С. И. Чечёта

Введение в дискретную теорию информации и кодирования

МЦНМО

С. И. Чечёта

Введение в дискретную теорию информации и кодирования

Учебное пособие

Допущено Учебно-методическим объединением вузов по образованию в области информационной безопасности в качестве учебного пособия для студентов высших учебных заведений, обучающихся по специальностям «Криптография» и «Компьютерная безопасность»

Москва
Издательство МЦНМО
2011

УДК 519.719.2
ББК 32.81В6
Ч57

Чечёта С. И.

Ч57 Введение в дискретную теорию информации и кодирования: учебное издание. — М.: МЦНМО, 2011. — 224 с.

ISBN 978-5-94057-701-0

В пособии рассматриваются три основные проблемы: количественная оценка информации, создаваемой дискретными источниками и передаваемой по дискретным каналам связи; оптимальное кодирование для дискретных источников; помехоустойчивое кодирование для дискретных каналов. Пособие нацелено прежде всего на выделение и объяснение математических понятий, моделей и результатов, которые лежат в основе теории, и тем самым на создание условий для дальнейшего самостоятельного изучения её специальных и более сложных разделов и практических приложений.

Предполагается, что читатель владеет необходимыми знаниями по математическому анализу, алгебре, теории вероятностей, дискретной математике. Для преподавателей и научных работников, студентов и аспирантов.

ББК 32.81В6

ISBN 978-5-94057-701-0

Оглавление

Предисловие

Предлагаемое учебное пособие основано на опыте преподавания теории информации и кодирования студентам-математикам, специализирующимся в области информационной безопасности. В небольшом по объёму пособии невозможно подробно рассмотреть все разделы современной теории информации и кодирования, поэтому наше изложение ограничено рамками программы учебной дисциплины. Основное внимание в пособии уделяется математической теории, в нем практически не затрагиваются вопросы приложений теории к реальным техническим системам.

Предполагается, что читатель в достаточной степени владеет необходимыми знаниями по основным математическим дисциплинам, изучение которых предшествует теории информации и кодирования: по математическому анализу, алгебре, теории вероятностей и математической статистике, математической логике и теории алгоритмов, дискретной математике.

Предпочтение отдаётся сравнительно простым моделям — дискретным источникам информации и дискретным каналам связи. Непрерывные модели не рассматриваются, поскольку это потребовало бы существенно увеличить объём пособия и выйти за пределы учебной программы. Кроме того, корректное изложение непрерывных моделей теории информации и кодирования должно опираться на более подробное изучение теории случайных процессов в курсе теории вероятностей. Эти соображения объясняют, почему в названии пособия используется термин «дискретная теория».

Приступая к изучению теории информации и кодирования, вначале необходимо понять, что является предметом новой дисциплины, поскольку возможное толкование термина *информация* очень широкое, и без необходимых пояснений может создаться неверное впечатление, что в рамках данной дисциплины рассматриваются все аспекты информации.

Что такое *информация*? Мы часто употребляем это слово, считая само собой разумеющимся, как его понимать. Однако точно и полно определить это понятие совсем непросто. Это сложный вопрос, сложная философская проблема, к решению которой имеют-

ся различные подходы, отражающие связь информации с материей и энергией.

Под информацией мы будем подразумевать некоторую характеристику, описывающую состояние (или изменение состояния) рассматриваемой системы; такая характеристика всегда выражается в форме каких-то *сигналов* или *знаков* на материальных носителях. Функционирование сложных систем, как правило, связано с *порождением*, *обработкой*, *передачей*, *хранением* и *восприятием* информации.

Нас будут интересовать сигналы и знаки, при помощи которых записывается информация. Системы сигналов, знаков и *слов* (наборов знаков) образуют *языки*, *естественные* или *искусственные*.

Семиотика — наука о знаковых системах, словах и языках, рассматривает проблему выражения информации в *текстах* (последовательностях слов) на трёх основных уровнях: *синтаксис* — внутренние свойства текстов, правила их записи, структура знаковых систем и языков; *семантика* — смысловое содержание текстов, отношения между текстами и реальными объектами, качествами, действиями, т. е. связь между текстом и источником информации; *прагматика* — потребительское значение текстов, их ценность или полезность для получателей информации.

Теория информации и кодирования связана только с первым уровнем, т. е. синтаксисом, и, как следует из названия, состоит из двух основных разделов — теории информации и теории кодирования. Теория информации рассматривает задачи о количественной оценке информации на основе частот встречаемости (вероятностей) различных знаков и слов, не затрагивая смысла (семантики) и ценности (прагматики) информации, а теория кодирования изучает методы преобразования информации для её экономного представления или для обнаружения и исправления ошибок, возникающих при её передаче и хранении. Можно сказать, что теория информации и кодирования — это раздел математической теории связи, рассматривающий три основные проблемы:

1) количественная оценка создаваемой и передаваемой информации, анализ количественных информационных характеристик *источников информации* и *каналов связи*;
2) методы кодирования для экономного (кратчайшего) представления сообщений, порождаемых различными источниками ин-

формации;

3) методы кодирования для надёжной (безошибочной) передачи сообщений по каналам связи с шумом.

Это относительно молодая наука, возникшая в середине XX века. Её появление именно как целостной научной дисциплины, а не совокупности разрозненных результатов и технических приёмов, обычно связывают с выходом в свет в 1948 году статьи «Математическая теория связи» выдающегося американского математика и инженера Клода Шеннона (см. сборник трудов [35]). Появление этой новой научной дисциплины было подготовлено всем предшествующим развитием науки и техники. Здесь следует упомянуть Р. Хартли, впервые предложившего логарифмическую меру количества информации, В. А. Котельникова, обосновавшего возможность точного представления непрерывной функции совокупностью её значений в отдельных точках отсчёта, А. Н. Колмогорова, давшего фундаментальное обоснование теории вероятностей, и многих других учёных.

Последовательность изложения материала в учебном пособии будет в основном следовать общей схеме системы передачи (хранения) информации, приведённой на рисунке:

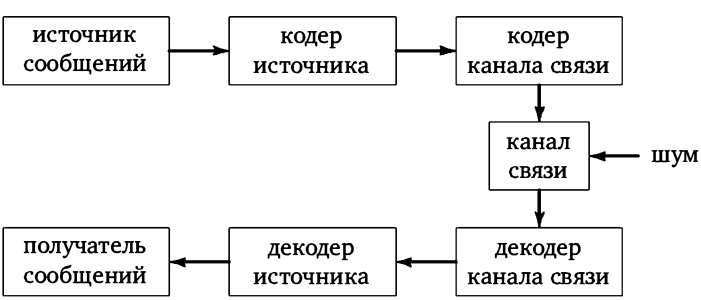

Рис. 1. Общая схема передачи информации

Будут последовательно рассмотрены меры количественной оценки информации, модели источников информации, методы экономного представления (кодирования) информации, порождаемой различными источниками, модели каналов связи и методы оценки их пропускной способности, методы кодирования и декодирования для

надёжной (безошибочной) передачи сообщений по каналам связи с шумом.

Перечислим кратко содержание учебного пособия по главам. При этом ещё раз отметим, что пособие действительно является только введением в обширную и активно развивающуюся область теории информации и кодирования. Принимая во внимание относительно небольшое время, отводимое на изучение дисциплины, мы ставим своей целью прежде всего выделение и объяснение тех понятий, моделей и результатов, которые лежат в основе теории, и тем самым создание условий для дальнейшего самостоятельного изучения её специальных и более сложных разделов и практических приложений. По этой причине многие важные понятия и результаты остались за рамками нашего пособия.

В главе 1 изложены основы количественного измерения информации для дискретных случайных величин — собственная, условная и взаимная информация, энтропия и средняя взаимная информация. Изучаемые здесь понятия и их свойства используются далее на протяжении всего пособия. Непрерывные случайные величины и их информационные характеристики не рассматриваются.

В главе 2 рассматриваются вероятностные модели дискретных источников сообщений и их информационные характеристики. Основное внимание уделяется источникам без памяти и марковским источникам. Для стационарных источников доказывается существование предельной энтропии. Центральное место в главе 2 занимает один из наиболее важных результатов теории информации, принадлежащий К. Шеннону, — свойство асимптотической равнораспределённости и его применения. За рамками главы остаются общие свойства эргодических источников, непрерывные источники, а также более сильные формулировки свойства асимптотической равнораспределённости и его обобщения.

Глава 3 вводит в основы теории кодирования дискретных источников при отсутствии шума. Рассматриваются однозначно декодируемые неравномерные коды, префиксные коды и кодовые деревья, условия их существования, энтропийные оценки средней длины кодового слова, оптимальные коды и способы их построения. Особенность изложения состоит в том, что подробно рассмотрено доказательство оптимальности кода Хаффмана в общем (не обязательно двоичном)

случае. Мы не касаемся многих разделов теории кодирования дискретных источников — моделирования источников с неизвестными вероятностными характеристиками, методов адаптивного кодирования и возникающих при этом алгоритмических вопросов, а также вопросов кодирования непрерывных источников с заданным критерием качества.

В главе 4 рассмотрены дискретные каналы связи с шумом. Основное внимание уделено понятию пропускной способности дискретного канала связи без памяти и способам её вычисления или оценки для некоторых частных случаев — симметричных каналов или каналов, представимых в виде соединения нескольких каналов. Геометрическое представление даёт наглядную иллюстрацию понятия пропускной способности. Каналы связи с памятью, нестационарные, непрерывные и полунепрерывные каналы не рассматриваются.

Глава 5 посвящена изложению двух принципиальных результатов К. Шеннона в теории информации — прямой и обратной теорем кодирования для каналов связи с шумом, указывающих на возможность достижения сколь угодно малой вероятности ошибочного декодирования, когда скорость передачи меньше пропускной способности, и отсутствие такой возможности, когда скорость передачи превышает пропускную способность. Рассматривается самый простой случай дискретных каналов связи без памяти, но при этом даётся полное доказательство основных теорем.

В заключительной главе 6 рассматриваются основы помехоустойчивого кодирования в дискретных каналах связи с шумом. Изучаются основные сведения о блоковых кодах, линейных кодах и циклических кодах. Рассматриваются методы декодирования для линейных кодов. За рамками главы остались эффективные алгоритмы декодирования для циклических кодов, свёрточные коды и многие другие вопросы.

В списке литературы приведены книги по теории информации и кодирования, рекомендуемые для первоначального ознакомления с теорией или дальнейшего более глубокого её изучения, а также учебники, на которые мы опираемся при использовании необходимых основных математических понятий и результатов. Например, в качестве основного руководства по алгебре мы ссылаемся на учебник М. М. Глухова, В. П. Елизарова и А. А. Нечаева [11], по теории вероятностей — на учебники А. А. Боровкова [5] и А. Н. Ширяева

[36], по комбинаторике и дискретной математике — на учебники В. Н. Сачкова [27] и С. В. Яблонского [37] и задачник Г. П. Гаврилова и А. А. Сапоженко [8].

Для первоначального ознакомления с теорией информации и кодирования можно рекомендовать книги М. Н. Аршинова и Л. Е. Садовского [2], А. Реньи [26], Р. В. Хэмминга [33], а также книгу А. М. Яглома и И. М. Яглома [38].

