется условная вероятность

$$P^{\overline{m}}_{\overline{m}^{(1)},\ldots,\overline{m}^{(r)},\ldots,\overline{m}^{(k)}} = \mathcal{P}\left(\mathrm{H} = \overline{m}/\Xi^{(1)} = \overline{m}^{(1)}, \ldots, \Xi^{(1)} = \overline{m}^{(r)}, \ldots, \Xi^{(k)} = \overline{m}^{(k)}\right),$$

которую будем называть *ядром композиции*. Далее, согласно теореме о полной вероятности, вычисляется $P^*(\overline{m})$ по формуле

$$P^*(\overline{m}) = \sum_{\overline{m}^{(1)},\ldots,\overline{m}^{(r)},\ldots,\overline{m}^{(k)}} P\left(\overline{m}^{(1)}, \ldots, \overline{m}^{(r)}, \ldots, \overline{m}^{(k)}\right) P^{\overline{m}}_{\overline{m}^{(1)},\ldots,\overline{m}^{(r)},\ldots,\overline{m}^{(k)}},$$

которая и определяет многомерную композицию распределения $P(\overline{m}^{(1)}, \ldots, \overline{m}^{(r)}, \ldots, \overline{m}^{(k)})$ с ядром композиции $P^{\overline{m}}_{\overline{m}^{(1)},\ldots,\overline{m}^{(r)},\ldots,\overline{m}^{(k)}}$.

Таким образом, задача сводится к вычислению ядра композиции $P^{\overline{m}}_{\overline{m}^{(1)},\ldots,\overline{m}^{(r)},\ldots,\overline{m}^{(k)}}$.

Легко видеть, что при равномерной расположенности t — 1-го слагаемого

$$P^{\overline{m}}_{\overline{m}^{(1)},\ldots,\overline{m}^{(r)},\ldots,\overline{m}^{(k)}} = \frac{1}{C_s^{\overline{m}^{(1)}} \ldots C_s^{\overline{m}^{(r)}} \ldots C_s^{\overline{m}^{(k)}}} \sum_{(*)} C_s^{\{\overline{m}_{\alpha_1},\ldots,\alpha_r,\ldots,\alpha_k\}}; \qquad (1.2)$$

где $(*)$ означает, что суммирование ведется по решениям системы уравнений:

$$\left.\begin{aligned} &\sum_{\alpha_1,\ldots,\alpha_{r-1},\ldots,\alpha_{r+1},\ldots,\alpha_t} m_{\alpha_1\ldots\alpha_{r-1}\alpha_r\alpha_{r+1}\ldots\alpha_k} = m^{(r)}_{\alpha_r} \\ &\qquad (r = \overline{1, k},\ \alpha_r = 0, 1, \ldots, a-1), \\ &\sum_{\sum_{r=1}^{k} \alpha_t = \beta} m_{\alpha_1\alpha_2\ldots\alpha_r\ldots\alpha_t} = m_\beta \quad (\beta = 0, 1, 2, \ldots, a-1); \end{aligned}\right\} (*)$$

при этом суммирование ведется по всем целочисленным неотрицательным решениям $\{m_{\alpha_1\ldots\alpha_r\ldots\alpha_k}\}$ системы уравнений (*), а числа стоящие, под знаком суммы указывают на количество «столбцов» из элементов $\alpha_1 \ldots \alpha_r \ldots \alpha_k$ в схеме (I.1):

$$C_s^{\{m_{\alpha_1\ldots\alpha_r\ldots\alpha_k}\}} = \frac{s!}{\prod\limits_{\alpha_1\ldots\alpha_r\ldots\alpha_k} m_{\alpha_1\ldots\alpha_r\ldots\alpha_k}!}.$$

Дальнейшие результаты для компактности изложения будут получены с помощью производящих.

§ 1.2. Соотношение между производящими и следствия

Особенность метода производящих применительно к стохастическим суммам состоит в том, что первые производящие не могут быть вычислены непосредственно, а приходится прибегать к так называемым вторым производящим.

Этот факт является обычным в теории вероятностей. Так, например, даже для такого элементарного распределения, как гипергеометрическое, не существует компактного выражения первой производящей, в то время как вторая производящая имеет довольно простой вид.

Докажем *основную* для дальнейшего изложения *теорему*, полагая, как и раньше, $w_0^n w_1^{n_1} \dots w_{n_{a-1}}^{n_{a-1}} = \overline{w^n}$.

Теорема I.1. Вторая производящая ядра композиции

$$g(\overline{v}; \overline{u^{(1)}}, \dots, \overline{u^{(r)}}, \dots, \overline{u^{(k)}}) = \sum_{\overline{m}^{(1)},\dots,\overline{m}^{(r)},\dots,\overline{m}^{(k)}} C_s^{\overline{m}^{(1)}} \dots C_s^{\overline{m}^{(r)}} \dots C_s^{\overline{m}^{(k)}} \times$$

$$\times g_{\overline{m}^{(1)},\dots,\overline{m}^{(r)},\dots,\overline{m}^{(k)}} \times (\overline{v})\, \overline{u^{(1)}}^{\overline{m}^{(1)}} \dots \overline{u^{(r)}}^{\overline{m}^{(r)}} \dots \overline{u^{(k)}}^{\overline{m}^{(k)}},$$

где $g_{\overline{m}^{(1)},\dots,\overline{m}^{(r)},\dots,\overline{m}^{(k)}}(\overline{v}) = \sum_{\overline{m}} P^{\overline{m}}_{\overline{m}^{(1)},\dots,\overline{m}^{(r)},\dots,\overline{m}^{(k)}}\, \overline{v}^{\overline{m}}$ имеет вид

$$g(\overline{v}; \overline{u^{(1)}}, \dots, \overline{u^{(r)}}, \dots, \overline{u^{(k)}}) =$$

$$= \Bigg(\sum_{\beta=0}^{a-1} \sum_{\alpha_1+\dots+\alpha_r+\dots+\alpha_k=\beta} u^{(1)}_{\alpha_1} \dots u^{(r)}_{\alpha_r} \dots u^{(k)}_{\alpha_k} v_\beta \Bigg)^s. \tag{I.3}$$

Доказательство. В самом деле,

$$g(\overline{v}; \overline{u^{(1)}}, \dots, \overline{u^{(r)}}, \dots, \overline{u^{(k)}}) = \Bigg(\sum_{\beta=0}^{a-1} \sum_{\alpha_1+\dots+\alpha_r+\dots+\alpha_k=\beta} u^{(1)}_{\alpha_1} \dots u^{(r)}_{\alpha_r} \dots u^{(k)}_{\alpha_k} v_\beta \Bigg)^s =$$

$$= \Bigg(\sum_{\alpha_1 \dots \alpha_r \dots \alpha_k} u^{(1)}_{\alpha_1} \dots u^{(r)}_{\alpha_r} \dots u^{(k)}_{\alpha_k} v_{\alpha_1+\dots+\alpha_r+\dots+\alpha_k} \Bigg)^s =$$

$$= \sum_{\sum_{\alpha_1 \dots \alpha_r \dots \alpha_k} m_{\alpha_1 \dots \alpha_r \dots \alpha_k} = s} C_s^{\{m_{\alpha_1 \dots \alpha_r \dots \alpha_k}\}} \times$$

$$\times \prod_{\alpha_1 \dots \alpha_r \dots \alpha_k} (u^{(1)}_{\alpha_1} \dots u^{(r)}_{\alpha_r} \dots u^{(k)}_{\alpha_k} v_{\alpha_1+\dots+\alpha_r+\dots+\alpha_k})^{m_{\alpha_1 \dots \alpha_r \dots \alpha_k}} =$$

$$= \sum_{\sum_{\alpha_1 \dots \alpha_r \dots \alpha_k} m_{\alpha_1 \dots \alpha_r \dots \alpha_k} = s} C_s^{\{m_{\alpha_1 \dots \alpha_r \dots \alpha_k}\}} \times$$

$$\times \prod_{\alpha_1 \dots \alpha_r \dots \alpha_k} \sum_{\alpha_2 \dots \alpha_k}^{m_{\alpha_1 \dots \alpha_k}} u^{(1)}_{\alpha_1} \sum_{\alpha_1 \dots \alpha_{r-1} \alpha_{r+1} \dots \alpha_k}^{m_{\alpha_1 \dots \alpha_k}} \dots u^{(r)}_{\alpha_r} \sum_{\alpha_1 \dots \alpha_{k-1}}^{m_{\alpha_1 \dots \alpha_k}} \dots u^{(k)}_{\alpha_k} v^{m_{\alpha_1 \dots \alpha_k}}_{\alpha_1+\dots+\alpha_k} =$$

$$= \sum_{\overline{m}^{(1)},\dots,\overline{m}^{(r)},\dots,\overline{m}^{(k)}} \sum_{\alpha_1 \dots \alpha_{r-1} \alpha_{r+1} \dots \alpha_k} C_s^{\{m_{\alpha_1 \dots \alpha_r \dots \alpha_k}\}} \times$$

$$\times \prod_{\alpha_1 \dots \alpha_r \dots \alpha_k} \sum_{m_{\alpha_1 \dots \alpha_r \dots \alpha_k} = m^{(r)}_{\alpha_r}} v^{m_{\alpha_1 \dots \alpha_k}}_{\alpha_1+\dots+\alpha_k} \overline{u^{(1)}}^{\overline{m}^{(1)}} \dots \overline{u^{(r)}}^{\overline{m}^{(r)}} \dots \overline{u^{(k)}}^{\overline{m}^{(k)}} =$$

$$= \sum_{\overline{m}^{(1)},\ldots,\overline{m}^{(r)},\ldots,\overline{m}^{(k)}} C_s^{\overline{m}^{(1)}} \ldots C_s^{\overline{m}^{(r)}} \ldots C_s^{\overline{m}^{(k)}} \times$$

$$\times \left(\sum_{\overline{m}} \frac{1}{C_s^{\overline{m}^{(1)}} \ldots C_s^{\overline{m}^{(r)}} \ldots C_s^{\overline{m}^{(k)}}} \sum_{(*)} C_s^{\{m_{\alpha_1}\ldots\alpha_k\}} \overline{v}^{\overline{m}} \right) \overline{u}^{(1)\overline{m}^{(1)}} \ldots \overline{u}^{(r)\overline{m}^{(r)}} \ldots \overline{u}^{(k)\overline{m}^{(k)}}$$

что и доказывает теорему.

В частном случае двух слагаемых ($r = 2$) имеем:

$$P^{\overline{m}}_{\overline{m}^{(1)},\overline{m}^{(2)}} = \frac{1}{C_s^{\overline{m}^{(1)}} C_s^{\overline{m}^{(2)}}} \sum_{\sum_{\alpha_2} m_{\alpha_1\alpha_2} = m^{(1)}_{\alpha_1}} \sum_{\sum_{\alpha_1} m_{\alpha_1\alpha_2} = m^{(2)}_{\alpha_2}} \sum_{\alpha_1+\alpha_2=\beta} C_s^{\{m_{\alpha_1\alpha_2}\}}; \tag{I.4}$$

$$g(\overline{v}, \overline{u}^{(1)}, \overline{u}^{(2)}) = \left(\sum_{\beta=0}^{a-1} \sum_{\alpha=0}^{a-1} u^{(1)}_{\alpha} u^{(2)}_{\beta-\alpha} v_{\beta} \right)^{s}. \tag{I.5}$$

Рассмотрим случай, когда стохастическую сумму можно рассматривать как стохастическую функцию со специальным видом матрицы переходов. Ниже будет доказана теорема, выясняющая возможность такой трактовки.

Теорема I.2. Пусть задана стохастическая сумма $H = \Xi^{(1)} \dotplus \Xi^{(2)}$, причем $\Xi^{(1)}$ и $\Xi^{(2)}$ независимы; одно из них равномерно расположено (например $\Xi^{(1)}$) и одно (например $\Xi^{(2)}$) полиномиально распределено, т. е.

$$\mathcal{P}(\Xi^{(2)} = \overline{m}^{(2)}) = P^{(2)}(\overline{m}^{(2)}) = C_s^{\overline{m}^{(2)}} \overline{p}^{(1)\overline{m}^{(2)}}, \text{ где } \overline{p}^{(2)} = (p_0^{(2)}, p_1^{(2)}, \ldots, p_{a-1}^{(2)}).$$

Тогда H можно рассматривать как стохастическую функцию $\Xi^{(1)}$, заданную матрицей переходов $\mathbf{p} = \| p^{(2)}_{\beta-\alpha} \|$, поскольку ее распределение $P_s(\overline{m})$, вычисленное согласно каждой из двух вероятностных схем, имеет один и тот же вид.

Доказательство. В самом деле, результат теоремы I.2 получается простой подстановкой в $g(\overline{v}; \overline{u}^{(1)}, \overline{u}^{(2)})$, на место $\overline{u}^{(2)}$ вероятностного вектора $\overline{p}^{(2)}(p_0^{(2)} + p_1^{(2)} + \ldots + p_{a-1}^{(2)} = 1;\ p_\alpha^{(2)} \geqslant 0)$.

Именно, с одной стороны, имеем для стохастических функций

$$g_{\overline{m}^{(1)}}(\overline{v}) = \sum_{\overline{m}} P^{\overline{m}}_{\overline{m}^{(1)}} \overline{v}^{\overline{m}} = \sum_{\overline{m}^{(2)}} P^{(2)}(\overline{m}^{(2)}) \sum_{\overline{m}} P^{\overline{m}}_{\overline{m}^{(1)},\overline{m}^{(2)}} \overline{v}^{\overline{m}} = \sum_{\overline{m}^{(2)}} P^{(2)}(\overline{m}^{(2)}) g_{\overline{m}^{(1)},\overline{m}^{(2)}}(\overline{v}).$$

Итак

$$g_{\overline{m}^{(1)}}(\overline{v}) = \sum_{\overline{m}^{(2)}} P^{(2)}(\overline{m}^{(2)}) g_{\overline{m}^{(1)},\overline{m}^{(2)}}(\overline{v}).$$

Пусть $P^{(2)}(\overline{m}^{(2)}) = C_s^{\overline{m}^{(2)}} \overline{p}^{(2)\overline{m}^{(2)}}$. Полагая $\overline{u}^{(2)} = \overline{p}^{(2)}$ в $g(\overline{v}; \overline{u}^{(1)}, \overline{u}^{(2)})$, получим

$$g(\overline{v}; \overline{u}^{(1)}, \overline{p}^{(2)}) = \sum_{\overline{m}^{(1)},\overline{m}^{(2)}} C_s^{\overline{m}^{(1)}} C_s^{\overline{m}^{(2)}} g_{\overline{m}^{(1)},\overline{m}^{(2)}}(\overline{v}) \overline{u}^{(1)\overline{m}^{(1)}} \overline{p}^{(2)\overline{m}^{(2)}} =$$

$$= \sum_{\overline{m}^{(1)}} C_s^{\overline{m}^{(1)}} \left(\sum_{\overline{m}^{(2)}} C_s^{\overline{m}^{(2)}} \overline{p}^{(2)\overline{m}^{(2)}} g_{\overline{m}^{(1)},\overline{m}^{(2)}}(\overline{v}) \right) \overline{u}^{(1)\overline{m}^{(1)}} = \sum_{\overline{m}^{(1)}} C_s^{\overline{m}^{(1)}} g_{\overline{m}^{(1)}}(\overline{v}) \overline{u}^{(1)\overline{m}^{(1)}}.$$

С другой стороны,

$$g(\overline{v};\ \overline{u}^{(1)},\ \overline{p}^{(2)}) = \left[\sum_{\alpha=0}^{a-1}\left(\sum_{\beta=0}^{a-1} p^{(2)}_{\beta-\alpha}v_\beta\right)u^{(1)}_\alpha\right]^s = \sum_{\overline{m}^{(1)}} C_s^{\overline{m}^{(1)}} \prod_{\alpha=0}^{a-1}\left(\sum_{\beta=0}^{a-1} p^{(2)}_{\beta-\alpha}v_\beta\right)^{m^{(1)}_\alpha} \overline{u}^{\overline{m}^{(1)}}.$$

Сравнивая результаты двух последних соотношений, заключаем, что

$$g_{\overline{m}^{(1)}}(\overline{v}) = \prod_{\alpha=0}^{a-1}\left(\sum_{\beta=0}^{a-1} p^{(2)}_{\beta-\alpha}v_\beta\right)^{m^{(1)}_\alpha}.$$

Таким образом, из предположения о полиноминальности одного из слагаемых мы получили условную производящую $g_{\overline{m}^{(1)}}(v)$ для стохастической суммы, аналогичную условной производящей для стохастической функции с матрицей переходов $\mathbf{p} = \| p^{(2)}_{\beta-\alpha} \|$ (см. (2.82′)), это и доказывает теорему I.2.