При изложении теоретического материала в учебном пособии были существенно использованы книги А. Акритаса [1], Р. Галлагера [9], В. Д. Колесника и Г. Ш. Полтырева [16], Р. Е. Кричевского [18], А. Я. Хинчина [32], Т. М. Ковера и А. Дж. Томаса [40], Р. Дж. МакЭлиса [46]. Например, в главе 1 мы следуем книгам [32], [16] и [18], в главе 3 — [37], в главе 5 — [40], в главе 6 — [1] и [46].

При составлении задач были использованы книги [1–4], [8–10], [13–17], [19], [20], [23–25], [28], [30], [31], [33], [34], [39–46], [48].

В книге принята двухпозиционная нумерация определений, утверждений, теорем, лемм, примеров, задач, рисунков и формул. Первая цифра номера обозначает главу, а число после точки — номер определения, теоремы и т. п. в данной главе; например, теорема 2.1 — первая теорема во второй главе, замечание 1.4 — четвёртое замечание в первой главе. Знак □ обозначает конец доказательства.

Для краткости во многих местах вместо слов *случайный, случайная* используется сокращение *сл.*, вместо термина *случайная величина* — сокращение *сл. в.*

Автор благодарен И. М. Арбекову, А. М. Гришину, А. А. Духину, А. М. Зубкову, А. Ю. Зубову, А. С. Котоусову, А. Б. Лосю, В. Н. Овчинникову, Б. А. Погорелову, В. В. Сизых, В. П. Ульяненко, А. В. Черемушкину, А. М. Шойтову за доброе внимание, обсуждение рукописи и полезные замечания.

Обозначения

Ниже приводится список основных обозначений, используемых в учебном пособии. Число в скобках означает номер страницы, на которой вводится обозначение.

$\mathscr{A} = \{a_1, \ldots, a_m\}$ — алфавит источника сообщений (множество сообщений) (15)

$\mathscr{A}^n$ — множество всех слов длины n из символов алфавита $\mathscr{A}$ (45)

$a^n = (a_{i_1}, \ldots, a_{i_n})$ — последовательность букв (слово) длины n (45)

$A = (A_n)_{n=1}^{+\infty} = (A_1, A_2, \ldots)$ — дискретный случайный процесс (бесконечная случайная последовательность) (46)

$A^k = (A_1, \ldots, A_k)$ — начальный отрезок сл. процесса A (сл. вектор) (46)

$\mathscr{A}_{n,\varepsilon}$ — множество ε-типичных последовательностей длины n (57)

$\mathscr{A}^\infty$ — множество всех бесконечных последовательностей из символов алфавита $\mathscr{A}$ (45)

$A\colon \Omega \to \mathscr{A}$ — сл. в. A (22)

$(\mathscr{A}, \vec{p})$ — дискретный источник без памяти (48)

$(\mathscr{A}, \vec{p}, Q)$ — простой марковский источник (49)

$\mathscr{B} = \{b_1, \ldots, b_D\}$ — кодовый алфавит (73)

$\mathscr{B}^* = \bigcup_{n \geqslant 1} \mathscr{B}^n$ — множество всех конечных слов в алфавите $\mathscr{B}$ (73)

$\mathscr{C} = \{x^n(1), \ldots, x^n(M)\}$ — код длины n и объёма M (137)

$\mathscr{C}^\perp$ — код, двойственный линейному коду $\mathscr{C}$ (171)

C^* — пропускная способность канала связи (113)

ДСК(p) — двоичный симметричный канал с параметром p (111)

D — дисперсия (61)

$\mathscr{D}$, $\mathscr{D}_n$ — декодер общего вида (137, 139)

$\mathscr{D}_{\mathsf{AP}}$ — декодер по методу максимальной апостериорной вероятности (140)

$\mathscr{D}_\varepsilon$ — декодер ε-типичных пар (148)

$\mathscr{D}_{\mathsf{L}}$ — декодер по методу максимального правдоподобия (140)

$\mathscr{D}_\rho$ — декодер в ближайшее кодовое слово (164)

$\mathscr{D}_S$ — декодер на основе таблицы стандартного расположения (178)

$d_{\mathscr{C}}$ — минимальное кодовое расстояние блокового кода $\mathscr{C}$ (166)

E — математическое ожидание (22)

e — вектор ошибок (177)

$\varphi : \mathscr{A} \to \mathscr{B}^*$ — алфавитное кодирование (73)

$\varphi(\mathscr{A}) = (\varphi(a_1), \ldots, \varphi(a_m))$ — код для алфавита $\mathscr{A}$ (73)

φ^* — продолжение алфавитного кодирования φ (73)

$\mathsf{G} = (\mathscr{V}, \mathscr{E})$ — граф с множеством вершин $\mathscr{V}$ и множеством рёбер $\mathscr{E}$ (75)

G — порождающая матрица линейного кода (170)

$g(x)$ — порождающий многочлен циклического кода (190)

$GF(q)$ — конечное поле из q элементов (169)

H — проверочная матрица линейного кода (171)

$H(A)$ — энтропия сл. в. A (22)

$H(A, B)$ — совместная энтропия сл. величин A и B (26)

$H(A/B=b_j)$ — условная энтропия сл. в. A при условии $\{B=b_j\}$ (26)

$H(A/B)$ — условная энтропия сл. в. A при условии сл. в. B (26)

H_k — энтропия на одну букву k-буквенного сообщения (53)

$H^{(k)}$ — условная энтропия k-й буквы (53)

$H(\vec{p})$ — энтропия распределения вероятностей $\vec{p}$ (22)

H_∞ — энтропия источника сообщений (54)

$\mathscr{H}_r$ — код Хэмминга (202)

$h(p)$ — функция двоичной энтропии (22)

$h(x)$ — проверочный многочлен циклического кода (191)

$I(a_i)$ — количество собственной информации (19)

$I(A;B)$ — средняя взаимная информация сл. величин A и B (30)

$I(a_i/b_j)$ — количество условной информации (20)

$I(a_i;b_j)$ — количество взаимной информации (21)

$I(A;B/C)$ — условная средняя взаимная информация сл. величин A и B при условии сл. в. C (32)

$\chi(S)$ — индикатор события S (63)

$\ell_i=\mathrm{len}(\varphi(a_i))$ — длина кодового слова для буквы a_i (73)

ℓ^φ — средняя длина кодового слова при алфавитном кодировании φ (84)

λ, $\lambda^{(n)}$ — средняя вероятность ошибочного декодирования (139)

λ_i — условная вероятность ошибочного декодирования (139)

$\mathbb{N}_0$ — множество неотрицательных целых чисел (163)

$(\Omega, \mathscr{F}, \mathsf{P})$ — вероятностное пространство (16)

$\mathsf{P}_{n_1, n_2, \ldots, n_k}$ — конечномерное распределение сл. процесса (46)

$\pi^{(n)}(y_{j_1}, \ldots, y_{j_n} / x_{i_1}, \ldots, x_{i_n})$ — переходные вероятности канала связи (109)

$\mathbb{R}$ — множество действительных чисел (25)

$\mathbb{R}_+$ — множество неотрицательных действительных чисел (168)

R_n — скорость передачи информации (137)

$\mathscr{R}_n$ — факторкольцо классов многочленов (187)

$\rho(a^n, b^n)$ — расстояние Хэмминга (163)

$\hat{\rho}(a^n, b^n)$ — обобщённое расстояние Хэмминга (168)

$S = (\mathbf{y}_{ij})$ — таблица стандартного расположения (177)

S_{m-1} — симплекс размерности $m-1$ (23)

$s(\mathbf{y})$ — синдром вектора $\mathbf{y}$ (179)

$U_{1, \ldots, n}$ — распределение на множестве $\mathscr{X}^n$ (входное распределение канала связи) (109)

$\mathscr{W}_{n, \varepsilon}$ — множество ε-типичных пар (x^n, y^n) (148)

$w(\mathbf{c})$ — вес Хэмминга вектора $\mathbf{c}$ (170)

$\mathscr{X} = \{x_1, \ldots, x_q\}$ — входной алфавит канала связи (109)

$x^n(i) = (x_1(i), \ldots, x_n(i))$ — кодовое слово (137)

X^n — сл. вектор на входе канала связи (110)

$(\mathscr{X}, \mathscr{Y}, \pi)$ — канал связи (111)

$\mathscr{Y} = \{y_1, \ldots, y_s\}$ — выходной алфавит канала связи (109)

Y^n — сл. вектор на выходе канала связи (110)

Глава 1
Энтропия и взаимная информация

§ 1.1. Введение

В этой главе рассматриваются меры количественной оценки информации, порождаемой *дискретными источниками*. В простейшем случае источник полностью характеризуется множеством сообщений $\mathscr{A} = \{a_1, \ldots, a_m\}$ и набором вероятностей их появления $\vec{p} = (p_1, p_2, \ldots, p_m)$.

Нас интересует, как можно задать количественную меру F, некоторым разумным образом характеризующую неопределённость выбора очередного сообщения источника. Такая мера может также рассматриваться как мера количества информации, получаемой при полном устранении неопределённости относительно сообщения источника. Нужная мера должна удовлетворять ряду естественных требований.

Во-первых, требование **монотонности** означает, что мера должна монотонно возрастать с ростом числа сообщений источника, но при этом невозможные сообщения (т. е. имеющие нулевую вероятность) не должны учитываться.

Во-вторых, требование **аддитивности** означает следующее. Пусть имеются два независимых источника, имеющих соответственно множества $\mathscr{A} = \{a_1, \ldots, a_m\}$ и $\mathscr{B} = \{b_1, \ldots, b_n\}$ равновероятных (для каждого из множеств) сообщений. В совокупности их можно рассматривать как один источник, одновременно реализующий пару сообщений (a_i, b_j). Естественно предполагать, что неопределённость такого объединённого источника должна равняться сумме неопределённостей исходных источников. Так как число возможных сообщений объединённого источника равно mn, искомая мера должна удовлетворять требованию

$$F(mn) = F(m) + F(n).$$

Указанным двум требованиям удовлетворяет логарифмическая мера неопределённости, предложенная Р. Хартли в 1928 г.,

$$F(m) = \log_b m.$$

Выбор основания $b > 1$ не имеет существенного значения и задаёт только единицу измерения неопределённости. Поскольку почти вся современная техника обработки и передачи информации строится на элементах, имеющих два устойчивых состояния, обычно выбирают $b = 2$. При этом единица измерения неопределённости называется *бит* (от английского *binary digit* — двоичная цифра) и представляет собой неопределённость в простейшей ситуации выбора одного из двух равновозможных сообщений. Если выбрать $b = 10$, то единица измерения называется *дит*, а в случае $b = e$ (основание натуральных логарифмов) — *нат*.

Мера Хартли не получила широкого применения, так как в ней не учитывается то, что вероятностное распределение $\vec{p}$ на множестве сообщений источника может быть неравномерным. Например, если $m = 2$, то при распределении $\vec{p} = (0{,}5, 0{,}5)$ мера среднего количества информации в сообщении, что представляется естественным, должна быть больше, чем при распределении $\vec{p} = (0{,}99, 0{,}01)$. Желательно, чтобы мера среднего количества информации, *в-третьих*, зависела от распределения вероятностей $\vec{p}$ таким образом, чтобы наибольшее значение количества информации достигалось при равномерном распределении, а наименьшее — при вырожденном распределении (когда одно из сообщений имеет вероятность 1, а все остальные сообщения — вероятность 0).

В-четвёртых, требование **симметричности** означает, что мера не должна зависеть от перенумерации сообщений источника.

В-пятых, требование **непрерывности** означает, что при малых изменениях распределения $\vec{p}$ мера не должна сильно изменяться.

Эти и некоторые другие соображения привели К. Шеннона к определению энтропии — одному из важнейших понятий теории информации. Точное определение, свойства и разнообразные применения этого понятия будут подробно рассмотрены ниже.

§ 1.2. Дискретные случайные величины

Здесь и далее, как правило, будем рассматривать дискретные случайные величины, принимающие значения из конечного или счётного множества.

Пусть $(\Omega, \mathscr{F}, \mathrm{P})$ — вероятностное пространство, где

Ω — пространство элементарных событий $\omega \in \Omega$;

$\mathscr{F}$ — класс подмножеств (событий) $F \subseteq \Omega$, являющийся σ-алгеброй, т. е. содержащий пустое множество (невозможное событие) и само пространство Ω (достоверное событие) и такой, что для любых событий $F_1, F_2, \ldots \in \mathscr{F}$ выполняются условия $F_1 \setminus F_2 \in \mathscr{F}$, $\bigcap_{n\geqslant 1} F_n \in \mathscr{F}$;

P — вероятностная мера, т. е. такое отображение $\mathsf{P}: \mathscr{F} \to [0, 1]$, что $\mathsf{P}(\Omega) = 1$ и для любых попарно непересекающихся $F_1, F_2, \ldots \in \mathscr{F}$ выполняется условие $\mathsf{P}\bigl(\bigcup_{n\geqslant 1} F_n\bigr) = \sum_{n\geqslant 1} \mathsf{P}(F_n)$.

Пусть $\mathscr{A} = \{a_i, i \in I\}$ — конечное или счётное множество.

Определение 1.1. Дискретной случайной величиной (сл. в.) A, определённой на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ и принимающей значения из множества $\mathscr{A}$, называется произвольное измеримое отображение $A: \Omega \to \mathscr{A}$, т. е. такое отображение, что для любого элемента $a_i \in \mathscr{A}$ справедливо включение $A^{-1}(a_i) = \{\omega: A(\omega) = a_i\} \in \mathscr{F}$.

Набор чисел $p_A(a_i) = \mathsf{P}(A = a_i) = \mathsf{P}\{\omega: A(\omega) = a_i\}$, $i \in I$, называется распределением сл. в. A.

Для краткости используют запись

$$A \sim \begin{pmatrix} a_1 & a_2 & \ldots \\ p_A(a_1) & p_A(a_2) & \ldots \end{pmatrix},$$

называемую вероятностной схемой.

Замечание 1.1. Отметим принятое здесь соответствие обозначений, которого будем придерживаться в дальнейшем. Одноимёнными буквами разного начертания обозначаем: $\mathscr{A}$ — конечное или счётное множество (алфавит), a — элемент (буква) этого множества (алфавита), $a \in \mathscr{A}$, и A — сл. в., принимающая значения из множества $\mathscr{A}$.

Пусть $\mathscr{B} = \{b_j, j \in J\}$ — другой алфавит и $B: \Omega \to \mathscr{B}$ — заданная на том же вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ сл. в., принимающая значения из $\mathscr{B}$, с распределением

$$p_B(b_j) = \mathsf{P}(B = b_j) = \mathsf{P}\{\omega: B(\omega) = b_j\}, \quad j \in J.$$

Рассмотрим совместное распределение сл. величин A и B

$$\begin{aligned} p_{AB}(a_i, b_j) &= \mathsf{P}(A = a_i, B = b_j) = \\ &= \mathsf{P}\{\omega: A(\omega) = a_i, B(\omega) = b_j\} = \\ &= \mathsf{P}\{A^{-1}(a_i) \cap B^{-1}(b_j)\}, \quad i \in I, j \in J. \end{aligned}$$

Из соотношений

$$\{A=a_i\}=\bigcup_{j\in J}\{A=a_i, B=b_j\}, \quad \{B=b_j\}=\bigcup_{i\in I}\{A=a_i, B=b_j\}$$

вытекают равенства

$$p_A(a_i)=\sum_{j\in J} p_{AB}(a_i, b_j), \qquad p_B(b_j)=\sum_{i\in I} p_{AB}(a_i, b_j). \tag{1.1}$$

Если $\mathscr{A}$ и $\mathscr{B}$ — конечные алфавиты, $|\mathscr{A}|=m$, $|\mathscr{B}|=n$, то совместное распределение удобно записать в виде следующей двумерной таблицы.

$\mathscr{A}\setminus\mathscr{B}$	b_1	…	b_j	…	b_n	
a_1						$p_A(a_1)$
⋮						⋮
a_i			$p_{AB}(a_i, b_j)$			$p_A(a_i)$
⋮						⋮
a_m						$p_A(a_m)$
	$p_B(b_1)$	…	$p_B(b_j)$	…	$p_B(b_n)$	1

В этой таблице элементы $p_A(a_i)$ последнего столбца представляют собой суммы элементов $p_{AB}(a_i, b_j)$ предшествующих n столбцов (в каждой строке); аналогичное замечание относится к элементам $p_B(b_j)$ последней строки.

Напомним, что события $F_1, F_2 \in \mathscr{F}$ называются независимыми, если $\mathsf{P}\{F_1 \cap F_2\}=\mathsf{P}\{F_1\}\cdot\mathsf{P}\{F_2\}$; сл. величины A и B называются независимыми, если для любых $i\in I$, $j\in J$ события $\{A=a_i\}$ и $\{B=b_j\}$ независимы. Независимость сл. величин A и B равносильна тому, что для любого $a_i\in\mathscr{A}$ и любого $b_j\in\mathscr{B}$ с условием $p_B(b_j)>0$ выполняется равенство $\mathsf{P}(A=a_i/B=b_j)=\mathsf{P}(A=a_i)$.

Замечание 1.2. Если дано совместное распределение p_{AB}, то можно однозначно найти частные распределения p_A и p_B. С другой стороны, для двух заданных частных распределений p_A и p_B совместное распределение p_{AB}, удовлетворяющее равенствам (1.1), всегда существует, но определяется не единственным образом. Совместное распределение

p_{AB} по заданным распределениям p_A и p_B определяется однозначно в том и только том случае, когда сл. величины A и B независимы.

Всё сказанное выше о распределении двух сл. величин легко обобщается на случай трёх и более сл. величин. В дальнейшем полагаем, что все рассматриваемые сл. величины заданы на одном вероятностном пространстве.

§ 1.3. Количество информации

Вначале рассмотрим некоторые наводящие соображения, поясняющие идею приводимых ниже точных определений количества информации и энтропии.

Предположим, что один из двух играющих выбирает наугад число a из множества $\mathscr{A} = \{1, 2, 3, \ldots, 2^n\}$, а другой участник игры стремится угадать это число, задав первому наименьшее количество вопросов, на которые можно дать только ответы «да» или «нет». Каково наименьшее число вопросов, приводящих к успеху при любом выборе числа a? Нетрудно доказать, что это наименьшее число вопросов равно n. Одна из оптимальных стратегий угадывающего состоит в последовательном делении множества возможных вариантов на две равные части с вопросом, в какой из этих двух частей лежит искомое число a. Если все варианты выбора числа a равновозможны, то естественно связать количество информации $I(a)$ о том, какое число было выбрано, с наименьшим числом вопросов, необходимых для его угадывания, и положить $I(a) = n = \log_2 |\mathscr{A}|$.

Если варианты выбора числа $a \in \mathscr{A} = \{1, 2, 3, \ldots, m\}$ не равновозможны, а число вариантов $m = |\mathscr{A}|$ не равно целой степени двойки, то предложить разумную меру количества информации $I(a)$ о выборе числа a не так просто. Ясно, однако, что если каждому варианту $a \in \mathscr{A}$ приписана некоторая вероятность $p(a)$, то величина $I(a)$ должна как-то зависеть от $p(a)$: редкий (маловероятный) вариант a должен давать больше информации, чем часто выбираемый, а если выбор некоторого a предопределён, т. е. $p(a) = 1$, то при этом мы вообще не получаем никакой новой информации и должны положить $I(a) = 0$.

Определение 1.2. Пусть дана сл. в. A. Количеством собственной информации (или, коротко, собственной информацией), заключён-

ной в событии $\{A = a_i\}$, называется величина

$$I(a_i) = -\log_2 p_A(a_i).$$

Замечание 1.3. В качестве основания логарифма выбрано число 2, но, как уже было отмечено выше, это не имеет принципиального значения; можно взять в качестве основания логарифма любое число, большее единицы. Данное определение согласуется с нашими предварительными рассуждениями об измерении количества информации, получаемой при определении случайно выбранного числа из заданного множества. Мы связываем количество собственной информации, заключённой в некотором событии (сообщении), только с его вероятностью, и тем самым исключаем из рассмотрения смысл этого сообщения и его ценность.

Очевидно, что собственная информация неотрицательна, и при этом $I(a_i) = 0$ тогда и только тогда, когда $p_A(a_i) = 1$.

Определение собственной информации естественным образом обобщается на наборы событий. Например, если даны две сл. величины A и B, то собственной информацией, заключённой в событии $\{A = a_i, B = b_j\}$, называется величина

$$I(a_i, b_j) = -\log_2 p_{AB}(a_i, b_j).$$

Определение 1.3. Пусть даны сл. величины A и B. Количеством условной информации (или, коротко, условной информацией), заключённой в событии $\{A = a_i\}$ при условии события $\{B = b_j\}$, называется величина

$$I(a_i/b_j) = -\log_2 \mathrm{P}(A = a_i/B = b_j).$$

Непосредственно из определений 1.2 и 1.3 легко выводятся следующие свойства:

1) $I(a_i/b_j) = I(a_i)$ тогда и только тогда, когда события $\{A = a_i\}$ и $\{B = b_j\}$ независимы;
2) $I(a_i, b_j) = I(a_i) + I(b_j/a_i) = I(b_j) + I(a_i/b_j)$.

Определение условной информации, данное выше для двух сл. величин, естественно обобщается на случай трёх и более сл. величин. Например, если даны три сл. величины A, B и C, то по определению

полагаем

$$I(a_i/b_j, c_k) = -\log_2 \mathsf{P}(A = a_i/B = b_j, C = c_k),$$
$$I(a_i, b_j/c_k) = -\log_2 \mathsf{P}(A = a_i, B = b_j/C = c_k).$$

В качестве упражнения докажите следующие равенства:

$$I(a_i, b_j, c_k) = I(a_i) + I(b_j/a_i) + I(c_k/a_i, b_j),$$
$$I(a_i, b_j/c_k) = I(a_i/c_k) + I(b_j/a_i, c_k).$$

Определение 1.4. Пусть даны сл. величины A и B. Количеством взаимной информации (или, коротко, взаимной информацией) между событиями $\{A = a_i\}$ и $\{B = b_j\}$ называется величина

$$I(a_i; b_j) = I(a_i) - I(a_i/b_j).$$

Отметим, что в обозначении для собственной информации $I(a_i, b_j)$ мы используем запятую, а для взаимной информации $I(a_i; b_j)$ — точку с запятой.