Следствие I.2.1. Если $\Xi^{(1)}$ и $\Xi^{(2)}$ распределены полиномиально, т. е. $P^{(1)}(\overline{m}^{(1)}) = C_s^{\overline{m}^{(1)}} \overline{p}^{\overline{m}^{(1)}}$, $P^{(2)}(\overline{m}^{(2)}) = C_s^{\overline{m}^{(2)}} \overline{p}^{\overline{m}^{(2)}}$, и хотя бы одно из них равномерно расположено, то их стохастическая сумма $\mathrm{H} = \Xi^{(1)} \dotplus \Xi^{(2)}$ тоже распределена полиномиально, т. е. $P(\overline{m}) = C_s^{\overline{m}} \overline{q}^{\overline{m}}$, где $q_\beta = \sum_{\alpha=0}^{a-1} p^{(1)}_\alpha p^{(2)}_{\beta-\alpha}$ $\beta = 0, 1, 2, \ldots, a-1$).

Доказательство. В самом деле, как следует из теоремы I.2, H можно рассматривать как стохастическую функцию полиномиально распределенного аргумента $\Xi^{(1)}$ с матрицей переходов $\mathbf{p} = \| p^{(2)}_{\beta-\alpha} \|$, а это приводит к сформулированному выше следствию (см. конец п. 2.8.1).

Можно дать прямое доказательство следствия I.2.1 без использования теоремы I.2. Подставив в $g(\overline{v};\ \overline{u}^{(1)},\ \overline{u}^{(2)})$ значения $\overline{u}^{(1)} = \overline{p}^{(1)}$ и $\overline{u}^{(2)} = \overline{p}^{(2)}$, получим результат следствия I.2.1, заключающийся в том, что если

$$P^{(1)}(\overline{m}^{(1)}) = C_s^{\overline{m}^{(1)}} \overline{p}^{\overline{m}^{(1)}} \quad \text{и} \quad P^{(2)}(\overline{m}^{(2)}) = C_s^{\overline{m}^{(2)}} \overline{p}^{\overline{m}^{(2)}},$$

$$P(\overline{m}) = C_s^{\overline{m}} \overline{q}^{\overline{m}},$$

где

$$q_\beta = \sum_{\alpha=0}^{a-1} p^{(1)}_\alpha p^{(2)}_{\beta-\alpha}.$$

В самом деле, с одной стороны,

$$g(\overline{v};\ \overline{p}^{(1)},\ \overline{p}^{(2)}) = \sum_{\overline{m}^{(1)},\ \overline{m}^{(2)}} C_s^{\overline{m}^{(1)}} C_s^{\overline{m}^{(2)}} g_{\overline{m}^{(1)},\overline{m}^{(2)}}(\overline{v})\, \overline{p}^{\overline{m}^{(1)}}\, \overline{p}^{\overline{m}^{(2)}} =$$

$$= \sum_{\overline{m}^{(1)},\overline{m}^{(2)}} C_s^{\overline{m}^{(1)}} \overline{p}^{\overline{m}^{(1)}}\, C_s^{\overline{m}^{(2)}} \overline{p}^{\overline{m}^{(2)}} \sum_{\overline{m}} P^{\overline{m}}_{\overline{m}^{(1)},\overline{m}^{(2)}}\, \overline{v}^{\overline{m}} =$$

$$= \sum_{\overline{m}} \sum_{\overline{m}^{(1)},\overline{m}^{(2)}} C_s^{\overline{m}^{(1)}} \overline{p}^{\overline{m}^{(1)}}\, C_s^{\overline{m}^{(2)}} \overline{p}^{\overline{m}^{(2)}}\, P^{\overline{m}}_{\overline{m}^{(1)},\overline{m}^{(2)}}\, \overline{v}^{\overline{m}} = \sum_{\overline{m}} P(\overline{m})\, \overline{v}^{\overline{m}},$$

а с другой стороны,

$$g(\bar{v};\bar{p}^{(1)},\bar{p}^{(2)})=\left[\sum_{\beta=0}^{a-1}\left(\sum_{\alpha=0}^{a-1}p_{\alpha}^{(1)}p_{\beta-\alpha}^{(2)}\right)v_{\beta}\right]^{s}=\left[\sum_{\beta=0}^{a-1}q_{\beta}v_{\beta}\right]^{s}=$$
$$=\sum_{\bar{m}}C_{s}^{\bar{m}}\bar{q}^{\bar{m}}\bar{v}^{\bar{m}}.$$

Сравнивая результаты двух последних соотношений, можно заключить, что $P(\bar{m})=C_{s}^{\bar{m}}\bar{q}^{\bar{m}}$, где $q_{\beta}=\sum_{\alpha=0}^{a-1}p_{\alpha}^{(1)}p_{\beta-\alpha}^{(2)}$, что и доказывает следствие I.2.1.

Следствие I.2.2. Достаточным условием равенства $P(\bar{m})=C_{s}^{\bar{m}}\left(\frac{1}{a}\right)^{s}$ при произвольности одного из $P^{(1)}(\bar{m}^{(1)})$ или $P^{(2)}(\bar{m}^{(2)})$ является равенство другого из них $P^{(1)}(\bar{m}^{(1)})=C_{s}^{\bar{m}^{(1)}}(1/a)^{s}$, или $P^{(2)}(\bar{m}^{(2)})=C_{s}^{\bar{m}^{(2)}}(1/a)^{s}$. (Мы снова предполагаем независимость $\Xi^{(1)}$ и $\Xi^{(2)}$ и равномерную расположенность хотя бы одного из них).

Доказательство. В самом деле, если одно из $P^{(1)}(\bar{m}^{(1)})$ или $P^{(2)}(\bar{m}^{(2)})$ полиномиально, например $P^{(2)}(\bar{m}^{(2)})=C_{s}^{\bar{m}^{(2)}}\bar{p}^{\bar{m}^{(2)}}$, то по теореме I.2 это эквивалентно заданию стохастической функции матрицей переходов $\mathbf{p}=\|p_{\beta-\alpha}^{(2)}\|$. Но из следствия 2.3.2 (см. п. 2.7.4.) легко показать, что для того, чтобы распределение стохастической функции имело вид $P(\bar{m})=C_{s}^{\bar{m}}\bar{p}^{\bar{m}}$, достаточно, чтобы $\mathbf{p}=\|p_{\beta}^{(2)}\|$.

Итак, достаточно требовать $\mathbf{p}=\|p_{\beta-\alpha}^{(2)}\|=\|p_{\beta}^{(2)}\|$.

Другими словами, достаточно требование $p_{\beta-\alpha}^{(2)}=p_{\beta}^{(2)}$, но $\sum_{\alpha=0}^{a-1}p_{\beta-\alpha}^{(2)}=1$, поэтому $\sum_{\alpha=0}^{a-1}p_{\beta-\alpha}^{(2)}=\sum_{\alpha=0}^{a-1}p_{\beta}^{(2)}=ap_{\beta}=1$, откуда $p_{\beta}^{(2)}=1/a$, что и доказывает следствие.

Обобщая следствия теоремы I.2 на случай k случайных слагаемых, получим следующие утверждения, *верные в предположении независимости k случайных слагаемых и равномерной расположенности $k-1$-го из них.*

Следствие I.2.3. Если k случайных слагаемых $\Xi^{(r)}(r=\overline{1,k})$ распределены полиномиально, т. е. $P^{(r)}(\bar{m}^{(r)})=C_{s}^{\bar{m}^{(r)}}\bar{p}^{\bar{m}^{(r)}}$, то и их стохастическая сумма $\mathrm{H}=\Xi^{(1)}+\Xi^{(2)}+\ldots+\Xi^{(t)}$ распределена полиномиально, т. е. $P(\bar{m})=C_{s}^{\bar{m}}\bar{q}^{\bar{m}}$, где $q_{\beta}=\sum_{\alpha_1+\ldots+\alpha_r+\ldots+\alpha_k=\beta}p_{\alpha_1}^{(1)}\ldots p_{\alpha_k}^{(r)}\ldots p_{\alpha_k}^{(k)}$.

Следствие I.2.4. Для равенства $P(\bar{m})=C_{s}^{\bar{m}}(1/a)^{s}$ достаточно, чтобы хотя бы одно из слагаемых имело распределение

$$P^{(r)}(\bar{m}^{(r)})=C_{s}^{\bar{m}^{(r)}}(1/a)^{s}.$$

Оба этих обобщенных следствия из теоремы I.2 доказываются последовательным суммированием случайных слагаемых по два.

§ I.3. Соотношения между моментами

Связь между моментами случайных слагаемых и их стохастической суммы легко устанавливается из рассмотрения производящих. Во избежание громоздких вычислений ограничимся случаем двух слагаемых ($k = 2$).

Дифференцированием $g(\bar{v}; \bar{u}^{(1)}, \bar{u}^{(2)})$ по компонентам $\bar{v}$ можно получить выражение для условных факториальных моментов ядра композиции $P_{\bar{m}^{(1)}, \bar{m}^{(2)}}^{\bar{m}}$; тогда имеем

$$\frac{\partial^h g(\bar{v}; \bar{u}^{(1)}, \bar{u}^{(2)})}{\partial v_0^{h_0} \dots \partial v_\beta^{h_\beta} \dots \partial v_{a-1}^{h_{a-1}}}\Bigg|_{v_0=\dots=v_\beta=\dots=v_{a-1}=1} =$$

$$= \frac{\partial^h g(\bar{v}; \bar{u}^{(1)}, \bar{u}^{(2)})}{\partial \bar{v}^{\bar{h}}}\Bigg|_{\bar{v}=\bar{e}} = \sum_{\bar{m}^{(1)}, \bar{m}^{(2)}} C_s^{\bar{m}^{(1)}} C_s^{m^{(2)}} \mathrm{E}_{\bar{m}^{(1)}, \bar{m}^{(2)}} \bar{m}^{(h)} \bar{u}^{\bar{m}^{(1)}}_{(1)} \bar{u}^{\bar{m}^{(2)}}_{(2)},$$

где

$$\mathrm{E}_{\bar{m}^{(1)}, \bar{m}^{(2)}} \bar{m}^{(\bar{h})} = \sum_{\bar{m}} \bar{m}^{(\bar{h})} P_{\bar{m}^{(1)}, \bar{m}^{(2)}}^{\bar{m}},$$

причем

$$\bar{m}^{(\bar{h})} = m_0 (m_0 - 1) \dots (m_0 - h_0 + 1) \dots m_{a-1} (m_{a-1} - 1) \dots (m_{a-1} - h_{a-1} + 1);$$

$$\sum_{\beta=0}^{a-1} h_\beta = h;$$

$$\frac{\partial^h g(\bar{v}; \bar{u}^{(1)}, \bar{u}^{(2)})}{\partial \bar{v}^{\bar{h}}}\Bigg|_{\bar{v}=\bar{e}} = \frac{\partial^h \left[\sum_{\beta=0}^{a-1} \sum_{\alpha=0}^{a-1} u_\alpha^{(1)} u_{\beta-\alpha}^{(2)} v_\beta \right]^s}{\partial \bar{v}^{\bar{h}}}\Bigg|_{\bar{v}=\bar{e}} =$$

$$= \sum_{\bar{m}^{(1)}, \bar{m}^{(2)}} A_{\bar{m}^{(1)}, \bar{m}^{(2)}} (\bar{h}) \bar{u}^{\bar{m}^{(1)}}_{(1)} \bar{u}^{\bar{m}^{(2)}}_{(2)}.$$

Сравнение коэффициентов двух тождественных полиномов от многих переменных при одинаковых степенях последних показывает, что

$$\mathrm{E}_{\bar{m}^{(1)}, \bar{m}^{(2)}} \bar{m}^{(\bar{h})} = \frac{A_{\bar{m}^{(1)}, \bar{m}^{(2)}} (\bar{h})}{C_s^{\bar{m}^{(1)}} C_s^{\bar{m}^{(2)}}},$$

$$\mathrm{E}\bar{m}^{(\bar{h})} = \sum_{\bar{m}^{(1)}, \bar{m}^{(2)}} P(\bar{m}^{(1)}, \bar{m}^{(2)}) \mathrm{E}_{\bar{m}^{(1)}, \bar{m}^{(2)}} \bar{m}^{(\bar{h})};$$

этим и заканчивается вычисление.

Таким образом, задача сводится к вычислению $A_{\bar{m}^{(1)}, \bar{m}^{(2)}} (\bar{h})$. Приведем соотношения для первых факториальных моментов при $h = 1$ и $h = 2$.