Утверждение 1.1. *Взаимная информация обладает следующими свойствами.*

1. (Симметричность.) $I(a_i; b_j) = I(b_j; a_i)$.
2. (Условие равенства нулю.) $I(a_i; b_j) = 0$ *тогда и только тогда, когда события* $\{A = a_i\}$ *и* $\{B = b_j\}$ *независимы.*
3. (Аддитивность.) $I(a_i, b_j) = I(a_i) + I(b_j) - I(a_i; b_j)$.

Доказательство утверждения опирается только на введённые определения и сводится к простым выкладкам.

Замечание 1.4. В отличие от собственной информации, которая всегда неотрицательна, взаимная информация может быть равной нулю (если события независимы), положительной и отрицательной. Например, для совместного распределения сл. величин A и B

$\mathscr{A} \setminus \mathscr{B}$	b_1	b_2	
a_1	1/8	3/8	1/2
a_2	1/4	1/4	1/2
	3/8	5/8	1

нетрудно подсчитать, что

$$\begin{cases} I(a_1; b_1) = 1 - \log_2 3 < 0, \\ I(a_2; b_1) = 2 - \log_2 3 > 0. \end{cases}$$

§ 1.4. Энтропия

Определение 1.5. Энтропией сл. в. A с конечным или счётным множеством значений $\mathscr{A} = \{a_i, i \in I\}$ называется величина

$$H(A) = -\sum_{i \in I} p_A(a_i) \log_2 p_A(a_i),$$

понимаемая как конечная сумма при конечном множестве $\mathscr{A}$ или как сумма соответствующего ряда при бесконечном множестве $\mathscr{A}$; если же ряд расходится, то полагают $H(A) = +\infty$.

Замечание 1.5. 1. Необходимо условиться, как в определении 1.5 понимать слагаемое $p \log_2 p$ при $p = 0$. Поскольку $\lim_{p \downarrow 0} p \log_2 p = 0$, полагаем по определению $0 \log_2 0 = 0$.

2. Если $p_1 = \ldots = p_m = \frac{1}{m}$, то $H(A) = -\sum_{i=1}^{m} \frac{1}{m} \log_2 \frac{1}{m} = \log_2 m$, т. е. в равновероятном случае энтропия равна среднему количеству информации по Хартли.

3. Так как по определению 1.2 количество собственной информации равно $I(a_i) = -\log_2 p_A(a_i)$, формулу для энтропии можно переписать в виде $H(A) = \sum_{i \in I} p_A(a_i) I(a_i) = \mathsf{E}\, I(A)$, т. е. энтропия равна среднему количеству собственной информации, или, что то же самое, математическому ожиданию новой неотрицательной сл. в. $I(A)$, принимающей значения $I(a_i)$ с вероятностями $p_A(a_i)$, $i \in I$.

4. Поскольку энтропия зависит только от распределения сл. в. A и не зависит от множества её значений, то для $A \sim \begin{pmatrix} a_1 & a_2 & \ldots \\ p_1 & p_2 & \ldots \end{pmatrix}$ иногда пишут $H(A) = H(\vec{p}) = -\sum_{i \in I} p_i \log_2 p_i$.

Пример 1.1. Пусть $m = 2$, $A \sim \begin{pmatrix} a_1 & a_2 \\ p_1 & p_2 \end{pmatrix}$, $0 \leqslant p_i \leqslant 1$, $p_1 + p_2 = 1$. Так как $p_2 = 1 - p_1$, энтропия $H(A)$ фактически представляет собой функцию от одного аргумента p_1. В этом примере для $m = 2$ введём и будем использовать в дальнейшем *функцию двоичной энтропии*

$$h(p) = H(p, 1-p) = -p \log_2 p - (1-p) \log_2 (1-p),$$

заданную в промежутке $0 \leqslant p \leqslant 1$. Нетрудно видеть, что $h(0) = h(1) = 0$, $h(1/2) = 1$. В интервале $0 < p < 1$ функция $h(p)$ дифференцируема, её производная $h'(p) = \log_2((1-p)/p)$ положительна при $0 < p < 1/2$, отрицательна при $1/2 < p < 1$, равна нулю при $p = 1/2$, и при этом

$$\lim_{p \downarrow 0} h'(p) = +\infty, \qquad \lim_{p \uparrow 1} h'(p) = -\infty.$$

Кроме того, вторая производная $h''(p) = -\log_2 e/(p(1-p))$ отрицательна при $0 < p < 1$, и поэтому функция $h(p)$ выпукла вверх.

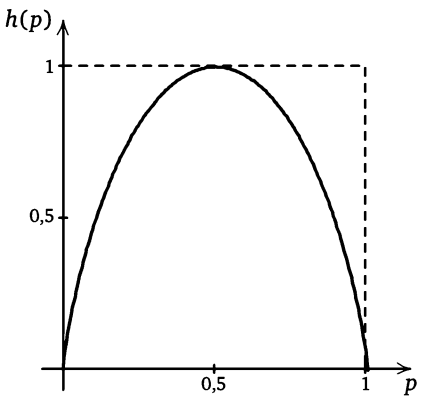

Рис. 1.1. График функции двоичной энтропии

В следующих двух утверждениях приведены некоторые свойства функции

$$H(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log_2 p_i,$$

заданной на симплексе $S_{m-1} = \{(p_1, \ldots, p_m)\colon p_i \geqslant 0, \sum_{i=1}^{m} p_i = 1\}$ размерности $m-1$.

Утверждение 1.2. *1. Функция $H(p_1, \ldots, p_m)$ непрерывна.*

2. Функция $H(p_1, \ldots, p_m)$ симметрична, т. е. не зависит от перестановки переменных $p_1, \ldots, p_m$.

3. $H(p_1, \ldots, p_m) \geqslant 0$, и равенство достигается в том и только том случае, когда одна из переменных $p_i = 1$, а остальные переменные равны нулю.

4. $H(p_1, \ldots, p_m) \leqslant \log_2 m$*, и равенство достигается в том и только том случае, когда* $p_1 = \ldots = p_m = \dfrac{1}{m}$.

Доказательство. Пункты 1 и 2 утверждения, а также неотрицательность функции $H(p_1, \ldots, p_m)$ очевидны из определения. Далее, в силу соглашения $0 \log_2 0 = 0$ имеем $H(1, 0, \ldots, 0) = 0$. Если же хотя бы одна из переменных p_i строго положительна и меньше единицы, то $-p_i \log_2 p_i > 0$, и поэтому $H(p_1, \ldots, p_m) > 0$.

Для доказательства пункта 4 нам понадобится известное из курса математического анализа неравенство

$$\ln p \leqslant p - 1, \tag{1.2}$$

которое справедливо для всех $p > 0$ и обращается в равенство только для $p = 1$. Неравенство (1.2) означает, что график строго выпуклой вверх функции $\ln p$ лежит ниже касательной $p - 1$, проведённой в точке $(1, 0)$.

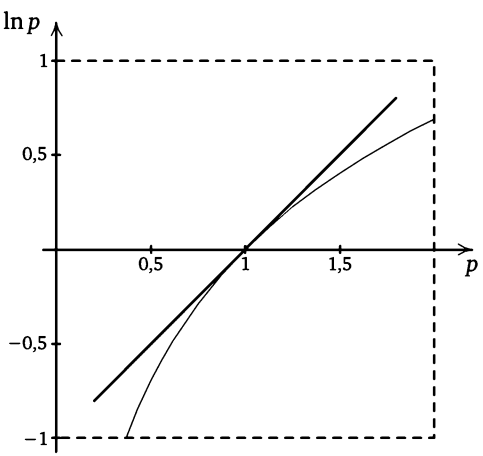

Рис. 1.2. График функции $\ln p$

Переходя к логарифму по основанию 2, перепишем (1.2) в виде

$$\log_2 p \leqslant (p - 1) \cdot \log_2 e. \tag{1.3}$$

Применяя это неравенство, получим:

$$\begin{aligned} H(p_1, \ldots, p_m) - \log_2 m &= -\sum_{i=1}^{m} p_i \log_2 p_i - \log_2 m \sum_{i=1}^{m} p_i = \\ &= \sum_{i=1}^{m} p_i \log_2 \frac{1}{mp_i} \leqslant \\ &\leqslant \log_2 e \cdot \sum_{i=1}^{m} p_i \left(\frac{1}{mp_i} - 1\right) = \\ &= \log_2 e \cdot \left(\sum_{i=1}^{m} \frac{1}{m} - \sum_{i=1}^{m} p_i\right) = \\ &= \log_2 e \cdot (1-1) = 0, \end{aligned}$$

так что

$$H(p_1, \ldots, p_m) - \log_2 m \leqslant 0,$$

и равенство достигается тогда и только тогда, когда $mp_i = 1$ для всех $1 \leqslant i \leqslant m$. □

Напомним, что множество $S \subseteq \mathbb{R}^m$ называется выпуклым, если для любых точек $x, y \in S$ и числа α, $0 \leqslant \alpha \leqslant 1$, точка $\alpha x + (1-\alpha)y$ также лежит в S. Действительная функция $f : S \to \mathbb{R}$ от m переменных называется выпуклой (вверх), если её область определения S есть выпуклое множество и для любых точек $x, y \in S$ и числа α, $0 \leqslant \alpha \leqslant 1$, справедливо неравенство $f(\alpha x + (1-\alpha)y) \geqslant \alpha f(x) + (1-\alpha) f(y)$. Если это неравенство строгое для всех α, отличных от нуля и единицы, то функция f называется строго выпуклой.

Известно, что если действительная функция одного переменного задана на некотором интервале, дважды дифференцируема на нём и её вторая производная всюду отрицательна, то эта функция строго выпукла.

Утверждение 1.3. *Функция $H(p_1, \ldots, p_m)$ строго выпукла.*

Доказательство. Нетрудно видеть, что $(m-1)$-мерный симплекс S_{m-1} является выпуклым множеством. Кроме того, функция $f(x) = -x \log_2 x$, определённая при $x > 0$, строго выпукла вверх, так как эта функция дважды дифференцируема, и её вторая производная $f''(x) = -(\log_2 e)/x$ всюду отрицательна. Поскольку $H(p_1, \ldots, p_m) =$

$= \sum_{i=1}^{m} f(p_i)$, функция $H(p_1, \ldots, p_m)$ строго выпукла как сумма строго выпуклых функций. □

Определение 1.6. Пусть даны сл. величины A и B.

1. Энтропией совместного распределения (или, коротко, совместной энтропией) сл. величин A и B называется величина

$$H(A, B) = -\sum_{i \in I} \sum_{j \in J} p_{AB}(a_i, b_j) \log_2 p_{AB}(a_i, b_j).$$

2. Если $b_j \in \mathscr{B}$ и $p_B(b_j) > 0$, то условной энтропией сл. в. A при условии события $\{B = b_j\}$ называется величина

$$H(A/B = b_j) = -\sum_{i \in I} \mathsf{P}(A = a_i/B = b_j) \log_2 \mathsf{P}(A = a_i/B = b_j).$$

3. Условной энтропией сл. в. A при условии сл. в. B называется величина

$$H(A/B) = -\sum_{i \in I} \sum_{j \in J} p_{AB}(a_i, b_j) \log_2 \mathsf{P}(A = a_i/B = b_j).$$

Замечание 1.6. 1. Принимая во внимание определения величин собственной информации $I(a_i, b_j)$ и условной информации $I(a_i/b_j)$, аналогично п. 3 замечания 1.5 получаем равенства

$$H(A, B) = \sum_{i \in I} \sum_{j \in J} p_{AB}(a_i, b_j) I(a_i, b_j) = \mathsf{E}\, I(A, B),$$

$$H(A/B) = \sum_{i \in I} \sum_{j \in J} p_{AB}(a_i, b_j) I(a_i/b_j) = \mathsf{E}\, I(A/B),$$

где $I(A, B)$ и $I(A/B)$ — новые сл. величины, принимающие соответственно значения $I(a_i, b_j)$ и $I(a_i/b_j)$ с вероятностями $p_{AB}(a_i, b_j)$, $i \in I$, $j \in J$.

2. Из определения 1.6 вытекает, что условная энтропия $H(A/B)$ является усреднением величин $H(A/B = b_j)$ по распределению сл. в. B:

$$H(A/B) = \sum_{j \in J} \mathsf{P}_B(b_j) H(A/B = b_j).$$

Следующие несколько утверждений описывают соотношения между значениями совместной и условной энтропии различных сл. величин.

Утверждение 1.4 (свойство аддитивности).

$$H(A, B) = H(A) + H(B/A) = H(B) + H(A/B).$$

Доказательство вытекает непосредственно из определений 1.5 и 1.6 и сводится к простым выкладкам. □

Утверждение 1.5 (правило цепочки).

$$H(A_1, A_2, \dots, A_N) = H(A_1) + H(A_2/A_1) + H(A_3/A_1, A_2) + \dots$$
$$\dots + H(A_N/A_1, A_2, \dots, A_{N-1}).$$

Доказательство. Доказательство несложно проводится методом математической индукции по $N \geqslant 2$, причём базис индукции $N = 2$ совпадает с утверждением 1.4. □

Утверждение 1.6. *Справедливо неравенство*

$$H(A/B) \leqslant H(A), \tag{1.4}$$

причём равенство имеет место тогда и только тогда, когда сл. величины A и B независимы.

Доказательство. Рассмотрим разность

$$H(A/B) - H(A) = -\sum_{i\in I}\sum_{j\in J} p_{AB}(a_i, b_j)\log_2 \mathsf{P}(A = a_i/B = b_j) +$$
$$+\sum_{i\in I} p_A(a_i)\log_2 p_A(a_i) =$$
$$= \sum_{i\in I}\sum_{j\in J} p_{AB}(a_i, b_j)\log_2 \frac{p_A(a_i)}{\mathsf{P}(A = a_i/B = b_j)} =$$
$$= \sum_{i\in I}\sum_{j\in J} p_{AB}(a_i, b_j)\log_2 \frac{p_A(a_i)p_B(b_j)}{p_{AB}(a_i, b_j)}.$$

Снова воспользуемся неравенством (1.3) и получим

$$H(A/B) - H(A) \leqslant \log_2 e \cdot \sum_{i\in I}\sum_{j\in J} p_{AB}(a_i, b_j)\left(\frac{p_A(a_i)p_B(b_j)}{p_{AB}(a_i, b_j)} - 1\right) =$$
$$= \log_2 e \cdot \left(\sum_{i\in I}\sum_{j\in J} p_A(a_i)p_B(b_j) - \sum_{i\in I}\sum_{j\in J} p_{AB}(a_i, b_j)\right) =$$
$$= \log_2 e \cdot (1 - 1) = 0,$$

так что $H(A/B) \leqslant H(A)$, причём равенство имеет место тогда и только тогда, когда $p_A(a_i)p_B(b_j) = p_{AB}(a_i, b_j)$ для любых $i \in I$, $j \in J$, что и означает независимость сл. величин A и B. □

Неравенства, аналогичные (1.4), справедливы для случая трёх и более сл. величин. Повторяя приведённые выше рассуждения практически дословно, с очевидными видоизменениями, можно доказать, например, следующее

Утверждение 1.7. *Справедливо неравенство*

$$H(B/A, C) \leqslant H(B/C), \tag{1.5}$$

причём равенство имеет место тогда и только тогда, когда B не зависит от A при условии C, т. е. для любых $a_i \in \mathscr{A}$, $b_j \in \mathscr{B}$ и $c_k \in \mathscr{C}$ с условием $p_{ABC}(a_i, b_j, c_k) > 0$ выполняется равенство

$$\mathsf{P}(B = b_j/A = a_i, C = c_k) = \mathsf{P}(B = b_j/C = c_k).$$

Замечание 1.7. Неравенству (1.4) и другим неравенствам такого типа можно дать нестрогое, но наглядное истолкование. Если $H(A)$ есть среднее количество информации, получаемое от наблюдения над сл. в. A, а $H(A/B)$ — условное среднее количество информации, получаемое от наблюдения над сл. в. A при условии, что предварительно уже известен результат наблюдения над сл. в. B, то естественно ожидать, что во втором случае (когда что-то уже заранее известно) мы получим от наблюдения над сл. в. A меньше информации.

Из утверждений 1.5 и 1.6 вытекает

Следствие 1.1 (свойство полуаддитивности).

$$H(A_1, A_2, \dots, A_N) \leqslant H(A_1) + H(A_2) + H(A_3) + \dots + H(A_N),$$

причём равенство имеет место тогда и только тогда, когда сл. величины $A_1, A_2, \dots, A_N$ независимы.

Теорема 1.1. *Пусть $\mathscr{A} = \{a_i, i \in I\}$ и $\mathscr{B} = \{b_j, j \in J\}$ — конечные или счётные множества, $\varphi : \mathscr{A} \to \mathscr{B}$ — произвольное отображение. Пусть A и B — сл. величины, принимающие значения в множествах $\mathscr{A}$ и $\mathscr{B}$ соответственно и связанные условием $B = \varphi(A)$.*

Тогда справедливо неравенство $H(B) \leqslant H(A)$, причём равенство имеет место тогда и только тогда, когда отображение φ взаимно однозначно на подмножестве $\{a_i \in \mathscr{A} : p_A(a_i) > 0\}$ элементов a_i с положительными вероятностями.

Доказательство. Для $b_j \in \mathscr{B}$ обозначим через $\varphi^{-1}(b_j)$ полный прообраз элемента b_j при отображении φ, т. е.

$$\varphi^{-1}(b_j) = \{a_i \in \mathscr{A} : \varphi(a_i) = b_j\}.$$

Тогда по определению сл. величин A и B

$$p_B(b_j) = \begin{cases} \sum_{i:\, \varphi(a_i)=b_j} p_A(a_i), & \text{если } \varphi^{-1}(b_j) \neq \varnothing; \\ 0, & \text{если } \varphi^{-1}(b_j) = \varnothing. \end{cases}$$

Поэтому

$$\begin{aligned} H(B) &= -\sum_{j \in J} p_B(b_j) \log_2 p_B(b_j) = \\ &= -\sum_{j:\, \varphi^{-1}(b_j) \neq \varnothing} p_B(b_j) \log_2 p_B(b_j) = \\ &= -\sum_{j:\, \varphi^{-1}(b_j) \neq \varnothing} \Big(\sum_{i:\, \varphi(a_i)=b_j} p_A(a_i) \Big) \log_2 \Big(\sum_{i:\, \varphi(a_i)=b_j} p_A(a_i) \Big). \end{aligned}$$

Заметим, что для различных индексов j с условием $\varphi^{-1}(b_j) \neq \varnothing$ во внешней сумме соответствующие области суммирования $\{i : \varphi(a_i) = b_j\}$ во внутренних суммах попарно не пересекаются и

$$\bigcup_{j:\, \varphi^{-1}(b_j) \neq \varnothing} \{i : \varphi(a_i) = b_j\} = I.$$

Кроме того, из областей суммирования во внутренних суммах можно исключить те индексы i, для которых $p_A(a_i) = 0$.

Ясно также, что для любых $p_1, p_2 \geqslant 0$

$$\begin{aligned} (p_1 + p_2) \log_2 (p_1 + p_2) &= p_1 \log_2 (p_1 + p_2) + p_2 \log_2 (p_1 + p_2) \geqslant \\ &\geqslant p_1 \log_2 p_1 + p_2 \log_2 p_2, \end{aligned}$$

и равенство имеет место тогда и только тогда, когда $p_1 = 0$ или $p_2 = 0$.

Следовательно,

$$H(B) = -\sum_{j:\,\varphi^{-1}(b_j)\neq\varnothing}\Big(\sum_{\substack{i:\,\varphi(a_i)=b_j,\\ p_A(a_i)>0}} p_A(a_i)\Big)\times\log_2\Big(\sum_{\substack{i:\,\varphi(a_i)=b_j,\\ p_A(a_i)>0}} p_A(a_i)\Big)\leqslant$$
$$\leqslant -\sum_{i:\,p_A(a_i)>0} p_A(a_i)\log_2 p_A(a_i) = H(A),$$

так что $H(B)\leqslant H(A)$, и равенство имеет место тогда и только тогда, когда все области суммирования $\{i:\varphi(a_i)=b_j,\ p_A(a_i)>0\}$ во внутренних суммах представляют собой одноэлементные множества. А это условие, в свою очередь, эквивалентно тому, что отображение φ взаимно однозначно на подмножестве $\{a_i\in\mathscr{A}:\ p_A(a_i)>0\}$. □

Замечание 1.8. Теореме 1.1 можно дать такое нестрогое истолкование: среднее количество информации, получаемое от наблюдения случайного исхода некоторого эксперимента, не увеличивается после обработки результатов наблюдения, и полностью сохраняется только в том случае, когда эта обработка не огрубляет результатов.

§ 1.5. Средняя взаимная информация

Пусть заданы сл. величины A и B. Рассмотрим новую сл. в. $I(A;B)$, принимающую значения $I(a_i;b_j)$ с вероятностями $p_{AB}(a_i;b_j)$, $i\in I$, $j\in J$. Согласно определению 1.4, $I(a_i;b_j)=I(a_i)-I(a_i/b_j)$, поэтому

$$I(A;B) = I(A) - I(A/B). \tag{1.6}$$

Определение 1.7. Средней взаимной информацией $\boldsymbol{I}(A;B)$ сл. величин A и B называется величина, равная математическому ожиданию сл. в. $I(A;B)$:

$$\boldsymbol{I}(A;B) = \mathsf{E}\,I(A;B) = \sum_{i\in I}\sum_{j\in J} p_{AB}(a_i,b_j)\,I(a_i;b_j).$$

Отметим, что в отличие от обозначения сл. в. $I(A;B)$, в обозначении средней взаимной информации $\boldsymbol{I}(A;B)$ используем полужирный шрифт.