Имеем

$$\mathrm{E}m_\beta = \frac{1}{s} \sum_{\alpha=0}^{a-1} \mathrm{E}m_\alpha^{(1)} m_{\beta-\alpha}^{(2)}, \tag{I.6}$$

при $\beta \neq \beta'$

$$\mathrm{E}m_\beta m_{\beta'} = \frac{1}{s(s-1)}\Bigg[\sum_{\alpha=0}^{a-1} \mathrm{E}m_\alpha^{(1)}(m_\alpha^{(1)}-1)\, m_{\beta-\alpha}^{(2)} m_{\beta'-\alpha}^{(2)} + \sum_{\substack{\alpha\neq\alpha' \\ \alpha-\alpha'\neq\beta-\beta'}} \mathrm{E}m_\alpha^{(1)} m_{\beta-\alpha}^{(1)} m_{\alpha'}^{(2)} m_{\beta'-\alpha'}^{(2)} + \sum_{\alpha=0}^{a-1} \mathrm{E}m_\alpha^{(1)} m_{\alpha+\beta'-\beta}^{(1)} m_{\beta-\alpha}^{(2)}(m_{\beta-\alpha}^{(2)}-1), \tag{I.7}$$

$$\mathrm{E}m_\beta(m_\beta - 1) = \frac{1}{s(s-1)}\Bigg[\sum_{\alpha=0}^{a-1} \mathrm{E}m_\alpha^{(1)}(m_\alpha^{(1)}-1)\, m_{\beta-\alpha}^{(2)}(m_{\beta-\alpha}^{(2)}-1) + 2\sum_{\alpha>\alpha'} \mathrm{E}m_\alpha^{(1)} m_{\beta-\alpha}^{(1)} m_{\alpha'}^{(2)} m_{\beta-\alpha'}^{(2)}\Bigg]. \tag{I.8}$$

Рассмотрим стохастические суммы в случае $a = 2$. Нами уже рассматривалось распределение $P^{\overline{m}}_{\overline{m^{(1)}},\,\overline{m^{(2)}}}$. В случае $a = 2$, полагая в нем $m_0^{(1)} = m^{(1)}$, $m_0^{(2)} = m^{(2)}$ и $m_0 = m$ и обозначая $P^{\overline{m}}_{\overline{m^{(1)}},\,\overline{m^{(2)}}} = P^{m(-)}_{m^{(1)},\,m^{(2)}}$, будем иметь

$$P^{m}_{m^{(1)},\,m^{(2)}} = \frac{1}{C_s^{m^{(1)}} C_s^{m^{(2)}}} C_s^{\frac{m+(m^{(1)}+m^{(2)}-s)}{2},\ \frac{s-m+(m^{(1)}-m^{(2)})}{2},\ \frac{s-m-(m^{(1)}-m^{(2)})}{2},\ \frac{m-(m^{(1)}+m^{(2)}-s}{2}} \tag{I.9}$$

где $m = m^{(1)} + m^{(2)} - s + 2l$ ($l = \overline{0,\ s - m^{(1)}}$), причем, не нарушая общности, можно полагать $m^{(1)} + m^{(2)} \geqslant s \geqslant m^{(1)} \geqslant m^{(2)}$.

Производящая $g(v;\ u^{(1)},\ u^{(2)})$ производящих $g_{m^{(1)},\,m^{(2)}}(v)$ распределений $P^{m}_{m^{(1)},m^{(2)}}$ имеет в нашем случае ($a = 2$) следующий вид:

$$g(v;\ u^{(1)},\ u^{(2)}) = [(u^{(1)}u^{(2)} + 1)v + u^{(1)} + u^{(2)}]^s = \sum_{m^{(1)},\ m^{(2)}} C_s^{m^{(1)}} C_s^{m^{(2)}} g_{m^{(1)},\ m^{(2)}}(v)\, u^{(1)\,m^{(1)}}\, u^{(2)\,m^{(2)}},$$

где

$$g_{m^{(1)},\ m^{(2)}}(v) = \sum_m P^{m}_{m^{(1)},\ m^{(2)}} v^m.$$

Прежде всего найдем среднее и дисперсию распределения $P^{m}_{m^{(1)},\ m^{(2)}}$, имеющего самостоятельный интерес. Для этого воспользуемся общими формулами (I.6) и (I.8). которые в нашем случае $a = 2$ имеют следующий вид для условных моментов:

$$\mathrm{E}_{m^{(1)},\ m^{(2)}} m = \frac{1}{s}[m^{(1)}m^{(2)} + (s - m^{(1)})(s - m^{(2)}];$$

$$\mathrm{E}_{m^{(1)},\ m^{(2)}} m(m-1) = \frac{1}{s(s-1)}[m^{(1)}(m^{(1)}-1)\, m^{(2)}(m^{(2)}-1) + (s - m^{(1)})(s - m^{(1)} - 1)(s - m^{(2)})(s - m^{(2)} - 1) + 2m^{(1)}(s - m^{(1)})\, m^{(2)}(s - m^{(2)})].$$

Но $D_{m^{(1)},\, m^{(2)}} m = E_{m^{(1)},\, m^{(2)}} m(m-1) + E_{m^{(1)},\, m^{(2)}} m - (E_{m^{(1)},\, m^{(2)}} m)^2$.

Подставив значения $E_{m^{(1)},\, m^{(2)}} m$ и $E_{m^{(1)},\, m^{(2)}} m(m-1)$ в последнее соотношение, будем иметь после несложных преобразований:

$$E_{m^{(1)} m^{(2)}} m = \frac{1}{s}[m^{(1)}m^{(2)} + (s - m^{(1)})(s - m^{(2)})]; \tag{I.10}$$

$$D_{m^{(1)},\, m^{(2)}} m = \frac{4}{s^2(s-1)} m^{(1)}m^{(2)}(s - m^{(1)})(s - m^{(2)}). \tag{I.11}$$

Для безусловных моментов первого и второго порядков, т. е. для первых моментов двух случайных слагаемых и их стохастической суммы, можно получить простые связи, но после большего числа преобразований, чем в случае условных моментов. Ввиду важности этих связей они приводятся ниже.

Рассмотрим для простоты случай независимых случайных слагаемых ($P(m^{(1)}, m^{(2)}) = P^{(1)}(m^{(1)}) P^{(2)}(m^{(2)})$). Отсюда имеем

$$Em = \frac{1}{s}[Em^{(1)}Em^{(2)} + (s - Em^{(1)})(s - Em^{(2)})] \tag{I.12}$$

и

$$\begin{aligned} Dm = {} & \frac{4}{s^2(s-1)} Em^{(1)}Em^{(2)}(s - Em^{(1)})(s - Em^{(2)}) + \\ & + \left[1 - \frac{4}{s(s-1)} Em^{(1)}(s - Em^{(1)})\right] Dm^{(2)} + \frac{4}{s(s-1)} Dm^{(1)}Dm^{(2)} + \\ & + \left[1 - \frac{4}{s(s-1)} Em^{(2)}(s - Em^{(2)})\right] Dm^{(1)} = Dm^{(1)} - 4\frac{Dm^{(1)}Dm^{(2)}}{s^2(s-1)} + \\ & + Dm^{(2)} + \frac{4}{s^2(s-1)}[Em^{(1)}(s - Em^{(1)}) - sDm^{(1)}][Em^{(2)}(s - Em^{(2)}) - sDm^{(2)}]. \end{aligned} \tag{I.13}$$

Из формул (I.12) и (I.13) можно получить как частный случай выведенные ранее формулы (I.10) и (I.11) для условных среднего и дисперсии. В самом деле, в этом частном случае $Em^{(1)} = m^{(1)}$, $Em^{(2)} = m^{(2)}$ и $Dm^{(1)} = Dm^{(2)} = 0$, что обеспечивает получение формул (I.10) и (I.8) из (I.12) и (I.13).

Как и в случае стохастических функций при $a = 2$ (см. § 2.9), представим средние и дисперсии частот $m^{(1)}$, $m^{(2)}$ и m в виде отклонений от биномиальных средних и дисперсий

$$Em^{(1)} = sp^{(1)} + \Delta^{(1)}, \quad Dm^{(1)} = sp^{(1)}q^{(1)} + \Delta^{(1)}; \quad Em^{(2)} = sp^{(2)} + \Delta^{(2)},$$

$$Dm^{(2)} = sp^{(2)}q^{(2)} + \Delta^{(2)}; \quad Em = sp + \Delta; \quad Dm = spq + \Delta,$$

где $q^{(1)} = 1 - p^{(1)}$; $q^{(2)} = 1 - p^{(2)}$; $q = 1 - p$; $p = p^{(1)}p^{(2)} + q^{(1)}q^{(2)}$. Тогда для Δ будем иметь согласно (I.12)

$$\Delta = 2\left[-\left(p^{(2)} - \frac{1}{2}\right)\Delta^{(1)} + \frac{\Delta^{(1)}\Delta^{(2)}}{s} - \left(p^{(1)} - \frac{1}{2}\right)\Delta^{(2)}\right]. \tag{I.14}$$

Для Δ будем иметь согласно (I.13)

$$\begin{aligned} \Delta = {} & 4\left\{\frac{\Delta^{(1)}\Delta^{(2)}}{s-1}\left[2\left(\frac{1}{2} - p^{(1)}\right) - \frac{\Delta^{(1)}}{s}\right]\left[2\left(\frac{1}{2} - p^{(2)}\right) - \frac{\Delta^{(2)}}{s}\right] + \right. \\ & + \left[\left(\left(\frac{1}{2} - p^{(1)}\right) - \frac{\Delta^{(1)}}{s-1}\right)^2 - \frac{1}{s}\left(\frac{\Delta^{(1)}}{s-1}\right)^2\right]\Delta^{(2)} + \frac{\Delta^{(1)}\Delta^{(2)}}{s(s-1)} + \\ & \left. + \left[\left(\left(\frac{1}{2} - p^{(2)}\right) - \frac{\Delta^{(2)}}{s-1}\right)^2 - \frac{1}{s}\left(\frac{\Delta^{(2)}}{s-1}\right)^2\right]\Delta^{(1)}. \right. \end{aligned} \tag{I.15}$$

Рассмотрим частные случаи формул (I.14) и (I.15) при частных предположениях относительно $p^{(1)}$ и $p^{(2)}$, являющиеся произвольными.

Пусть $p^{(1)} = p^{(2)} = \frac{1}{2}$; тогда:

$$\delta = 2\frac{\delta^{(1)}\delta^{(2)}}{s}; \tag{I.16}$$

$$\Delta + \delta^2 = \frac{4}{s(s-1)}(\Delta^{(1)} + \delta^{(1)^2})(\Delta^{(2)} + \delta^{(2)^2}). \tag{I.17}$$

Формулы (I.16) и (I.17) легко обобщаются на случай k независимых слагаемых, т. е.:

$$\delta = \left(\frac{2}{s}\right)^{k-1}\delta^{(1)}\delta^{(2)}\ldots\delta^{(r)}\ldots\delta^{(k)}; \tag{I.18}$$

$$\Delta + \delta^2 = \left(\frac{4}{s(s-1)}\right)^{k-1}(\Delta^{(1)} + \delta^{(1)^2})(\Delta^{(2)} + \delta^{(2)^2})\ldots(\Delta^{(r)} + \Delta^{(r)^2})\ldots$$
$$\ldots(\Delta^{(k)} + \delta^{(k)^2}), \tag{I.19}$$

где δ и Δ — отклонения от $s/2$ и $s/4$ соответственно среднего и дисперсии стохастической суммы независимых слагаемых; $\delta^{(r)}$ и $\Delta^{(r)}$ $(r = \overline{1, k})$ — аналогичные отклонения у слагаемых.

В случае k одинаково распределенных независимых случайных слагаемых $\delta^{(r)} = \delta^{(1)}$ и $\Delta^{(r)} = \Delta^{(1)}$ $(r = \overline{1, k})$ формулы (I.18) и (I.19) упрощаются и имеют вид:

$$\delta = \left(\frac{2}{s}\right)^{k-1}\delta^{(1)^k}; \tag{I.20}$$

$$\Delta + \delta^2 = \left(\frac{4}{s(s-1)}\right)^{k-1}(\Delta^{(1)} + \delta^{(1)^2})^k. \tag{I.21}$$

Дополнение II

ТЕОРИЯ ОСУЩЕСТВИМОСТИ

§ II.1. Введение

До недавнего времени возникавшие в статистической теории связи задачи, по крайней мере в постановках, не выходили за рамки математической статистики или общей теории решений (их решение часто было сопряжено с большими аналитическими трудностями). Однако в последнее время все чаще возникают практические задачи, выходящие уже в постановках за рамки существующих математических теорий. В этом, по-видимому, проявляется тенденция обгона быстро развивающейся кибернетической практикой возможностей существующих математических теорий.

Несмотря на различное происхождение этих статистических задач их можно считать разновидностями одной проблемы — отыскания оптимальных решающих процедур в сложных ситуациях в условиях неполных вероятностных знаний анализируемых явлений. При этом в известных постановках предполагаются полные вероятностные знания, например заранее известные вероятностные свойства шумов и сигналов.

Решение новых статистических задач с меньшими начальными данными, требует использования новых математических методов, какими являются теоретические методы, учитывающие возможность использования быстродействующих электронных машин дискретного действия. Необходимо получить общие аналитические или алгоритмические соотношения, связывающие параметры чисто статистических теорий (a — отношение сигнал/шум; α — вероятность ложной тревоги; $1-\beta$ — вероятность правильного обнаружения; T — время наблюдения) с параметрами практики использования электронных машин (V *операций/сек* — быстродействие; M — объем памяти; $P(T)$ — надежность, т. е. вероятность безотказной работы системы за время T).

Необходимость получения оптимальных соотношений такого рода особенно остро возникает там, где требуется создание сложных и дорогостоящих радиотехнических систем.

По своей идее теория оптимальной машинной реализации статистических процедур является развитием теории потенциальной помехоустойчивости (В. А. Котельникова) в условиях возможного использования машинной реализации идеального приемника. Выводы из этой теории позволяют судить о потенциальной осуществимости машинной реализации статистических процедур.

В настоящем дополнении развивается вероятностная теория функционирования сложных систем обработки информации и принятия решений при наличии шумов. При этом такие системы рассматриваются как некоторые кибернетические устройства (КУ), осуществляющие оптимальные алгоритмы без конкретизации их вещественной и энергетической реализации. В такой системе выделяются три фактора:

A — внутренняя иерархическая структура соподчиняющихся частей системы; S^* и S — алгоритмы (стратегии) внутреннего и внешнего поведения системы, соответственно. Алгоритм S^* используется для поддержания бесперебойного функционирования системы, а алгоритм S — для осуществления некоторой внешней цели. Структура системы A может перестраиваться (усовершенствоваться) в ходе работы системы, однако здесь на период осуществления системой определенной цели она считается неизменной.

Качество работы системы оценивается вероятностью осуществления $P(T)$ системой определенной цели за время T. Эта вероятность зависит от выбора факторов (S^*, A, S). Оптимальными для фиксированного T будем называть такие факторы, при которых $P(T)$ достигает максимума. Устанавливается, что при достаточно общих предположениях $P(T)$ как функция T имеет единственный максимум, отличный от единицы, а именно $\max\limits_{T} P(T) = P(T_0)$. Величина этого максимума $P(T_0)$ определяет при оптимальных в точке T_0 факторах (S^*, A, S) потенциальную возможность системы осуществить поставленную перед ней цель. Задача теории осуществимости состоит в построении оптимальных факторов (S^*, A, S) и в выражении $P(T)$ через параметры этих факторов (V, M и др.).

§ II.2. Основные соотношения теории осуществимости

II.2.1. *Вводные замечания.* В этом параграфе рассматриваются принципы, по которым можно судить об эффективности работы сложных кибернетических устройств (КУ) при наличии шумов. На основе этих принципов далее развивается теория таких систем. Она основывается на синтезе разработанной теории обнаружения сигналов на фоне шумов и практики использования больших электронных машин дискретного действия для обработки информации. Огромные возможности таких машин для решения разнообразных кибернетических задач все же ограничены, и необходима точная математическая формулировка этих ограничений. При этом существенна разработка в некотором смысле оптимальных принципов использования электронных машин для обработки информации при наличии шумов.

В самом деле, техническое решение современных сложных кибернетических задач требует дорогостоящей аппаратуры и энергии. Если предварительная оценка показывает, что затраты на них оправданы, то оптимальное решение не является столь актуальным. Если же предварительные оценки показывают, что решение важной задачи на имеющемся уровне техники невозможно, то возникает вопрос о том, в какой мере предварительные оценки исходят из учета оптимального использования имеющихся возможностей. Если это не так, то отрицательное заключение о возможности решения данной задачи ставится под сомнение. Для возможности построения количественной теории рассматриваемых вопросов необходимо точно определить понятие осуществимости решения той или иной задачи, а также установить перечень важнейших параметров, определяющих уровень техники, таких, как быстродействие, объем памяти и др.

II.2.2. *Основные определения.* Приведем общие определения рассматриваемых объектов и ситуаций. При этом для упрощения изложения не будем давать точных аксиоматических определений в современном математическом смысле, хотя это и можно сделать.

Под сложным кибернетическим устройством (КУ) понимается система, которая действует автоматически без участия человека по за-

ранее составленной программе с различной степенью гибкости вплоть до самоанализа, самообучения и самоусовершенствования. В общем виде КУ мыслится как *иерархическая система* A соподчиняющихся разнородных частей $A_{ijk}\ldots$

$$A=(A_i),\quad A_i=(A_{ij}),\quad A_{ij}=(A_{ijk})\ \text{и т. д.}$$

Кибернетическое устройство может быть использовано для решения определенной задачи (определенной цели). При этом КУ использует некоторую *стратегию* S (способ поведения). КУ должно по возможности бесперебойно функционировать хотя бы на период решения внешней задачи. Это достигается определенной *организацией* S^* самоконтроля КУ, рассматриваемого в данном случае в виде иерархической системы соподчиняющихся частей.

В целом при постановке перед КУ определенной задачи (цели) мы располагаем для ее решения тремя факторами: S^*, A и S, называемыми далее (S^*, A, S)-реализацией решения. Здесь фактор A является определенной схемой функциональной структуры КУ, быть может, и изменяющейся в ходе его функционирования. Организация S^* и стратегия S являются некоторыми процедурами, совершаемыми КУ для достижения определенной цели. Могут быть разные степени внешней активности КУ, и возможен крайний случай внешне пассивного КУ, единственная цель которого состоит в бесперебойном внутреннем функционировании. Однако здесь мы будем заниматься активными КУ, организация и стратегия которых подчинены достижению определенных внешних целей.

По-видимому, невероятностная постановка вопроса об эффективности работы КУ не оправдана, так как в реально действующих КУ и вне их присутствуют мешающие их работе шумы. Поэтому естественно характеризовать качество (S^*, A, S)-реализации решения КУ поставленной задачи *вероятностью решения* P задачи с его помощью. При этом вполне естественно считать, что эта вероятность зависит от времени T бесперебойной работы КУ, а также от конкретного вида (S^*, A, S)-реализации.

Важным является следующее представление вероятности решения

$$P(T)=P_1(T)P_2(T), \tag{II.1}$$

где $P_1(T)$ — условная вероятность решения при бесперебойной работе КУ за время T; $P_2(T)$ — вероятность бесперебойной работы КУ за то же время.

При фиксированном T будем называть (S^*, A, S)-реализацию из некоторого класса K оптимальной, если в этом классе нет ни одной (S^*, A, S)-реализации, приводящей к большей вероятности решения $P(T)$, чем оптимальная реализация. Из сказанного видно, что может быть не только одна оптимальная (S^*, A, S)-реализация. В случае, если (S^*, A, S)-реализация оптимальна при всех T $(0\leqslant T<\infty)$, то будем называть ее *равномерно-оптимальной.*

Организацию S^*, иерархическую структуру A и стратегию S, приводящие к оптимальной или равномерно-оптимальной (S^*, A, S)-реализации, будем называть *соответственно оптимальными или равномерно-оптимальными.*

Теперь можно определить, что мы будем иметь в виду, говоря об осуществимости решения задачи при данном уровне имеющихся технических средств. К ним относится, в частности, *быстродействие* V,

т. е. число элементарных операций, производимых дискретно работающим КУ в секунду. Этот параметр имеет особое значение, связывая «мысленные» кибернетические эксперименты с физическим временем.

Если имеется (S^*, A, S)-реализация решения данной задачи, приводящая за время T к решению с вероятностью $P(T)$, то мы будем говорить, что решение задачи $P(T)$-осуществимо.

Пусть из практических соображений выбрана допустимая вероятность $P^{(0)}$ решения задачи за приемлемое время $T^{(0)}$. Тогда пару значений $(P^{(0)}, T^{(0)})$ будем называть *порогами осуществимости*. Основания к конкретному выбору значений $(P^{(0)}, T^{(0)})$ здесь не рассматриваются, очевидна лишь тенденция выбора значения $P^{(0)}$, близкого к единице, а $T^{(0)}$ — не слишком большого.

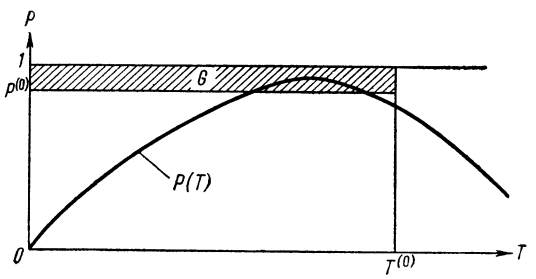


Рис. II.1. Качественное поведение функции осуществимости и пороги осуществимости

Пусть решение задачи $P(T)$-осуществимо. Мы скажем, что оно *осуществимо*, если для некоторого T одновременно имеем $P(T) \geqslant P^{(0)}$ и $T \leqslant T^{(0)}$. Наоборот, решение задачи *не осуществимо*, если для $P(T)$-осуществимой оптимальной (S^*, A, S)-реализации нарушается хотя бы одно из указанных неравенств. В последнем случае функция $P(T)$ может быть названа функцией оптимальной осуществимости, так как она определяет предельные возможности осуществления решения задачи.

До количественного определения функции осуществимости $P(T)$ можно привести ряд качественных соображений относительно ее поведения.

Ясно, что с ростом времени T, отводимого на решение задачи, вероятность $P_1(T)$ решения (при условии бесперебойной работы КУ) не убывает, а вероятность $P_2(T)$ бесперебойной КУ не возрастает. В результате функция осуществимости $P(T)$, согласно (II.1), оказывается произведением неубывающей и невозрастающей с ростом T функций. Значит, существует максимум $P(T)$, вообще говоря отличный от единицы.

Качественное поведение функции осуществимости $P(T)$ и задание порогов осуществимости можно изобразить на плоскости (T, P) (рис. II.1). Задание порогов осуществимости выделяет допустимую область G значений переменных T и P (заштрихована на рис. II.1). Решение осуществимо в случае, когда функция осуществимости $P(T)$ имеет хотя бы одну общую точку с областью G. При этом естественно строить оптимальную (S^*, A, S)-реализацию, основываясь на точке области G с минимальным значением T.

Теория осуществимости ставит целью дать оценку осуществимости решения общих технических задач, например задачи обнаружения сигналов и различения образов на фоне шумов, и других при использовании электронных машин дискретного действия, эффективность которых ограничена такими параметрами, как быстродействие, память и пр.

В заключение остановимся на связи теории осуществимости с другими математическими областями. Современная теория решений, переплетающаяся с теорией игр, ищет оптимальные стратегии S в предположении бесперебойности работы решающих схем.

Теория надежности и теория синтеза надежных конечных автоматов, состоящих из ненадежных частей, ищет оптимальные [в смысле максимизации $P_2(T)$] организации S^* и иерархические структуры A, отвлекаясь от внешних проявлений КУ. Теория осуществимости в теоретическом плане должна явиться синтезом двух указанных направлений. Как упоминалось ранее, она основывается на практике переработки информации на фоне шумов с помощью электронных машин дискретного действия. Для дальнейшего развития теории необходимы допущения о конкретном виде зависимости $P(T)$ от T.

II.2.3. *Основные соотношения в экспоненциальном случае.* В теории надежности доказывается универсальность экспоненциального закона зависимости от T вероятности $P_2(T)$, называемой *надежностью*. Можно показать, что экспоненциальная зависимость вероятности $P_1(T)$ от T также носит универсальный характер.

Ранее было приведено доказательство экспоненциального характера $P_1(T)$ для случаев оптимального обнаружения сигналов на фоне шумов (гл. 3) и в теории оптимального кодирования (гл. 5, 6, 7). Здесь развивается теория осуществимости в предположении об экспоненциальном характере зависимости вероятностей $P_1(T)$ и $P_2(T)$ от T.

$$P_1(T) = 1 - e^{-f_1 T} \tag{II.2}$$

и

$$P_2(T) = e^{-T/f_2}, \tag{II.3}$$

где коэффициенты f_1 и f_2 будем называть внешней и внутренней эффективностью стратегии S и организации S^*, соответственно (f_2 равно среднему времени между отказами КУ).

Статистические теории не учитывают надежности систем, осуществляющих статистические процедуры (они предполагаются абсолютно надежными, $f_2 = \infty$) и здесь с ростом T вероятность решения задачи $P(T) = P_1(T)$ стремится к единице. В теории осуществимости (см. п. II.2.2) существует предельно большое значение $\overline{P} = P(T)$, вообще говоря, отличное от единицы. В случае оптимальной реализации эта величина $\overline{P} = P(T_1) = \max\limits_T P(T)$ имеет предельно большое значение вероятности решения задачи. В рассматриваемом экспоненциальном случае можно получить выражение $\overline{P}$ через параметры f_1 и f_2.

Приступим к вычислению $\overline{P}$. Согласно (II.1), (II. 2) и (II.3), вероятность $P(T)$ решения задачи имеет вид

$$P(T) = (1 - e^{-f_1 T})\, e^{-T/f_2},$$

откуда

$$P'(T) = e^{-(f_1 + f_2^{-1})T}\,[f_1 - f_2^{-1}(e^{f_1 T} - 1)] = f_1 e^{-(f_1 + f_2^{-1})T} - f_2^{-1} P(T)$$

и

$$P''(T) = e^{-(f_1 + f_2^{-1})T}\,[-f_1(f_1 + f_2^{-1}) - f_1 f_2^{-1} + f_2^{-2}(e^{f_1 T} - 1)] =$$

$$= -f_1(f_1 + f_2^{-1})\, e^{-(f_1 + f_2^{-1})T} - f_2^{-1} P'(T).$$

Приравнивая $P'(T)$ нулю, находим точку $T_1 = T(f_1\ f_2)$ экстремума

$$T_1 = T(f_1 f_2) = \frac{1}{f_1}\ln(1 + f_1 f_2) = f_2 \frac{1}{f_1 f_2}\ln(1 + f_1 f_2), \tag{II.4}$$

в которой $P(T)$ обращается в максимум, равный

$$P(T_1) = \overline{P}(f_1, f_2) = \left(1 - \frac{1}{1 + f_1 f_2}\right)(1 + f_1 f_2)^{-\frac{1}{f_1 f_2}} = f_1 f_2 (1 + f_1 f_2)^{-\frac{1+f_1 f_2}{f_1 f_2}}, \tag{II.5}$$

а

$$P'(T_1) = 0$$

и

$$P''(T_1) = -f_1 f_2^{-1}(1 + f_1 f_2)^{-\frac{1}{f_1 f_2}} < 0. \tag{II.6}$$

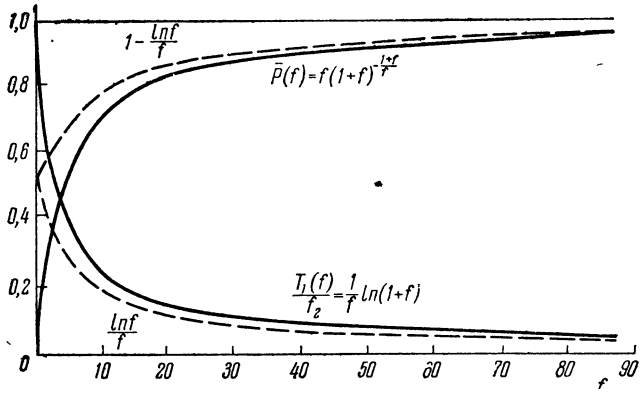


Рис. II.2. Зависимости $\overline{P}(f)$ и $T_1(f)/f_2$ от f

Введем параметр $f = f_1 f_2$ (назовем его эффективностью КУ), через который можно выразить $\overline{P}$ и T_1.

Из (II.4), (II.5) и (II.6) имеем:

$$T_1 = T_1(f) = f_1^{-1}\ln(1 + f) = f_2 \frac{1}{f}\ln(1 + f);$$

$$\overline{P} = P(T_1) = \overline{P}(f) = \left(1 - \frac{1}{1 + f}\right)(1 + f)^{-\frac{1}{f}} = f(1 + f)^{-\frac{1+f}{f}}$$

и

$$P''(T_1) = -\frac{f_1^2}{f}(1 + f)^{-\frac{1}{f}} = -\frac{1}{f_2^2}(1 + f)\overline{P}(f).$$

Параметр f меняется в пределах $0 \leqslant f < \infty$, при этом $T_1(f)/f_2$ меняется в пределах $1 \geqslant T_1(f)/f_2 \geqslant 0$ и $\overline{P}(f)$ изменяется в пределах $0 \leqslant \overline{P}(f) \leqslant 1$.

На рис. II.2 приведена зависимость предельного значения вероятности $\overline{P} = \overline{P}(f)$ решения задачи и потребного на это времени $T_1(f)/f_2$, нормированного средним временем между отказами f_2, в зависимости от эффективности f КУ. Асимптотические формулы при $f \to 0$ и $f \to \infty$ имеют вид:

$$T_1 = T_1(f) = \begin{cases} f_2\left(1 - \frac{f}{2}\right), & \text{при } f \to 0, \\ f_2 \frac{\ln f}{f}, & \text{при } f \to \infty; \end{cases} \tag{II.7}$$

$$\overline{P} = \overline{P}(f) = \begin{cases} f/e + o(f), & \text{при } f \to 0, \\ 1 - \frac{\ln f}{f} + O(f^{-1}), & \text{при } f \to \infty. \end{cases} \tag{II.8}$$

Зная величины $P(T_1)$ и $P''(T_1)$, можно вычислить значения $P(T)$ в окрестности точки T_1 по формуле:

$$P(T) \approx P(T_1) + \frac{(T-T_1)^2}{2} P''(T_1) =$$

$$= \overline{P}(f)\left\{1 - \frac{\left[T/f_2 - \frac{1}{f}\ln(1+f)\right]^2}{2}(1+f)\right\}.$$

Основную практическую ценность имеют значения вероятности $\overline{P}(f)$, близкие к единице.