Утверждение 1.8. *Справедливы равенства*

$$\boldsymbol{I}(A;B) = H(A) - H(A/B) =$$
$$= H(B) - H(B/A) = H(A) + H(B) - H(A,B).$$

Доказательство. Принимая во внимание формулу (1.6), замечания 1.5 и 1.6, получим

$$I(A;B) = \mathsf{E}\, I(A;B) = \mathsf{E}\, I(A) - \mathsf{E}\, I(A/B) = H(A) - H(A/B).$$

Второе выражение для $I(A;B)$ получается аналогично в силу симметричности взаимной информации (см. утверждение 1.1). Третье выражение следует из первых двух и утверждения 1.4. □

Из утверждений 1.6 и 1.8 вытекает

Следствие 1.2. *Средняя взаимная информация неотрицательна,*

$$I(A;B) \geqslant 0,$$

причём равенство имеет место тогда и только тогда, когда сл. величины A и B независимы.

Замечание 1.9. Отметим связь следствия 1.2 с замечанием 1.4 — хотя взаимная информация может быть положительной, отрицательной или равной нулю, её среднее значение всегда неотрицательно.

Соотношения между величинами $I(A;B)$, $H(A)$, $H(A/B)$, $H(B)$ и $H(B/A)$ иллюстрирует следующий рисунок.

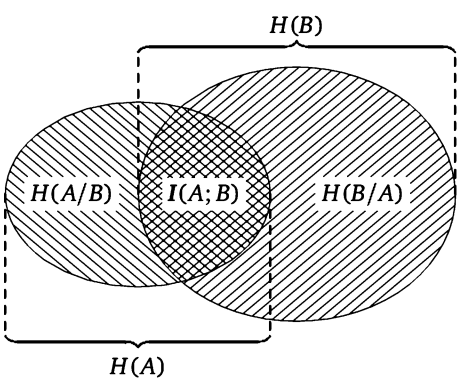

Рис. 1.3. Средняя взаимная информация и энтропия

При рассмотрении последовательного соединения каналов связи в главе 4 нам понадобится некоторое обобщение понятия средней взаимной информации для трёх сл. величин. Пусть даны сл. величины A, B и C, заданные на одном вероятностном пространстве и при-

нимающие значения в множествах $\mathscr{A} = \{a_i, i \in I\}$, $\mathscr{B} = \{b_j, j \in J\}$ и $\mathscr{C} = \{c_k, k \in K\}$ соответственно.

Определение 1.8. Условной средней взаимной информацией двух сл. величин A и B при условии сл. в. C называется величина

$$I(A; B/C) = H(A/C) + H(B/C) - H(A, B/C).$$

Утверждение 1.9. *Справедливо равенство*

$$I(A; (B, C)) = I(A; C) + I(A; B/C).$$

Доказательство. Складывая выражения для $I(A; C)$ и $I(A; B/C)$ почленно, получим

$$\begin{aligned}
&I(A; C) + I(A; B/C) = \\
&\quad = H(A) + H(C) - H(A, C) + H(A/C) + H(B/C) - H(A, B/C) = \\
&\quad = H(A) + (H(C) + H(B/C)) - (H(A, C) - H(A/C)) - H(A, B/C) = \\
&\quad = H(A) + H(B, C) - H(C) - H(A, B/C) = \\
&\quad = H(A) + H(B, C) - H(A, B, C) = \\
&\quad = I(A; (B, C)). \quad \square
\end{aligned}$$

Утверждение 1.10. *Справедливо неравенство* $I(A; B/C) \geqslant 0$.

Доказательство. По определению 1.8

$$\begin{aligned}
I(A; B/C) &= H(A/C) + H(B/C) - H(A, B/C) = \\
&= -\sum_{i \in I,\, k \in K} p_{AC}(a_i, c_k) \log_2 \mathsf{P}(A = a_i / C = c_k) - \\
&\qquad - \sum_{j \in J,\, k \in K} p_{BC}(b_j, c_k) \log_2 \mathsf{P}(B = b_j / C = c_k) + \\
&\qquad + \sum_{i \in I,\, j \in J,\, k \in K} p_{ABC}(a_i, b_j, c_k) \log_2 \mathsf{P}(A = a_i, B = b_j / C = c_k) = \\
&= -\sum_{i \in I,\, j \in J,\, k \in K} p_{ABC}(a_i, b_j, c_k) \log_2 \frac{\mathsf{P}(A = a_i / C = c_k)\mathsf{P}(B = b_j / C = c_k)}{\mathsf{P}(A = a_i,\, B = b_j / C = c_k)},
\end{aligned}$$

причём можно считать, что написанные суммы распространены на такие множества индексов i, j, k, для которых $p_{ABC}(a_i, b_j, c_k) > 0$, и поэтому все условные вероятности корректно определены. При

этом выражение под знаком логарифма равно

$$\frac{p_C(c_k)}{p_{ABC}(a_i, b_j, c_k)} \cdot \frac{p_{AC}(a_i, c_k)}{p_C(c_k)} \cdot \frac{p_{BC}(b_j, c_k)}{p_C(c_k)}.$$

Воспользуемся неравенством (1.3) и получим соотношения:

$$I(A; B/C) =$$

$$= - \sum_{i \in I,\, j \in J,\, k \in K} p_{ABC}(a_i, b_j, c_k) \log_2 \frac{p_{AC}(a_i, c_k) p_{BC}(b_j, c_k)}{p_C(c_k) p_{ABC}(a_i, b_j, c_k)} \geqslant$$

$$\geqslant \log_2 e \sum_{i \in I,\, j \in J,\, k \in K} p_{ABC}(a_i, b_j, c_k) \left(1 - \frac{p_{AC}(a_i, c_k) p_{BC}(b_j, c_k)}{p_C(c_k) p_{ABC}(a_i, b_j, c_k)}\right) =$$

$$= \log_2 e \sum_{i \in I,\, j \in J,\, k \in K} p_{ABC}(a_i, b_j, c_k) -$$

$$- \log_2 e \sum_{k \in K} \frac{1}{p_C(c_k)} \sum_{i \in I} p_{AC}(a_i, c_k) \sum_{j \in J} p_{BC}(b_j, c_k) =$$

$$= \log_2 e \left(1 - \sum_{k \in K} \frac{1}{p_C(c_k)} p_C(c_k) p_C(c_k)\right) = 0,$$

так что $I(A; B/C) \geqslant 0$. □

Утверждение 1.11. *Равенство $I(A; B/C) = 0$ справедливо тогда и только тогда, когда B не зависит от A при условии C, т. е. для любых $a_i \in \mathscr{A}$, $b_j \in \mathscr{B}$ и $c_k \in \mathscr{C}$ с условием $p_{ABC}(a_i, b_j, c_k) > 0$ выполняется равенство*

$$\mathsf{P}(B = b_j / A = a_i, C = c_k) = \mathsf{P}(B = b_j / C = c_k).$$

Доказательство. Из доказательства утверждения 1.10 и условия обращения неравенства (1.3) в равенство следует, что $I(A; B/C) = 0$ тогда и только тогда, когда для всех значений индексов i, j, k, для которых $p_{ABC}(a_i, b_j, c_k) > 0$, справедливо равенство

$$\frac{p_{AC}(a_i, c_k) p_{BC}(b_j, c_k)}{p_C(c_k) p_{ABC}(a_i, b_j, c_k)} = 1.$$

Это равенство, в свою очередь, равносильно равенству

$$\frac{p_{ABC}(a_i, b_j, c_k)}{p_{AC}(a_i, c_k)} = \frac{p_{BC}(b_j, c_k)}{p_C(c_k)},$$

или

$$\mathsf{P}(B = b_j / A = a_i, C = c_k) = \mathsf{P}(B = b_j / C = c_k).$$ □

§ 1.6. Аксиоматическое определение энтропии

Следуя К. Шеннону, мы определили энтропию формулой

$$H(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log_2 p_i.$$

Почему именно так, а не иначе? Возможны ли другие определения энтропии, отвечающие перечисленным выше в § 1.1 естественным требованиям? Есть два подхода к ответам на эти вопросы.

Во-первых, можно показывать теоретическую значимость и практическую полезность введённого определения энтропии, доказывая различные теоремы, важные как для развития теории, так и для приложений. Этого подхода мы в основном будем придерживаться в дальнейшем.

Во-вторых, на заданные вопросы можно отвечать в рамках так называемого аксиоматического подхода, то есть доказывать, что формула для энтропии строго выводится из небольшого числа аксиом, отражающих желательные свойства этой функции как количества информации, получаемой в среднем от наблюдения случайного исхода некоторого эксперимента. Имеется несколько таких доказательств (например, доказательства К. Шеннона, А. Я. Хинчина, Д. К. Фаддеева). Развитие этого вопроса было связано с естественным стремлением вывести формулу энтропии из возможно меньшего числа простых аксиом, однако доказательства становились более сложными. Мы приведём здесь один из возможных вариантов аксиоматического подхода.

Предположим, что неотрицательная функция H задана на объединении симплексов $\bigcup_{m\geqslant 2} S_{m-1}$. Сформулируем три аксиомы, которым должна удовлетворять эта функция.

A1 (непрерывность). Функция $H(p_1, \ldots, p_m)$ непрерывна по совокупности переменных $(p_1, \ldots, p_m)$ в симплексе S_{m-1} для любого $m \geqslant 2$.

A2 (монотонность). Последовательность $F(m) = H\left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$, $m \geqslant 2$, не убывает.