На рис. II. 3 и II. 4 приведены значения $\overline{P}(f)$, близкие к единице, вычисленные по асимптотической формуле (II.8) при $f \to \infty$, начиная с $f = 150$, когда эта формула дает почти полное совпадение с точной. Согласно асимптотической формуле (II.7), значение T_1/f_2 вычисляется вычитанием из единицы значения $\overline{P}(f)$.

Рассмотрим несколько числовых примеров.

Пример II. 1. Пусть среднее время между отказами КУ $f_2 = 1$ *сутки* $= 24 \cdot 60 \cdot 60 \approx 8 \cdot 10^5$ *сек*. Предположим, что для осуществления некоторой цели используется КУ с внешней эффективностью $f_1 = 0{,}01$ $^1/_{сек}$. Пороги осуществимости $T^{(0)} = 15$ *мин* и $P^{(0)} = 0{,}995$. Определим, осуществимо ли достижение цели, поставленной перед КУ.

Решение. Подсчитаем параметр эффективности КУ $f = f_1 \cdot f_2 = 10^{-2} \cdot 8 \cdot 10^5 = 8 \cdot 10^3$. По графику рис. II.4 найдем по $f = 8 \cdot 10^3$ значение $\overline{P}(8 \cdot 10^3) = 0{,}9989$. Далее, по $\overline{P}(8 \cdot 10^3)$ находим величину $T_1(f) = f_2(1 - \overline{P}(f)) = 8 \cdot 10^5 \cdot 10^{-3} = 8 \cdot 10^2$ *сек* $= 13$ *мин* 20 *сек*. Таким образом, в нашем случае $\overline{P}(f) \geqslant P^{(0)}$ и $T_1(f) < T^{(0)}$, т. е. поставленная цель осуществима с помощью рассмотренного КУ.

Пример II. 2. Определить среднее время f_2 безотказной работы КУ и внешнюю эффективность f_1 его работы, при которых цель, поставленная перед ним, была осуществлена при порогах осуществимости $P^{(0)} = 0{,}99$ и $T^{(0)} = 50$ *сек*.

Решение. Определим по графику рис. II. 3 значение f, приводящее к $\overline{P}(f) \geqslant P^{(0)} = 0{,}99$. Оно равно $f = f_1 \cdot f_2 \geqslant 650$, и ему соответствует величина $T_1(f) = f_2[1 - \overline{P}(f)] = f_2\, 0{,}01 \leqslant 50$ *сек*. Отсюда $f_2 \leqslant 5000$ *сек* $=$ $= 1$ *час* 23 *мин* 20 *сек* и $f_1 \geqslant 650/5000 = 0{,}13$ $^1/_{сек}$.

В заключение отметим, что полученная простая зависимость максимальной вероятности осуществления $\overline{P}(f)$ и соответствующего времени осуществления $T_1(f)$ от эффективности КУ $f = f_1 f_2$ нацеливает дальнейшие

исследования на конкретизацию множителей параметра f, а именно параметра f_1 — внешней эффективности работы КУ и параметра f_2 — среднего времени безотказной работы.

II. 2.4. Выводы. 1. Вероятность достижения сложной системой поставленной цели при наличии шумов и возможных сбоев не может быть более $\overline{P}(f) < 1$ за время $T_1(f)$.

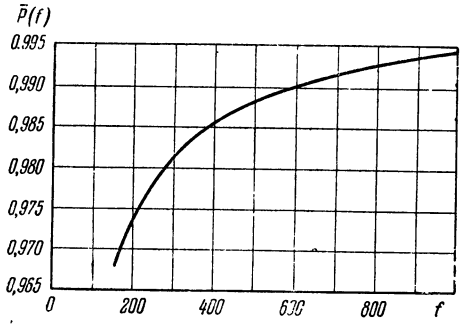


Рис. II.3. Асимптотическая зависимость $\overline{P}(f) \approx 1 - T_1(f)/f_2$ от f, при $\overline{P}(f) \to 1$ ($f = 150 \div 1000$)

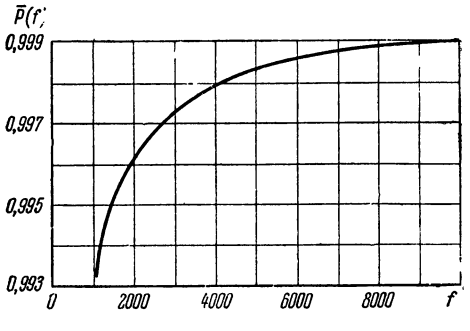


Рис. II.4. Асимптотическая зависимость $P(f) \approx 1 - T_1(f)/f_2$ от f, при $\overline{P}(f) \to 1$ ($f = 1000 \div 10\,000$)

2. Вероятность $P(f)$ и время $T_1(f)/f_2$, нормированное средним временем между отказами f_2, выражаются через параметр $f = f_1 \cdot f_2$—эффективность КУ, где f_1 — внешняя эффективность его работы. Эта зависимость в наиболее интересном для практики случае, когда значение $\overline{P}(f)$ близко к единице ($f \to \infty$), имеет вид

$$1 - \overline{P}(f) \approx T_1(f)/f_2 \approx \ln f / f.$$

§ II.3. Учет ограниченности быстродействия кибернетического устройства

II.3.1. *Вводные замечания.* В § II.1 мы столкнулись с параметром быстродействия V, который был определен как число операций в секунду. В общем случае параметр быстродействия определяется как число элементарных операций в секунду, где под элементарными операциями могут подразумеваться элементарные акты в ходе осуществления кибернетическим устройством (КУ) определенных целей. Параметр быстродействия V не следует смешивать с параметром быстротечности не зависящих от нас физических процессов.

Параметр быстродействия V связывает «мысленные» эксперименты теории осуществимости с реальным физическим временем. Это позволяет определять реальное время T, необходимое для осуществления КУ некоторой цели с удовлетворяющей нас вероятностью $P = P_V(T)$. Ясно, что с ростом V и при фиксированном P происходит уменьшение T, чем повышается возможность реализации процедуры. Вычисление вероятности $P(T)$ успешного проведения статистической процедуры за время T, пропорциональное числу наблюдений, является задачей математической статистики. Однако последняя не учитывает быстродействия аппаратуры, проводящей процедуру. Для возможности учета быстродействия необходим более подробный анализ факторов, мешающих решению задачи.

В самом деле, в различных ситуациях статистических решений, выбора между гипотезами, оценки параметров, поиска при наличии помех и т. д.

решение задачи осложняется двумя видами факторов, имеющих принципиальное различие. К первому виду мешающих факторов относятся природные шумы, рост уровня которых понижает вероятность решения. Увеличить последнюю можно за счет увеличения времени наблюдения или количества попыток достижения цели. Ко второму виду мешающих факторов относятся неизвестные нам, но постоянные на протяжении статистической процедуры параметры ситуации.

Например, в задаче обнаружения сигнала на фоне помех к первому виду факторов относятся помехи, а ко второму неизвестные параметры сигнала. В задаче оценки параметров сигнала на фоне помех при ограниченном числе наблюдений оцениваются лишь вероятные диапазоны их возможных значений, точные значения при ограниченном числе наблюдений установить нельзя. В задачах поиска и погони к первому виду факторов, как и в задаче обнаружения, относятся помехи, а ко второму — параметры активного и пассивного объектов.

В дальнейшем изложении мы будем придерживаться терминологии задач обнаружения. Пусть необходимо произвести выбор между гипотезами $H_0(a = a_0)$ и $H_1(a = a_1 \neq a_0)$ о значениях параметра a плотности вероятности $f_{a,b}(x)$ случайной величины $\xi_{a,b}$, зависящей от проверяемого параметра a и от векторного параметра $b = (b_1, \ldots, b_k)$ с постоянными, но неизвестными значениями компонент. Обычно задают априорное распределение $P(b)$ параметра b, и с его помощью усредняют плотность $f_{a,b}(x)$, получая плотность

$$f_a(x) = \underbrace{\int \ldots \int}_{k} f_{a,b}(x)\, P(c)\, db_1 \ldots db_k, \tag{II.9}$$

относительно которой и решают задачу выбора между гипотезами H_0 и H_1. Однако усреднение (II.9) существенно усложняет задачу обнаружения, «сближая» гипотезы.

Указанный эффект хорошо известен в задачах когерентного и некогерентного обнаружения сигнала на фоне белого шума. Во втором случае усреднение по фазе при отношении сигнал/шум меньше единицы приводит к возведению его в квадрат. В достаточно общем случае можно показать, что усреднение по параметрам приводит к «сближению» гипотез. Во всяком случае имеются разнообразные многоканальные системы обнаружения с разрешением по неизвестным параметрам сигнала, которые позволяют избежать указанный нежелательный эффект. Однако их использование приводит к новым трудностям, например к возрастанию вероятности ошибки первого рода (принятие шума за сигнал).

Принципиальные технические трудности, связанные с реализацией многоканальных систем, состоят в необходимости обработки огромного потока информации, возрастающего с ростом числа M каналов. Одновременная обработка такого потока требует большого числа M приборов, а последовательная обработка одним прибором требует большого быстродействия и памяти. Отметим, что аналогичная ситуация имеет место при реализации схем оптимального по Шеннону декодирования (см. гл. 9).

Особо остро встает вопрос о быстродействии в случае, когда необходимо вынести решение сразу же после анализа быстротечного процесса. Ясно, что при неограниченном быстродействии и памяти можно иметь существенные выигрыши от использования многоканальных систем. С точки зрения параметров осуществимости $P(T)$ и T представляет интерес оценка выигрыша от использования многоканальных систем разрешения по неизвестным параметрам при ограниченном быстродействии и памяти.

II.3.2. *Последовательный во времени статистический анализ компонент.* Пусть требуется найти значения компонент неизвестного векторного

параметра $b = (b_1, \ldots, b_i, \ldots, b_k)$, о которых известно[1], что они заключены в пределах $\underline{b_i} \leqslant b_i < \overline{b_i}$ $(i = \overline{1, k})$. Ограниченная разрешающая способность приборов приводит к неразличимости значений $\Delta b_i \leqslant h_i$. Тогда каждая из компонент b_i определяется дискретным набором $N_i = (\overline{b_i} - \underline{b_i}) h_i$ своих значений, а вектор b определяется $N = \sum_{i=1}^{k} N_i$ своими значениями.

Будем называть N_i вариантностью b_i, а N — вариантностью b. Мультипликативное выражение вариантности вектора b через вариантности его компонент является основным препятствием в реализации многоканальных систем разрешения по многомерным векторным параметрам. Далее считаем, что все векторы имеют дискретные компоненты.

Введем обозначения $b_{(i)} = (b_1, \ldots, b_i)$ и $b^{(i)} = (b_{i+1}, \ldots, b_k)$ $(i = \overline{1, k-1})$, полагая $b_{(k)} = b^{(0)} = b$. Зададим априорные вероятности $p(b)$, из которых легко получаем вероятности $p(b^{(i)})$. Рассмотрим условно плотность вероятности $f_b(x)$ и усредненные плотности вероятности случайной величины ξ:

$$f_0(x) = \sum f_b(x)\, p(b); \quad f_i(x) = \sum_{b^{(i)}} f_b(x) \frac{p(b)}{p(b_{(i)})} \quad (i = \overline{1, k-1}); \quad f_k(x) = f_b(x).$$

Можно показать, что если $b_{(i)} = b^*_{(i)}$ является истинным значением набора компонент b, то условная плотность вероятности случайной величины ξ имеет вид

$$f(x) = \begin{cases} f_i(x), & \text{если } b_{(i)} = b^*_{(i)}, \\ [f_0(x) - p(b^*_{(i)})\, f_i(x)]/[1 - p(b^*_{(i)})], & \text{если } b_{(i)} \neq b^*_{(i)} \ (i = \overline{1, k}). \end{cases}$$

Поэтому из-за структуры $f(x)$, как и в случае статистического декодирования, вместо выбора между $\prod_{j=1}^{i} N_j$ гипотезами $H^*_{b_{(i)}}(b_{(i)} = b^*_{(i)})$ можно ставить задачу о $\prod_{j=1}^{i} N_j$ выборах между двумя гипотезами $H_1^{(i)}(b_{(i)} = b^*_{(i)})$ и $H_0^{(i)}(b_{(i)} \neq b^*_{(i)})$.

При асимптотически растущем объеме выборки n_i вероятность $\beta_i = 1 - \delta_i$ ошибки второго рода не зависит от вероятности $\alpha_i = \gamma_i$ ошибки первого рода (см. гл. 3) и асимптотически равна

$$\beta_i \approx e^{-d_i n_i},$$

где

$$d_i = I_i(2:1) = \int_{-\infty}^{\infty} \left[\ln \frac{f_i(x)}{f_0(x)}\right] f_i(x)\, dx$$

являются числами Кулебака-Леблера [21] (их мы будем называть силами соответствующих критериев).

Далее найдем условия, при которых с точки зрения теории оптимальной реализации целесообразно проведение многоканального решения задачи выбора между k гипотезами $H_0^{(k)}$ и $H_1^{(k)}$.

Если при усреднении по компонентам параметра $b^{(1)}$ задача выбора между двумя гипотезами абсолютно надежным КУ $P(T)$-осуществима, то $P(T)$ и T вычисляются из следующих соотношений [34]

$$P(T) = 1 - \beta \approx \Phi[\Phi^{-1}(\alpha) + \sqrt{2d_1 n}] \approx 1 - e^{-d_1 n} = 1 - e^{-d_1 v T},$$

[1] Например, с помощью статистической оценки параметров.

где $n = vT$ ($T = n/v$) — число испытаний; d_1 — параметр силы критерия; v — быстродействие визуальной оценки результата; α и β — вероятности ошибок 1-го и 2-го родов.

Итак, решение

$$(1 - e^{-d_1 vT})\text{-осуществимо.} \qquad \text{(II.10)}$$

При многоканальном решении той же задачи вариантность b оказывается равной $N = \prod_{i=1}^{k} N_i$ (равна числу каналов), при этом используется машина с быстродействием $V \gg v$ и силой критерия на истинном варианте $d_k \gg d_1$. В этой ситуации реализация анализируемого процесса (выборка) как бы размножается в N экземплярах, соответствующих N вариантам b. Лишь в одном экземпляре, соответствующем истинному значению $b = b^*$, как отмечалось ранее, имеет место гипотеза $H_1^{(k)}$ с параметром силы критерия $d_k \gg d_1$. В остальных экземплярах имеет место гипотеза $H_0^{(k)}$.

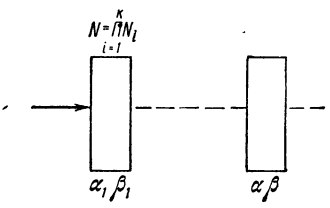


Рис. II.5. Последовательный во времени анализ компонент

Произведем выбор между двумя гипотезами $H_0^{(k)}$ и $H_1^{(k)}$ для каждого из N вариантов с заданными вероятностями ошибок 1-го и 2-го родов α_k и β_k соответственно. Тогда случайное число μ_0 ложных вариантов, принятых за истинные, будет иметь биномиальное распределение

$$P(\mu_0 = m) = P_{N-1, \alpha_k}(m) = C_{N-1}^{m} \alpha_k^m (1 - \alpha_k)^{N-1-m} \qquad \text{(II.11)}$$

и обнаружение истинного варианта будет иметь вероятность

$$1 - \beta_k = \Phi[\Phi^{-1}(\alpha_k) + \sqrt{2d_k n}] \approx 1 - e^{-d_k n}.$$

При этом окончательный визуальный просмотр μ_0 или $\mu_0 + 1$ (в зависимости от сохранения истинного варианта) должен проводиться с малым визуальным быстродействием, благодаря чему можно просматривать выборки меньшего объема $n_k < n$ с существенно более мощными (практически достоверными) критериями. Следует отметить, что непрерывный по времени статистический анализ (отбор) N вариантов объема n будет приводить через случайные интервалы времени к «кандидатам» на истинные варианты. Накапливая последние в некотором числе (в устройстве задержки), достаточном для выполнения условий центральной предельной теоремы, можно затем выпускать эти варианты через регулярные промежутки времени, равные среднему времени $1/\alpha_k$ между поступлениями (рис. II.5). Эта операция приведет к сравнительно небольшой задержке. Учитывая, что при больших $N\alpha_k > 100$ распределение (II.11) вырождается в δ-функцию, можно считать эти интервалы равными T/α_k, а их число μ_0 равным $N\alpha_k$.

Для дальнейшего изложения существенно то обстоятельство, что между поступлениями очередных вариантов может осуществляться их визуальный просмотр на медленной визуальной скорости и, таким образом, указанные оба отбора могут быть осуществлены одновременно.

Рассмотренные выше соображения приводят к следующим соотношениям:

$$VT = nN; \quad vT = n_k \mu_0; \quad \mu_0 = N\alpha; \quad 1 - \beta = (1 - \beta_k) \cdot 1 \qquad \text{(II.12)}$$

или

$$VT = nN = \alpha_k n_k N \frac{V}{v};\ \alpha_k = \frac{v}{V} \cdot \frac{n}{n_k};\ n = \frac{V}{N} T; \qquad \text{(II.12')}$$

$$1 - \beta = 1 - e^{-d_k \frac{V}{N} T}.$$

Таким образом, в рассматриваемом случае решение задачи

$$\left(1 - e^{-d_k \frac{V}{N} T}\right)\text{-осуществимо.} \qquad \text{(II.13)}$$

Сравнивая (II.10) с (II.13), можно заключить, что многоканальное решение целесообразно при условии $d_k (V/N) > d_1 v$, т. е. при быстродействии V,

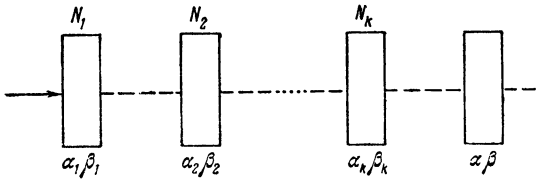


Рис. II.6. Одновременный во времени анализ компонент

удовлетворяющем соотношению

$$V > N \frac{d_1}{d_k} v. \qquad \text{(II.14)}$$

Следует заметить, что с ростом числа измерений k вектора b его вариантность растет по экспоненциальному закону $N > (\min\limits_{1 \leqslant i \leqslant k} N_i)^k$ и поэтому условие (II.14) требует быстродействия V, как правило, выходящего за пределы имеющегося уровня техники.

II.3.3. *Одновременный статистический анализ компонент.* Можно существенно уменьшить экспоненциальный рост вариантности N способами, позволяющими производить одновременный анализ сразу всех компонент $b = (b_1, \ldots, b_i, \ldots, b_k)$. В самом деле, правильная фиксация компонент b

$$(b_1),\ (b_1, b_2), \ldots, (b_1, b_2, \ldots, b_i), \ldots, (b_1, b_2, \ldots, b_k)$$

приводит к задачам выбора между двумя все более «далекими» гипотезами $H_0^{(i)}$ и $H_1^{(i)}$ с силами критерия

$$d_1 \leqslant d_2 \leqslant \ldots \leqslant d_i \leqslant \ldots \leqslant d_k, \qquad \text{(II.15)}$$

вероятностями ошибок 1-го и 2-го родов

$$(\alpha_1, \beta_1),\ (\alpha_2, \beta_2), \ldots, (\alpha_i, \beta_i), \ldots, (\alpha_k, \beta_k)$$

и соответственно объемами выборок

$$n = n_0 \geqslant n_1 \geqslant n_2 \geqslant \ldots \geqslant n_i \geqslant \ldots \geqslant n_{k-1} \geqslant n_k. \qquad \text{(II.16)}$$

Легко обобщить предыдущие рассмотрения так, чтобы иметь одновременный анализ всех вариантов компонент b (все варианты последующей компоненты успевают анализироваться между поступлениями «кандидатов» на правильную фиксацию всех предыдущих компонент) (рис. II.6). Соот-

ношения между параметрами задачи, обеспечивающие такую возможность, обобщают соотношения (II.12) и имеют вид

$$VT = nN_1 = \alpha_1 n_1 N_1 N_2 = \alpha_1\alpha_2 n_2 N_1 N_2 N_3 = \ldots \qquad \text{(II.17)}$$
$$\ldots = \alpha_1 \ldots \alpha_{k-1} n_{k-1} N_1 \ldots N_k = \alpha_1 \ldots \alpha_k n_k N_1 \ldots N_k \ (V/v),$$

откуда

$$n = (V/N_1)T;$$
$$\left.\alpha_1 = \frac{1}{N_2}\frac{n}{n_1};\ \alpha_2 = \frac{1}{N_3}\frac{n_1}{n_2};\ \ldots;\ \alpha_{k-1} = \frac{1}{N_k}\frac{n_{k-2}}{n_{k-1}};\ \alpha_k = \frac{v}{V}\frac{n_{k-1}}{n}\right\} \qquad \text{(II.18)}$$
$$1 - \beta_i = 1 - e^{-d_i n_{i-1}} \ (i = \overline{1, k};\ n_0 = n).$$

Используя формулу Буля, получим

$$1 - \beta > \sum_{i=1}^{k} (1 - \beta_i) - (k - 1) = 1 - \sum_{i=1}^{k} \beta_i. \qquad \text{(II.19)}$$

Оценим далее величину функции осуществимости $P(T) = 1 - \beta$ для рассматриваемого случая. Если k не растет экспоненциально с ростом T (а это имеет место в известных практических приложениях), то из (II.19) имеем

$$1 - \beta > 1 - \sum_{i=1}^{k} \beta_i > 1 - k \max_{1 \leqslant i \leqslant k} \beta_i \approx 1 - e^{-\min\limits_{1 \leqslant i \leqslant k} d_i n_{i-1}}. \qquad \text{(II.20)}$$

Из соотношения (II.20) следует, что объемы выборок $n_i \ (i = \overline{1, k})$, находящиеся в нашем распоряжении, нужно выбирать так, чтобы $d_1 n = d_2 n_1 = \ldots = d_k n_{k-1} = \text{const}$, что всегда возможно из-за соотношений (II.15) и (II.16). В этом случае $1 - \beta \gtrsim 1 - k\beta_1 \approx 1 - \beta_1 = 1 - e^{-d_1 n}$.

Используя (II.18), получим, что в рассматриваемом случае решение задачи

$$\left(1 - e^{-d_1 \frac{V}{N_1} T}\right)\text{-осуществимо.} \qquad \text{(II.21)}$$

Теперь можно оценить целесообразность одновременного анализа по сравнению с последовательным анализом во времени компонент b. Для этого нужно, чтобы

$$\frac{d_1}{N_1} > \frac{d_k}{N}. \qquad \text{(II.22)}$$

Неравенство (II.22), как правило, имеет место на практике, несмотря на то, что $d_1 \leqslant d_k$, поскольку $N \gg N_1$. Заметим, что при одновременном анализе компонент b свободные параметры процедур выбора между гипотезами α_i используются только для удовлетворения требованию (II.17) согласования потоков вариантов между последовательными критериями (см. рис. II.6). Таким образом, внешняя эффективность f_1 работы КУ, связанная с большой вариантностью N, при оценке параметров имеет, согласно (II.10), (II.13) и (II.21), вид

$$f_1 = \max\left\{d_1 v,\ d_k \frac{V}{N},\ d_1 \frac{V}{N_1}\right\}, \qquad \text{(II.23)}$$

где v и V $(v \ll V)$ — визуальное и машинное быстродействия; N и N_1 $(N \gg N_1)$ — вариантности в целом b и его первой компоненты; d_1 и d_k $(d_1 \leqslant d_k)$ — силы статистических критериев, при правильной фиксации одной компоненты и всех компонент.

Последние параметры выражаются через величины, имеющие простой статистический смысл. Так, для параметрического случая близких гипотез [34]

$$d = \Delta_1^2 I_2(a_0),$$

где Δ_1 — разность между гипотетическими значениями параметра, а $I_2(a_0)$ — «информация» Фишера. Как правило, на практике имеем

$$d_1 v < d_k \frac{V}{N} \ll d_1 \frac{V}{N_1}$$

и поэтому из (II.23) получим

$$f_1 = d_1 \frac{V}{N_1}.$$

Другими словами, эффективность работы КУ прямо пропорциональна силе d_1 первого критерия и быстродействию V и обратно пропорциональна вариантности (числа значений) N_1 первой компоненты оцениваемого вектора b.

II.3.4. *Выводы.* 1. Современная постановка задачи оценки k-мерного параметра b с помощью машин с быстродействием V сводится к анализу большого числа вариантов $N = \prod_{i=1}^{k} N_i$ значений параметра $b = (b_1, \ldots, b_i, \ldots, b_k)$.

2. Эта задача допускает существенное упрощение одновременным статистическим анализом компонент b.

3. Внешняя эффективность f_1 работы КУ при этом имеет вид

$$f_1 = d_1 \frac{V}{N_1},$$

где d_1 — сила однокомпонентного статистического критерия; N_1 — вариантность (число значений) первой из анализируемых компонент b; V — быстродействие КУ.

§ II.4. Учет ограниченности объема памяти кибернетического устройства

II.4.1. *Введение.* Во многих задачах различения и поиска сигналов на фоне шумов, поддержания нормального функционирования сложных кибернетических систем и других задачах возникает следующая ситуация.

Имеется N элементов фазового пространства (N каналов, N частей системы и т. д.), являющихся одновременно действующими источниками информации, и известны различные предположения об их индивидуальной и совместной природе. Для распознания их характера предназначается КУ с объемом памяти, ограниченным M ячейками. Содержательными здесь оказываются задачи, когда $M \ll N$ (это соотношение и имеет основной практический интерес).

Если $M \geqslant N$, то каждый элемент можно жестко закоммутировать с одной из ячеек и, накопив в ней достаточное количество данных, вынести статистическое решение о природе элемента. В случае $M \ll N$ этого сделать нельзя, и поэтому обычные статистические методы здесь непригодны. Впервые такого рода задачи рассматривались в [34], где введенное понятие информационного потока можно интерпретировать для рассматриваемого случая как N элементов фазового пространства с шумовым или сигнальным характером флуктуаций в них.

В настоящее время нет общей оптимальной постановки задачи, однако можно определить следующие естественные функции КУ при решении такого рода задач (рис. II.7):

а) преобразование многомерного информационного потока с одновременными компонентами в одномерный векторный процесс с последовательными во времени компонентами, что достигается последовательным во времени просмотром КУ элементов фазового пространства;

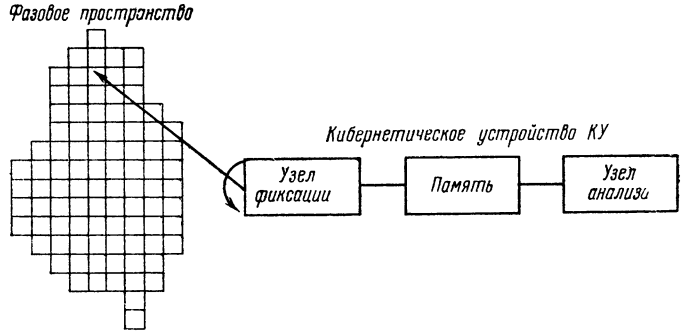


Рис. II.7. Функции кибернетического устройства

б) фиксация «подозрительных» элементов, требующих дополнительного анализа (узел фиксации КУ);

в) накопление во времени данных за несколько просмотров от фиксированных элементов в соответствующих ячейках памяти (память КУ);

г) анализ результатов накопления и окончательное решение о шумовом или сигнальном характере элемента (узел анализа КУ).

В случае подвижных образов КУ должно существенно быстрее осуществлять периодический просмотр элементов фазового пространства, а узел фиксации должен по какому-либо правилу выделять «подозрительные» элементы. Эта операция преобразует информационный поток в поток вызовов в смысле теории массового обслуживания. После этого устанавливается временная жесткая связь между некоторыми элементами фазового пространства и ячейками памяти КУ. В них происходит накопление данных за несколько просмотров вплоть до вынесения окончательного решения в узле анализа о шумовом или сигнальном характере элементов. После вынесения решения (вообще говоря, за случайное время) ячейка очищается и может воспринимать новые данные. Это создает динамический режим загрузки памяти КУ.

Можно поставить задачу отыскания оптимальных процедур, которые должны различать образы при ограниченном времени наблюдения, заданных вероятностях ошибок и заданной близкой к нулю вероятности переполнения памяти КУ (переполнение памяти приводит к сбою в работе). Однако точная формулировка такой задачи еще не определена (не ясна формулировка оптимальности критерия «подозрительности» элементов; см. п.II.4.2). Поэтому ниже решаются задачи оптимизации общей процедуры отдельно в узле фиксации и в узле анализа (см. рис. II.7). Вообще говоря, оптимальность «в частях» может не быть оптимальностью «в целом».

II.4.2. *Постановка задачи.* Информационный поток $I(t)$ можно интерпретировать [34] как N элементов некоторого фазового пространства, в каждом из которых можно измерить величину случайного процесса $I_i(t) = \xi_i(t)$ $(i = \overline{1, N})$. Пусть некоторая часть X_i значений $\xi_i(t)$ рассматривается как значения «подозрительные», а остальные как значения обычные.

Предположим, что поступил вызов из i-го элемента, если значение $\xi_i(t) = x_i(t) \in X_i$. Например, можно определить множество значений X_i, как интервал $(X_i\ \infty)$. После вызова из i-го элемента в момент t вызов подключается к одной из свободных ячеек памяти и происходит накопление значений $\xi_i(t)$ в течение некоторого, вообще говоря, случайного времени ν_i (время обслуживания) до вынесения определенных статистических заключений о характере элемента, после чего ячейка очищается.

Далее удобна дискретизация по времени рассмотренной непрерывной схемы. Произведем дискретизацию по времени с интервалом $\Delta t = h > 0$, идущим на один просмотр, т. е. рассмотрим моменты $0, h, 2h, \ldots, s = lh, \ldots, t = kh, \ldots$, отождествляя моменты вызовов и моменты начала и окончания обслуживания, например, с правыми концами интервалов дискретизации, в которые они попали.