A3 (разложение). Если $m = m_1 + \ldots + m_k$, $m_i \geqslant 2$,

$$(p_1, \ldots, p_m) = (\underbrace{q_{11}, \ldots, q_{1\,m_1}}_{m_1}, \underbrace{q_{21}, \ldots, q_{2\,m_2}}_{m_2} \ldots, \underbrace{q_{k1}, \ldots, q_{k\,m_k}}_{m_k}),$$

$$\alpha_1 = q_{11} + \ldots + q_{1\,m_1} > 0, \ \ldots, \ \alpha_k = q_{k1} + \ldots + q_{k\,m_k} > 0,$$

то

$$H(p_1, \ldots, p_m) = H(\alpha_1, \ldots, \alpha_k) + \sum_{i=1}^{k} \alpha_i H\left(\frac{q_{i1}}{\alpha_i}, \ldots, \frac{q_{i\,m_i}}{\alpha_i}\right). \qquad (1.7)$$

Замечание 1.10. Информационный смысл аксиом A1, A2 очевиден. Аксиоме A3 можно дать такое нестрогое истолкование. Предположим, что множество $\mathscr{A}$ возможных значений сл. в. A разбивается на k подмножеств, мощности которых равны $m_1, \ldots, m_k$, $m = m_1 + \ldots + m_k$, и распределение сл. в. A представляется в виде

$$(p_1, \ldots, p_m) = (q_{11}, \ldots, q_{1\,m_1}, q_{21}, \ldots, q_{2\,m_2}, \ldots, q_{k1}, \ldots, q_{k\,m_k}),$$

причём суммарные вероятности подмножеств положительны,

$$\alpha_i = q_{i1} + \ldots + q_{i\,m_i} > 0,\ i = 1, \ldots, k.$$

Одно наблюдение (эксперимент) над сл. в. A можно провести двумя способами. Первый способ (за один шаг) даёт непосредственно значение сл. в. $A = a$, и среднее количество информации при этом равно значению $H(p_1, \ldots, p_m)$. Второй способ (за два шага) состоит в том, что на первом шаге определяется, в какое из k подмножеств попадает значение сл. в., а на втором шаге определяется элемент выбранного подмножества. При втором способе среднее количество информации на первом шаге равно $H(\alpha_1, \ldots, \alpha_k)$, и если выбрано i-е подмножество (с вероятностью α_i), то условное распределение вероятностей на этом подмножестве имеет вид $\left(\frac{q_{i1}}{\alpha_i}, \ldots, \frac{q_{i\,m_i}}{\alpha_i}\right)$, а среднее количество информации при выборе элемента из i-го подмножества равно $H\left(\frac{q_{i1}}{\alpha_i}, \ldots, \frac{q_{i\,m_i}}{\alpha_i}\right)$. Поэтому сумма $\sum_{i=1}^{k}$ в правой части (1.7) выражает среднее количество информации на втором шаге эксперимента, а всё соотношение (1.7) — равенство количеств информации, независимо от способа наблюдения над сл. в. A.

Теорема 1.2. *Если неотрицательная функция H, заданная на множестве $\bigcup_{m \geqslant 2} S_{m-1}$, удовлетворяет аксиомам* A1, A2 *и* A3, *то либо H тождественно равна нулю, либо*

$$H(\alpha_1, \ldots, \alpha_k) = -K \sum_{i=1}^{k} \alpha_i \log_2 \alpha_i,$$

где K — положительная постоянная.

Доказательство. Пусть $r \geqslant 2$, $n \geqslant 2$, $m = r^n$ и $m_1 = \ldots = m_r = r^{n-1}$. Тогда в силу аксиомы А3

$$F(r^n) = H\left(\frac{1}{r^n}, \ldots, \frac{1}{r^n}\right) = H\left(\frac{1}{r}, \ldots, \frac{1}{r}\right) + \sum_{i=1}^{r} \frac{1}{r} H\left(\frac{1}{r^{n-1}}, \ldots, \frac{1}{r^{n-1}}\right) =$$
$$= F(r) + F(r^{n-1}),$$

и методом индукции получаем равенство

$$F(r^n) = nF(r), \tag{1.8}$$

верное также при $n = 1$.

Если для некоторого $r_0 \geqslant 2$ верно $F(r_0) = 0$, то в силу неотрицательности функции H и монотонности последовательности $F(m)$ (аксиома А2) получаем, что $F(r) = 0$ для всех r, $2 \leqslant r \leqslant r_0$. С другой стороны, для любого $r > r_0$ существует такое $n \geqslant 1$, что $r \leqslant {r_0}^n$. Поэтому в силу аксиомы А2 и свойства (1.8) получаем $F(r) \leqslant F({r_0}^n) = nF(r_0) = 0$, т. е. снова $F(r) = 0$. Таким образом, доказано, что либо все члены последовательности $F(m) = H\left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$, $m \geqslant 2$, положительны, либо все они равны нулю.

Рассмотрим случай, когда последовательность $F(m)$, $m \geqslant 2$, положительна. Пусть даны натуральные числа $s \geqslant r \geqslant 2$, $u \geqslant 1$. Тогда существует, и притом единственное, целое $n \geqslant 1$ такое, что $r^n \leqslant s^u < r^{n+1}$. Отсюда в силу аксиомы А2 получаем неравенства

$$F(r^n) \leqslant F(s^u) \leqslant F(r^{n+1}),$$

из которых в силу (1.8) следует, что

$$nF(r) \leqslant uF(s) \leqslant (n+1)F(r),$$

или

$$\frac{n}{u} \leqslant \frac{F(s)}{F(r)} \leqslant \frac{n}{u} + \frac{1}{u}. \tag{1.9}$$

Для логарифмической функции, рассуждая аналогично, получим

$$\frac{n}{u} \leqslant \frac{\log_2 s}{\log_2 r} \leqslant \frac{n}{u} + \frac{1}{u}. \tag{1.10}$$

Из неравенств (1.9) и (1.10) вытекает, что

$$\left| \frac{F(s)}{F(r)} - \frac{\log_2 s}{\log_2 r} \right| \leqslant \frac{1}{u}$$

для любых $u \geqslant 1$. Следовательно, для любых $s \geqslant r \geqslant 2$ имеем

$$\frac{F(s)}{F(r)} - \frac{\log_2 s}{\log_2 r} = 0,$$

или

$$\frac{F(s)}{\log_2 s} = \frac{F(r)}{\log_2 r}.$$

Таким образом, для положительной последовательности $F(m)$, $m \geqslant 2$, отношение $\dfrac{F(m)}{\log_2 m}$ принимает некоторое постоянное значение $K > 0$, и мы получаем равенство

$$F(m) = K \log_2 m. \tag{1.11}$$

Пусть $\vec{\alpha} = (\alpha_1, \ldots, \alpha_k)$ — вектор из симплекса S_{k-1}. Рассмотрим вначале случай, когда все координаты $\alpha_1, \ldots, \alpha_k$ — положительные рациональные числа. Приведём числа α_i к общему знаменателю:

$$\alpha_i = \frac{m_i}{m}, \quad m_i \geqslant 2, \quad i = 1, \ldots, k, \quad \sum_{i=1}^{k} m_i = m.$$

Применяя аксиому А3, получим

$$\begin{aligned} F(m) &= H\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) = \\ &= H\left(\frac{m_1}{m}, \ldots, \frac{m_k}{m}\right) + \sum_{i=1}^{k} \frac{m_i}{m} H\left(\frac{1}{m_i}, \ldots, \frac{1}{m_i}\right) = \\ &= H(\vec{\alpha}) + \sum_{i=1}^{k} \alpha_i F(m_i), \end{aligned}$$

или

$$H(\vec{\alpha}) = F(m) - \sum_{i=1}^{k} \alpha_i F(m_i). \tag{1.12}$$

Из установленных выше свойств последовательности $F(m)$ и равенства (1.12) следует, что либо $H(\vec{\alpha}) = 0$, либо

$$\begin{aligned} H(\vec{\alpha}) &= K \log_2 m - \sum_{i=1}^{k} \alpha_i K \log_2 m_i = \\ &= -K \sum_{i=1}^{k} \alpha_i \log_2 \frac{m_i}{m} = -K \sum_{i=1}^{k} \alpha_i \log_2 \alpha_i. \end{aligned}$$

Перейдём к случаю, когда все $\alpha_1, \ldots, \alpha_k$ — рациональные, и среди них есть $s \geqslant 1$ чисел, равных нулю. Пусть a — наименьшее из чисел

$\alpha_1, \ldots, \alpha_k$, не равных нулю. Для целого $n > \frac{s}{(k-s)a}$ рассмотрим вектор $\vec{\alpha}^{(n)} = (\alpha_1^{(n)}, \ldots, \alpha_k^{(n)})$, где

$$\alpha_i^{(n)} = \begin{cases} \frac{1}{n}, & \text{если } \alpha_i = 0, \\ \alpha_i - \frac{s}{(k-s)n}, & \text{если } \alpha_i > 0. \end{cases}$$

Очевидно, что $\alpha_1^{(n)}, \ldots, \alpha_k^{(n)}$ — положительные рациональные числа, сумма которых равна 1. Как показано выше, либо $H(\vec{\alpha}^{(n)}) = 0$, либо

$$H(\vec{\alpha}^{(n)}) = -K \sum_{i=1}^{k} \alpha_i^{(n)} \log_2 \alpha_i^{(n)}.$$

Последовательность векторов $\vec{\alpha}^{(n)}$ сходится к вектору $\vec{\alpha}$, поэтому в силу аксиомы A1

$$H(\vec{\alpha}) = H\big(\lim_{n\to\infty} \vec{\alpha}^{(n)}\big) = \lim_{n\to\infty} H(\vec{\alpha}^{(n)}),$$

и мы получаем либо $\lim_{n\to\infty} H(\vec{\alpha}^{(n)}) = 0$, либо

$$\lim_{n\to\infty} H(\vec{\alpha}^{(n)}) = -K \lim_{n\to\infty} \sum_{i=1}^{k} \alpha_i^{(n)} \log_2 \alpha_i^{(n)} = -K \sum_{i=1}^{k} \alpha_i \log_2 \alpha_i.$$

Для завершения доказательства осталось рассмотреть случай, когда в векторе $\vec{\alpha} \in S_{k-1}$ некоторые координаты являются иррациональными числами. Тогда найдётся последовательность векторов $\vec{\alpha}^{(n)} \in S_{k-1}$, $n \geqslant 1$, с рациональными координатами, сходящаяся к $\vec{\alpha}$, и мы снова воспользуемся соотношением $H(\vec{\alpha}) = \lim_{n\to\infty} H(\vec{\alpha}^{(n)})$. □

Замечание 1.11. Отметим, что в теории информации рассматривают и другие определения энтропии. Например, энтропия Реньи порядка $\alpha \neq 1$ задаётся формулой

$$H_\alpha(p_1, \ldots, p_m) = \frac{1}{1-\alpha} \log_2\Big(\sum_{i=1}^{m} p_i^\alpha\Big),$$

а энтропия порядка α и типа β задаётся формулой

$$H(p_1, \ldots, p_m; \alpha, \beta) = \frac{1}{2^{1-\beta} - 1}\Big[\sum_{i=1}^{m} (p_i^\alpha)^{(\beta-1)/(\alpha-1)} - 1\Big],$$
$$\alpha > 0, \quad \beta > 0, \quad \alpha \neq \beta, \quad \alpha \neq 1, \quad \beta \neq 1.$$

§ 1.7. Задачи

1.1. Исход какого из опытов обладает большей неопределённостью:

а) внутри правильного треугольника случайно ставится точка, которая может оказаться внутри или вне вписанного в него круга;

б) внутри круга случайно ставится точка, которая может оказаться внутри или вне вписанного в него правильного треугольника.

1.2. В двух урнах имеется по 15 шаров, причём в первой урне 5 красных, 7 белых и 3 чёрных, а во второй — 4 красных, 4 белых и 7 чёрных. Из каждой урны случайно вынимают по одному шару и фиксируют цвет шара. Определить, для какой из урн исход опыта является более определённым.

1.3. В урне m чёрных шаров и $n - m$ белых. Случайно выбирают без возвращения два шара. Пусть X обозначает цвет первого шара, а Y — второго. Вычислить $H(X)$, $H(Y)$, $H(Y/X)$. Ответить на тот же вопрос, если выбор производится с возвращением.