Пусть плотность вероятности «напряжения» $\xi_i(t)$ в i-м элементе, в зависимости от его шумового или сигнального характера, имеет вид $f^{(i)}_{a_\varepsilon}(x)$, где $\varepsilon = 0$ в шумовом и $\varepsilon = 1$ в сигнальном случае $(a_0 \neq a_1)$. Ясно, что в этом случае в каждом просмотре условные вероятности того, что i-й элемент окажется «подозрительным», равны

$$p^{(i)}_\varepsilon = \int_{X_i}^{\infty} f^{(i)}_{a_\varepsilon}(x)\,dx \quad (\varepsilon = 0,1).$$

Пусть шумовые элементы имеют в момент t индексы $\mathscr{E}^{(t)}_0 = (i_1, \ldots, i_{N_0})$, а сигнальные $\mathscr{E}^{(t)}_1 (i_{N_0+1}, \ldots, i_N)$; тогда, как легко показать, среднее значение числа вызовов $\varkappa(t)$ в момент t равно

$$\mathrm{E}\,\varkappa(t) = \sum_{i \in \mathscr{E}^{(t)}_0} p^{(i)}_0 + \sum_{i \in \mathscr{E}^{(t)}_1} p^{(i)}_1. \tag{II.24}$$

Вместе с тем случайное число ρ_i просмотров, необходимое для того чтобы сигнальный элемент оказался «подозрительным» (число просмотров, при котором *впервые* будет превзойден порог X_i), имеет геометрическое распределение

$$\mathscr{P}(\rho_i = r) = (1 - p^{(1)}_1)^{r-1} p^{(i)}_1 \approx p^{(i)}_1\, e^{-p^{(i)}_1 r}$$

со средним

$$\mathrm{E}\,\rho_i \approx \frac{1}{p^{(i)}_1}. \tag{II.25}$$

Соотношения (II.24) и (II.25) обнаруживают противоположные требования к назначению порогов X_i. В самом деле, увеличением X_i, приводящим к уменьшению $p^{(i)}_\varepsilon$, можно добиться согласно (II.24) столь малого значения $\mathrm{E}\varkappa(t) \ll M$, что вероятность переполнения памяти (см. п. II.4.3) может быть сделана сколь угодно малой. Но при этом согласно (II.25) будет возрастать средняя задержка обнаружения сигнального элемента. Можно дать точную экстремальную постановку и решение задачи выбора X_i. Однако здесь мы этого делать не будем. Ниже будет решена задача оценки загрузки памяти в рассматриваемом случае при произвольных, но фиксированных X_i (или $p^{(i)}_\varepsilon$), состоящая в следующем.

Пусть заданы вероятности вызовов $p^{(i)}_\varepsilon$ при одном просмотре и множества сигнальных и шумовых элементов $\mathscr{E}^{(t)}_1$ и $\mathscr{E}^{(t)}_0$ в момент t, соответ-

ственно, а также заданы распределения времени ν_t обслуживания каждого из вызовов (время для вынесения статистического решения) $\mathcal{P}(\nu_t < n) = F^{(t)}(n)$. Требуется найти распределение случайного числа $\mu(t)$ ячеек памяти в момент t. Ниже приводится решение этой неоднородной задачи ($p_\varepsilon^{(t)}$ и $F^{(t)}(n)$ в общем случае зависят от i) в стационарном случае, когда $t \to \infty$. До этого решалась лишь однородная задача [34]. Необходимость более общей постановки задачи диктуется рядом новых ее приложений.

II.4.3. *Оценка загрузки памяти в неоднородном случае.* Пусть число ячеек памяти $M \to \infty$, число элементов фазового пространства N и момент дискретного времени $t = kh$. Можно показать, что число $\mu(t)$ занятых ячеек в момент $t = kh$ имеет вид

$$\mu(t) = \sum_{s=h}^{t} \sum_{i \in R_s'} \sum_{j=0}^{\varepsilon_i(s)} \varepsilon_{ij}(s, t), \tag{II.26}$$

где R_s' — совокупность элементов, незакоммутированных с ячейками памяти в момент s; $\varepsilon_i(s)$ — случайная функция, определенная на элементах R_s', равная нулю, если элемент неподозрителен, или равная единице в противном случае; $\varepsilon_{ij}(s, t)$ — случайная функция, определенная на подозрительном элементе, возникшем в момент s, равная нулю, если он обслужен (закоммутирован со своей ячейкой) до момента t, или равная единице в противном случае; $\varepsilon_{i0}(s, t)$ — случайная функция, тождественно равная нулю.

Упростим соотношение (II.26). В самом деле, число элементов в R_s', как легко видеть, равно $N - \mu(s - h)$ (ясно, что $\mu(0) = 0$ и в начальный момент h число элементов в R_s' равно N).

В дальнейшем изложении мы будем рассматривать лишь случай малой вероятности переполнения памяти, для чего необходимо иметь с большой вероятностью соотношение

$$\mu(t) < M \ll N.$$

Таким образом, с большой вероятностью можно полагать, что

$$N - \mu(s - h) > N - M \approx N.$$

Поэтому вместо соотношения (II.26) можно полагать

$$\mu(t) = \sum_{s=h}^{t} \sum_{i=1}^{N} \sum_{j=0}^{\varepsilon_i(s)} \varepsilon_{ij}(s, t), \tag{II.27}$$

где под знаком сумм стоят независимые между собой и от $\varepsilon_i(s)$ случайные величины $\varepsilon_{ij}(s, t)$.

Таким образом, последняя сумма (II.27) представляет собой вырожденный случай суммы случайного числа случайных слагаемых, и к ней применима теория, изложенная в гл. 2. В однородном случае из (II.27) имеем [34]

$$\mu(t) = \sum_{s=h}^{t} \sum_{k_s=1}^{\varkappa(s)} \varepsilon_{k_s}(s, t).$$

Найдем распределение случайной величины $\mu(t)$. Как правило, на практике основная масса элементов имеет шумовой характер, т. е. $N_0 \approx N$. Поэтому

будем вести расчет загрузки памяти для случая, когда все элементы являются шумовыми. Введем производящие функции случайных величин

$$\varepsilon_i(s) = \begin{Bmatrix} 0 & 1 \\ 1-p_0^{(1)} & p_0^{(1)} \end{Bmatrix}; \quad \varepsilon_{i0}(s,t) \equiv 0; \quad \varepsilon_{i1} = \begin{Bmatrix} 0 & 1 \\ F^{(i)}(t-s) & 1-F^{(i)}(t-s) \end{Bmatrix};$$

$$h_s^{(i)}(x) = 1 + p_0^{(i)}(x-1); \quad h_{s,t,0}^{(i)}(x) \equiv 1; \quad h_{s,t,1}^{(i)}(x) = 1 + [1 - F^{(1)}(t-s)](x-1),$$

соответственно.

Найдем производящую функцию $G_t(x)$ случайной величины $\mu(t)$. Для этого заметим, что производящая функция $g_{s,t}^{(i)}(x)$ внутренней суммы (II.27) имеет согласно (1.47) вид

$$g_{s,t}^{(i)}(x) = (1-p_0^{(i)}) h_{s,t,0}^{(i)}(x) + p_0^{(i)} h_{s,t,1}^{(i)}(x) = 1 - p_0^{(i)}[1 - F^{(i)}(t-s)](x-1).$$

Из-за независимости внутренних сумм (II.27) между собой имеем

$$G_t(x) = \prod_{s=h}^{t} \prod_{i=1}^{N} \{1 + p_0^{(i)}[1 - F^{(i)}(t-s)](x-1)\}.$$

Отсюда дифференцированием по x получим

$$\mathrm{E}\mu(t) = G_t'(1) = \sum_{s=h}^{t} \sum_{i=1}^{N} p_0^{(i)}[1 - F^{(i)}(t-s)] = \sum_{i=1}^{N} p_0^{(i)} \sum_{s=h}^{t} [1 - F^{(i)}(t-s)]$$

и

$$\mathrm{D}\mu(t) = G_t''(1) + G_t'(1) - (G_t'(1))^2 = \mathrm{E}\mu(t) - \sum_{i=1}^{N} (p_0^{(i)})^2 \sum_{s=h}^{N} [1 - F^{(i)}(t-s)]^2 \leqslant \mathrm{E}\mu(t).$$

Рассмотрим предельный случай больших $N \to \infty$, когда $p_0^{(i)} = \dfrac{c_0^{(i)}}{N} \to 0$ малы. Легко показать, что в этом случае

$$G_t(x) = \exp\left\{ \sum_{i=1}^{N} p_0^{(i)} \sum_{s=h}^{t} [1 - F^{(i)}(t-s)](x-1) \right\} + O(N^{-2}). \qquad \text{(II.28)}$$

Из (II.28) следует, что предельным распределением для $\mu(t)$ при $N \to \infty$ является распределение Пуассона

$$\mathcal{P}(\mu(t) = m) = \frac{\Lambda^m(t)}{m!} e^{-\Lambda(t)},$$

определяемое единственным параметром

$$\Lambda(t) = \mathrm{E}\mu(t) = \mathrm{D}\mu(t) = \sum_{i=1}^{N} p_0^{(i)} \sum_{s=h}^{t} [1 - F^{(i)}(t-s)]. \qquad \text{(II.29)}$$

Из структуры $\Lambda(t)$, определяемой соотношением (II.29), следует, что с ростом t $\Lambda(t)$ является неубывающей функцией. Определим ее предел при $t \to \infty$ (стационарный случай). Заметим, что

$$\lim_{t\to\infty} \sum_{s=h}^{\infty} [1 - F^{(i)}(t-s)] = \sum_{s=h}^{\infty} [1 - F^{(i)}(s)] = \sum_{s=h}^{\infty} s\Delta F^{(i)}(s) = \mathrm{E}_0 \nu_i, \qquad \text{(II.30)}$$

где $E_0 \nu_i$ — среднее число просмотров, доставляющих данные в ячейку памяти, для вынесения решения о шумовом или сигнальном характере i-го элемента, если он имеет шумовой характер.

Итак, с учетом соотношения (II.30) заключаем из (II.29), что при $t \to \infty$ $\Lambda(t)$ стремится слева к выражению

$$\Lambda = \Lambda(\infty) = \sum_{i=1}^{N} p_0^{(i)} E_0 \nu_i. \qquad \text{(II.31)}$$

II.4.4. *Оценка вероятности непереполнения памяти.* Рассмотрение стационарного случая для оценки вероятности P_1 непереполнения памяти объема M из-за максимального значения $\Lambda(\infty)$ приводит к оценке этой вероятности снизу. Проведем такую оценку.

В самом деле, при больших $\Lambda = \Lambda(\infty)$, (а это как раз имеет место на практике) случайное число $\mu = \mu(\infty)$ занятых ячеек памяти, распределенное по Пуассону, переходит в (Λ, Λ)-нормально распределенную величину. Поэтому вероятность P_1 непереполнения памяти в стационарном режиме при одном просмотре равна

$$P_1 = \mathcal{P}(\mu < M = \Lambda + u_p \sqrt{\Lambda}) = \Phi(u_p) = p(M, \Lambda), \qquad \text{(II.32)}$$

где u_p—p-квантиль нормального распределения (при $u_0 = 5$ имеем $p \approx 1 - 10^{-6}$).

Так как случайные величины $\mu(t)$ при различных t являются зависимыми, то, используя формулу Буля, можно оценить снизу вероятность P_T непереполнения памяти за $n = T/h$ просмотров. Эта оценка имеет вид

$$P_T > 1 - \frac{T}{h}(1 - p(M, \Lambda)) \approx e^{-\frac{1}{h}(1-p(M,\Lambda))T}, \qquad \text{(II.33)}$$

$$\text{при } \frac{T}{h}(1-p(M, \Lambda)) \leqslant 1.$$

Ясно, что оценка (II.33) тем более является нижней оценкой вероятности непереполнения памяти в нестационарном режиме и что переполнение памяти КУ на некоторое время может привести к расстройству его работы и невыполнению поставленной перед ним цели. В некотором смысле переполнение памяти КУ можно рассматривать как сбой в его работе. Поэтому вероятность осуществимости $P(T)$ (§ II.2) должна дополнительно умножаться на P_T или на нижнюю оценку P_T (II.33). Но это эквивалентно переходу от параметра f_2 внутренней эффективности КУ, учитывающему лишь его надежность, к параметру внутренней эффективности f_2' КУ, учитывающему еще и возможность переполнения памяти КУ. Легко показать что

$$f_2' = \frac{f_2}{1 + r f_2},$$

где

$$r = \frac{1 - p(M, \Lambda)}{h}, \qquad \text{(II.34)}$$

причем

$$p(M, \Lambda) = \Phi\left[\frac{M - \Lambda}{\sqrt{\Lambda}}\right], \qquad \text{(II.35)}$$

а Λ определяется соотношением (II.31).

Перейдем теперь к выводу условий, налагаемых на параметры задачи в связи с необходимостью непереполнения ограниченной памяти. Из соот-

ношения (II.32) следует, что при $u_p \gg 3$ для непереполнения памяти необходимо, чтобы

$$\Lambda \approx M - u_p \sqrt{M},$$

т. е. чтобы при $M \to \infty$ $\Lambda \approx M$.

Учитывая соотношение (II.31), получим следующее основное ограничение, накладываемое на параметры $p_0^{(i)}$ и $E_0 \nu_i$, связанное с ограниченностью объема памяти

$$\sum_{i=1}^{N} p_0^{(i)} E_0 \nu_i = M. \tag{II.36}$$

Кроме соотношения (II.36), важнейшей характеристикой является время γ_i задержки решения о сигнальном характере i-го элемента, когда он на самом деле имеет сигнальный характер. В этом случае общая случайная задержка γ_i имеет вид

$$\gamma_i = \rho_i + \nu_i, \tag{II.37}$$

где ρ_i — величина задержки вызова подозрительности от сигнального элемента; ν_i — число испытаний, необходимых для вынесения окончательного решения о сигнальном характере сигнального элемента.

В соответствии с (II.37) и (II.25) имеем

$$E\gamma_i = \frac{1}{p_1^{(i)}} + E_1 \nu_i. \tag{II.38}$$

Пусть $\mathcal{P}_i$ — априорная вероятность сигнального характера i-го элемента; тогда среднее время $E\gamma$ задержки окончательного решения о сигнальном элементе будет иметь вид

$$E\gamma = \sum_{i=1}^{N} \mathcal{P}_i E\gamma_i = \sum_{i=1}^{N} \mathcal{P}_i / p_1^{(i)} + \sum_{i=1}^{N} \mathcal{P}_i E_1 \nu_i.$$

II.4.5. *Оценка загрузки памяти в однородном случае.* Если распределение вероятностей «напряжений» во всех элементах фазового пространства не зависит от элементов $f_{a_\varepsilon}^{(i)}(x) = f_{a_\varepsilon}(x)$ (однородный случай), то соотношения (II.36) и (II.38) упрощаются и имеют вид:

$$N p_0 E_0 \nu = M \tag{II.39}$$

и

$$E\gamma = \frac{1}{p_1} + E_1 \nu. \tag{II.40}$$

Эти соотношения использовались в [34] для учета ограниченного объема памяти КУ воспроизведения образов. В рассматриваемом однородном случае можно проанализировать не только стационарный ($t \to \infty$), но и нестационарный случай ($t = \text{const}$) загрузки памяти КУ. При этом возможен переход к непрерывному времени. В результате получаются соотношения, обобщающие некоторые известные соотношения теории массового обслуживания.

Проанализируем соотношения (II.39) и (II.40) [качественно результаты анализа касаются и более общих соотношений (II.36) и (II.38)].

В самом деле, при фиксированных значениях параметров a_0 и a_1 (отношение сигнал/шум), а также параметров N и M для удовлетворения

соотношения (II.39) нужно выбрать первичный порог X дискретизации так, чтобы

$$p_0 = \frac{M}{N}\frac{1}{\mathrm{E}_0\nu}.$$

Малое отношение M/N и малое отношение сигнал/шум ($a_0 \approx a_1$) приводит к малым $p_0 \approx p_1$. В результате этого основную роль в соотношении (II.40) начинает играть первое слагаемое порядка $(N/M)\,\mathrm{E}_0\nu$. Таким образом, в задачах различения информационных потоков основной вклад в среднюю задержку $\mathrm{E}\gamma$ решения о сигнальном характере сигнального элемента вносит величина отношения M/N. В обычных же задачах обнаружения ($M \to \infty$) основное значение имеет отношение сигнал/шум.

II.4.6. *Выводы.* 1. Учет ограниченности объема памяти КУ, осуществляющего статистическую процедуру различения информационных потоков, выходит за рамки существующих оптимальных задач математической статистики.

2. Возникающие при этом задачи родственны по своим постановкам задачам теории массового обслуживания, и их решение сводится к оценке распределения числа занятых ячеек памяти с помощью производящих функций.

3. Учет возможности переполнения памяти КУ, рассматриваемой как сбой КУ, приводит к уменьшению внутренней эффективности f_2 КУ (среднего времени между сбоями) в $1+rf_2$ раз, где коэффициент r определяется соотношением (II.34), (II.35) и (II.31).

4. Основной вклад в среднее значение задержки решения о сигнальном характере элемента фазового пространства вносит отношение N/M, где N — число элементов фазового пространства, а M — объем памяти КУ.

ОСНОВНЫЕ ПРИНЯТЫЕ ОБОЗНАЧЕНИЯ

$\alpha = \overline{1, a}$ — входные символы (обозначение $\alpha = \overline{1, a}$ означает $\alpha = 1,2, \ldots, a$);

$\beta = \overline{1, b}$ — выходные символы (вообще говоря, $a \neq b$);

$x = (\alpha_1, \ldots, \alpha_n)$ — входные слова длины n (R — их совокупность);

$y = (\beta_1, \ldots, \beta_n)$ — выходные слова длины n (R^* — их совокупность);

$\overline{p} = (p_\alpha)$—вектор абсолютных вероятностей $\alpha \left(\sum_{\alpha=1}^{a} p_\alpha = 1 \right)$;

$\mathbf{p} = \| p_\alpha^\beta \| = (\overline{p_\alpha})$— матрица условных вероятностей β;

$\overline{q} = (q_\beta)$ — вектор абсолютных вероятностей $\beta \left(\sum_{\beta=1}^{b} q_\beta = 1;\ \overline{p}\,\mathbf{p} = \overline{q} \right)$;

$\overline{m} = (m_\alpha)$ — числа символов α во входном слове x;

$\sum_{\alpha=1}^{a} m_\alpha = n,\ m_\alpha = n\mu_\alpha,\ \sum_{\alpha=1} \mu_\alpha = 1,\ \overline{\mu} = (\mu_\alpha),\ \mu_\alpha \geqslant 0;$

$\mathbf{m} = \mathbf{m}(x, y) = \| m_{\alpha\beta} \|$ — матричное «расстояние» между x и y, где $m_{\alpha\beta}$ — число пар (α, β) в паре (x, y);

$\sum_{\beta=1}^{b} m_{\alpha\beta} = m_\alpha,\ m_{\alpha\beta} = n\mu_\alpha\mu_\alpha^\beta,\ \sum_{\beta=1}^{b} \mu_\alpha^\beta = 1\ (\alpha = \overline{1, a}),\ \mu = \| \mu_\alpha^\beta \| = (\overline{\mu_\alpha});$

$\overline{m^*} = (m_\beta^*)$ — числа символов β в выходном слове y;

$\sum_{=1}^{b} m_\beta^* = n,\ m_\beta^* = n\nu_\beta,\ \sum_{\alpha=1}^{a} \nu_\beta = 1,\ \overline{\nu} = (\nu_\beta) = \overline{\ }$;

$N(\mathcal{E})$ — число элементов множества $\mathcal{E}$ (события);

$\mathcal{P}(\mathcal{E})$ — вероятность появления какого-либо из элементов $\mathcal{E}$;

$\mathcal{P}(\mathcal{E}/\mathcal{E}')$ — та же вероятность при условии, что имеет место событие $\mathcal{E}'$);

$\mathcal{E}_\sigma^{\overline{m}} \subset R$ — множество x-ов, определенных в моменты времени $\sigma = (1, 2, \ldots, n)$ с фиксированным $\overline{m}$;

$\mathcal{E}_x^{\mathbf{m}} \subset (x, R^*)$ — множество y-ов с фиксированным расстоянием $\mathbf{m}(x, y)$ от x, равным $\mathbf{m}$;

$\mathcal{E}_\sigma^{\overline{m}^*} \subset R^*$ — множество y-ов, определенных в моменты времени $\sigma = (1, 2, \ldots, n)$ с фиксированным $\overline{m}^*$;

$C_n^{\overline{m}} = n!/\prod_{\alpha=1}^{a} m_\alpha!$ — полиномиальный коэффициент;

$\mathbf{u}^{\mathbf{m}} = \prod_{\alpha\beta} u_{\alpha\beta}^{m_{\alpha\beta}}$;

$h(\bar{u}, \bar{v}) = -\sum_{\beta=1}^{b} u_\beta \ln v_\beta$ — скалярная функция, определенная на паре векторов $\bar{u} = (u_\beta)$ и $\bar{v} = (v_\beta)$; [здесь $\sum_{\beta=1}^{b} u_\beta = \sum_{\beta=1}^{b} v_\beta = 1$; $u_\beta, v_\beta \geqslant 0$; $h(\bar{u},\bar{v}) \geqslant h(\bar{v}, \bar{v}) = h(\bar{v})$];

$k_{\bar{p}\bar{d}}(\varepsilon)$ — функция ε при фиксированных $\bar{p} = (p_\alpha)$ и $\bar{d} = (d_\alpha)$, определяемая соотношениями $k_{\bar{p}\bar{d}}(\varepsilon) = \lambda[\gamma'(0) + \varepsilon] - \gamma(\lambda)$ [здесь $\lambda = \lambda(\varepsilon)$ — корень уравнения $\gamma'(0) + \varepsilon = \gamma'(\lambda)$, причем $\gamma(\lambda) = \ln \sum_{\alpha=1}^{a} p_\alpha e^{d_\alpha \lambda}$];

$\mathfrak{M}(\varepsilon)$ — множество матриц $\mu = \| \mu_\alpha^\beta \| = (\bar{\mu}_\alpha)$, определяемых условиями $h(\bar{\mu}_\alpha, \bar{p}_\alpha) - h(\bar{p}_\alpha) \leqslant \varepsilon_\alpha$ ($\alpha = \overline{1, a}$), причем $k_{\bar{p}_\alpha, \bar{d}_\alpha}(\varepsilon_\alpha) = \varepsilon^2/\mu_\alpha$ ($\mu_\alpha > 0$) и $\bar{d}_\alpha = (-\ln p_\alpha^\beta)$;

$\mathrm{E}\xi$ — среднее значение (математическое ожидание) случайной величины ξ;

$\mathrm{D}\xi = \mathrm{E}(\xi - \mathrm{E}\xi)^2$ — дисперсия случайной величины ξ.

Буквы, набранные жирным шрифтом, означают матрицы.

Литература

1 В. А. Котельников. *Теория потенциальной помехоустойчивости.* Госэнергоиздат, 1956.
2. К. Шеннон. *Статистическая теория передачи электрических сигналов.* Сборник переводов «Теория передачи электрических сигналов при наличии помех». ИЛ, 1953.
3. К. Шеннон. *Связь при наличии шума.* Сборник переводов «Теория информации и ее приложения». Физматгиз, 1959
4. Б. С. Флейшман. *Конструкция оптимального кода в простейшем случае бинарного канала.* Научн. докл. высш. школы, серия радиотехника и электроника, № 1, 1958.
5. Б. С. Флейшман *О конструктивном доказательстве основной теоремы Шеннона в простейшем бинарном случае.* Труды Всесоюзного совещания по теории вероятностей и математической статистике, сентябрь 1958 г. Изд. АН Арм. ССР, 1960.
6 Б. С. Флейшман. *Построение оптимального в смысле Шеннона кода в простейшем случае бинарного канала с шумами.* Сборник трудов НТО радиотехники и электросвязи им. А. С. Попова, вып. 3. Госэнергоиздат, 1959.
7. Б. С. Флейшман. *Комбинаторика расположений.* Ученые записки Моск. обл. педагог. ин-та, т 57, труды кафедр математики, вып. 4, 1957.
8. А. А. Марков. *Об испытаниях, связанных в цепь ненаблюдаемыми событиями.* Известия Российской академии наук, (6), 6, 1912.
9. В. И. Романовский. *О цепных корреляциях.* Ташкент, Изд. Комитета наук при СНК Уз. ССР, 1939.
10. C. Shannon *Certain results in coding theory for noisy channels.* Information and Control, 1, 6-25, 1957. (Есть русский перевод: К. Шеннон. *Некоторые результаты теории кодирования для каналов с шумами.* Периодический сборник переводов иностранных статей «Математика», № 2, ИЛ, 1959).
11. E. Netto. *Lehrbuch der Combinatorik.* Leipzig. 1901.
12. P. J. Mac Magon. *Combinatory analysis.* Cambridge University Press, 1915.
13. C. Riordan. *Introduktion in combinatorial analysis.* N. Y., Wiley a. sons, 1958.
14. К. Берж. *Теория графов и ее применения.* ИЛ, 1962.
15. Г. Зейферт, В. Трельфалль. *Комбинаторная топология.* ГОНТИ, 1938.
16. Д. Б. Юдин, Е. Г. Гольштейн. *Задачи и методы линейного программирования.* Изд-во «Сов. радио», 1961.
17. А. Н. Колмогоров. *Основные понятия теории вероятностей.* ОНТИ, 1936.
18. R. Mises. *Uber Aufteilungs- und Besetzungswahrscheinlichkeiten.* Revue de la Faculté des Sciences de l'universite d'Istambul, N. S. 4, 1939.
19. Б. В. Гнеденко. *Курс теории вероятностей,* изд. 3. Физматиздат, 1962.
20. H. Chernov. *A measure of asymptotic efficiency for tests of a hypotheses, based on the sum of observations.* Ann. Mat. Statistics, v. 23, 1952.
21. S. Kullback. *Certain inequalites in information theory and the Cramer — Rao inequality.* Ann. Mat. Statistics, v. 25, 1954.
22. И. Н. Санов. *О вероятности больших отклонений случайных величин.* Мат. сборник, т. 42, № 1, 1957.
23. В. Феллер. *Введение в теорию вероятностей и ее приложения.* ИЛ, 1952.
24. W. Hamming. *Error detecting and error correcting codes.* RSTJ, April, 1950. (Есть русский перевод в сборнике «Коды с обнаружением и исправлением ошибок», ИЛ, 1956).
25. В. И. Романовский. *Дискретные цепи Маркова.* Гостехиздат, 1949.
26 В. И. Романовский. *Математическая статистика.* ГОНТИ, 1938.
27. У. Р. Эшби. *Введение в кибернетику.* ИЛ, 1960.
28. M. Frechet. *Les probabilités associeés a un systéme d'evenements compatibles et dependants. Actualités* Scientifiquens et industrielles, № 859, 1940; № 942, 1943, Paris.
29. С. Н. Бернштейн. *Теория вероятностей.* Гостехиздат, 1946.
30. А. А. Марков. *Избранные труды.* Изд-во АН СССР, 1951.
31. J. Neyman, E. Pearson. *The testing of statistical hypotheses in relation to probability a priori.* Proc. Cambr. Phil. Soc., v. 29, 1933.
32. C R. Rao *Advanced statistical methods in biometric research.* N. Y., Wiley a. sons, 1952.

33. А. Вальд. *Последовательный анализ.* Физматгиз, 1960.
34. А. Е. Башаринов, Б. С. Флейшман. *Методы статистического последовательного анализа и их приложения.* Изд-во «Сов. радио», 1962.
35. R. M. Fano. *Transmission of information.* N. Y.— London, M. I. T. and Wiley a. sons, 1961.
36. G. A. Barnard. *Simple proofs of simple cases of the coding theorem.* Imperial College, London, 3 symposium of theory Information, september, 1956. (Есть русский перевод: Г Бернард. *Простое доказательство простых случаев теоремы кодирования.* Труды 3-й международной конференции по теории информации. Сборник статей «Теория передачи сообщений» под ред. В. И. Сифорова. ИЛ, 1957).
37. В. И. Сифоров. *К теории идеального кодирования бинарной передачи.* «Радиотехника и электроника», т. 1, вып. 4, Изд-во АН СССР, 1956.
38. P. Elias. *Coding for two noisy channels.* Departament of Electrical Engineering and Research Laboratory of Electronics, Massachusetts Inst. of Technology, Cambridge, Massachusetts, 1955. (Есть русский перевод: П. Элайес. *Кодирование для двух каналов с шумами.* Труды 3-й международной конференции по теории информации. Сборник статей «Теория передачи сообщений» под ред. В. И. Сифорова. ИЛ, 1957).
39. А. Г. Постников. *Арифметическое моделирование случайных процессов.* Труды Мат. ин-та им. В. А. Стеклова, т. 7, 1960.
40. A. B. Fontaine, W. W. Peterson. *On coding for the binary simmetric channel.* Communications and Electronics, N 6. p. 638—643, 646—647, 1958.
41. Б. С. Флейшман. *Сравнение трех оптимальных кодов, имеющих различные способы построения.* Научн. докл. высш школы, серия радиотехника и электроника, № 1, 1958.
42. J. M. Wozencraft. B. Reiffen. *Sequential decoding.* N. Y.—London, M. I. T. and Wiley a. sons, 1961.
43. П. Элайес. *Кодирование в реальных системах связи.* Кибернетический сборник, № 4, ИЛ, 1962.
44. Н. Н. Воробьев. *Сложение независимых случайных величин на конечных абелевых группах.* Мат. сборник, т. 34, № 1, 1954.

1р.06к.