1.4. В урне r красных шаров, b чёрных и w белых. Случайный эксперимент X состоит в том, что выбирают $k \geqslant 2$ шаров и фиксируют их цвета. В каком случае величина $H(X)$ больше — при выборе с возвращением или без возвращения?

1.5. Пусть сл. в. X имеет равномерное распределение на множестве целых чисел от 1 до 1050, и $Y = (Y_1, Y_2)$, где Y_1 — остаток от деления X на 5, а Y_2 — остаток от деления X на 7. Вычислить $H(X)$, $H(Y)$, $H(X/Y)$.

1.6. Пусть сл. в. A есть результат случайного бросания «честной» игральной кости, грани которой помечены числами 1, 2, 3, 4, 5, 6. Сл. в. X равна 1, если A чётна, и 0 в противном случае. Сл. в. Y равна 1, если A делится на 3, и 0 в противном случае. Вычислить $I(X; Y)$.

1.7. Производится два независимых случайных бросания «честной» игральной кости, грани которой помечены числами 1, 2, 3, 4, 5, 6. Сл. в. X есть результат первого бросания, а сл. в. Y равна сумме первого и второго бросаний. Вычислить $I(X; Y)$.

1.8. Вычислить энтропию следующих распределений:

а) биномиального: $p_k = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$,

б) геометрического: $p_k = p(1-p)^k$, $k = 0, 1, \ldots$,

в) пуассоновского: $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, \ldots$

В каждом случае найти максимальное значение энтропии при изменении параметра распределения.

1.9. Пусть сл. в. X есть число бросаний «честной» монеты до первого появления герба. Вычислить $H(X)$. Найти «кратчайшую» последовательность вопросов вида «принадлежит ли значение X множеству S», позволяющую определить значение X. Сравнить $H(X)$ и среднее число вопросов, необходимое для определения значения X.

1.10. Найти распределение $\vec{p} = (p_1, p_2, \ldots, p_m)$, доставляющее максимум энтропии $H(\vec{p})$ при условии, что величина $p_1 = \alpha$ фиксирована. Вычислить максимальное значение энтропии $H(\vec{p})$.

1.11. Для неотрицательной целочисленной сл. в. X найти распределение, доставляющее максимум энтропии $H(X)$ при условии, что математическое ожидание $\mathsf{E}\, X$ равно заданному числу $\lambda > 0$. Вычислить максимальное значение энтропии $H(X)$.

1.12. Полагая $p = (1+\delta)/2$, $|\delta| < 1$, получить представление функции двоичной энтропии h в виде степенного ряда

$$h\Big(\frac{1+\delta}{2}\Big) = 1 - \frac{1}{\ln 4} \sum_{k \geqslant 1} \frac{\delta^{2k}}{k(2k-1)} = 1 - \frac{1}{\ln 4}\Big(\delta^2 + \frac{\delta^4}{6} + \frac{\delta^6}{15} + \ldots\Big).$$

1.13. Доказать, что если целые $n, k \to +\infty$ так, что $\frac{k}{n} \to p$, $0 < p < 1$, то для биномиального коэффициента $\binom{n}{k}$ справедлива асимптотическая формула

$$\binom{n}{k} \sim 2^{nh(p)(1+o(1))} \cdot \Big(\sqrt{2\pi np(1-p)}\Big)^{-1}.$$

1.14. Вычислить математическое ожидание $\mathsf{E}\, h(\xi)$, если сл. в. ξ имеет равномерное распределение на отрезке $[0, 1]$.

1.15. Вычислить математическое ожидание $\mathsf{E}\, H(\vec{p})$, если случайный вектор $\vec{p} = (p_1, p_2, \ldots, p_m)$ равномерно распределён на симплексе S_{m-1}.

1.16. Показать, что энтропия дискретной сл. в. может быть бесконечной:

а) доказать, что величина $C_\alpha = \sum_{n=2}^{+\infty} \frac{1}{n(\log_2 n)^\alpha}$ конечна при $\alpha > 1$ и бесконечна при $0 \leqslant \alpha \leqslant 1$;

б) для сл. в. X с распределением

$$\mathrm{P}_X(n) = (C_\alpha n(\log_2 n)^\alpha)^{-1}, \quad n = 2, 3, \ldots,$$

доказать, что $H(X) < +\infty$ при $\alpha > 2$ и $H(X) = +\infty$ при $1 < \alpha \leqslant 2$.

1.17. Найти два таких распределения вероятностей $\vec{p} = (p_1, p_2, \ldots, p_m)$ и $\vec{q} = (q_1, q_2, \ldots, q_n)$, что $\vec{p} \neq \vec{q}$, $p_1 \geqslant \ldots \geqslant p_m > 0$, $q_1 \geqslant \ldots \geqslant q_n > 0$, и $H(\vec{p}) = H(\vec{q})$.

1.18. Доказать, что $\sum_{i=1}^m p_i \log_2 \frac{p_i}{q_i} \geqslant 0$ для произвольных распределений вероятностей $\vec{p} = (p_1, p_2, \ldots, p_m)$ и $\vec{q} = (q_1, q_2, \ldots, q_m)$.

1.19. Доказать, что

$$H(p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m) \leqslant$$
$$\leqslant H\Big(p_1, \ldots, \frac{p_i + p_j}{2}, \ldots, \frac{p_i + p_j}{2}, \ldots, p_m\Big).$$

1.20. Пусть распределения вероятностей $\vec{p}$ и $\vec{q}$ связаны формулой $\vec{q} = \vec{p}\, A$, где A — дважды стохастическая матрица. Доказать, что $H(\vec{q}) \geqslant H(\vec{p})$, и при этом равенство имеет место тогда и только тогда, когда матрица A — подстановочная.

1.21. Пусть $\vec{p} = (p_1, p_2, \ldots, p_n)$ — распределение вероятностей, $q_m = 1 - \sum_{j=1}^m p_j$, $0 \leqslant m \leqslant n$. Доказать неравенство

$$H(p_1, p_2, \ldots, p_n) \leqslant H(p_1, p_2, \ldots, p_m, q_m) + q_m \log_2(n - m).$$

При каких условиях это неравенство обращается в равенство?

1.22. Вычислить $H(X)$, $H(Y)$, $H(X/Y)$, $H(Y/X)$, $H(X, Y)$, $I(X; Y)$ для сл. в. X, Y, совместное распределение которых задано матрицей:

а) $\frac{1}{20} \cdot \begin{pmatrix} 3 & 2 \\ 9 & 6 \end{pmatrix}$; б) $\frac{1}{3} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$; в) $\frac{1}{8} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$;

г) $\frac{1}{36} \cdot \begin{pmatrix} 1 & 3 & 0 \\ 9 & 4 & 5 \\ 0 & 2 & 12 \end{pmatrix}$; д) $\frac{1}{24} \cdot \begin{pmatrix} 7 & 1 & 0 \\ 1 & 6 & 1 \\ 0 & 1 & 7 \end{pmatrix}$;

е) $\frac{1}{32}\cdot\begin{pmatrix}4&2&1&1\\2&4&1&1\\2&2&2&2\\8&0&0&0\end{pmatrix}$; ж) $\frac{1}{25}\cdot\begin{pmatrix}1&1&1&1&1\\2&1&2&0&0\\2&0&1&1&1\\0&3&0&2&0\\0&0&1&1&3\end{pmatrix}$.

1.23. Пусть сл. в. X и Y независимы и имеют распределения

$$X\sim\begin{pmatrix}0&1&3&4\\\frac{1}{8}&\frac{1}{8}&\frac{1}{4}&\frac{1}{2}\end{pmatrix},\quad Y\sim\begin{pmatrix}-2&2\\\frac{3}{8}&\frac{5}{8}\end{pmatrix}$$

и $Z=X+Y$. Вычислить $H(Z/Y)$.

1.24. Сл. в. X принимает четыре значения $\pm1,\pm2$ с равными вероятностями, и $Y=X^2$. Вычислить $I(X;Y)$ и ковариацию между X и Y.

1.25. Пусть сл. в. X и Y независимы, X равномерно распределена на множестве $\{1,2,\dots,8\}$, а $Y\sim\begin{pmatrix}1&2&3&\dots\\\frac{1}{2}&\frac{1}{4}&\frac{1}{8}&\dots\end{pmatrix}$. Вычислить $H(X+Y,X-Y)$.

1.26. Две равных по силам команды проводят серию из семи игр без ничьих, до четырёх побед одной из команд. Результаты игр считаем независимыми. Пусть X — сл. в., обозначающая результат серии игр, например, AAAA, BABABAB, BBBAAAA, и Y — сл. в., равная числу игр до окончания серии, $4\leqslant Y\leqslant 7$. Вычислить $H(X)$, $H(Y)$, $H(X/Y)$, $H(Y/X)$, $H(X,Y)$, $I(X;Y)$.

1.27. Пусть X и Y — зависимые сл. в., $H(X)=8$, $H(Y)=12$. Какие значения может принимать $H(Y/X)$, если $H(X/Y)$ изменяется в максимально возможных пределах?

1.28. Сл. в. X принимает три значения 0, 1 и 2 с равными вероятностями. Насколько большой может быть величина $I(X;Y)$, если Y — произвольная сл. в., принимающая два значения 0 и 1.

1.29. Сл. в. X принимает значения в конечном множестве чисел. Каковы общие соотношения в виде неравенств между $H(X)$ и $H(Y)$, если $Y=2^X$; $Y=\cos X$.

1.30. Сл. в. X принимает значения в конечном множестве $\mathscr{X}$. Пусть $\mathscr{S}\subseteq\mathscr{X}$, и $\mathsf{P}(X\in\mathscr{S})=\alpha$. Сл. в. Y равна 1, если $X\in\mathscr{S}$, и равна 0 в противном случае. Вычислить $I(X;Y)$.

1.31. Пусть X и Y — сл. в., и g — некоторая функция, заданная на множестве значений сл. в. Y. Доказать неравенство $H(X/g(Y)) \geqslant H(X/Y)$. При каких условиях это неравенство обращается в равенство?

1.32. Пусть X и Y — сл. в., принимающие значения в множествах $\mathscr{X}$ и $\mathscr{Y}$ соответственно. Доказать, что если $H(Y/X) = 0$, то Y является функцией от X, т. е. для любого элемента $x \in \mathscr{X}$ с условием $\mathsf{P}_X(x) > 0$ найдётся единственный элемент $y \in \mathscr{Y}$ такой, что $\mathsf{P}_{XY}(x, y) > 0$.

1.33. Является ли $I(X; Y)$ строго выпуклой вверх функцией от распределения сл. в. X?

1.34. Пусть X и Y — числовые сл. в. с конечными множествами значений, и $Z = X + Y$.

а) Доказать, что $H(Z/X) = H(Y/X)$;

б) доказать, что если X и Y независимы, то $H(Z) \geqslant H(Y)$, $H(Z) \geqslant H(X)$;

в) привести пример зависимых X и Y, для которых $H(Z) < H(X)$ и $H(Z) < H(Y)$;

г) при каких условиях $H(Z) = H(X) + H(Y)$?

1.35. Пусть X и Y — сл. в., принимающие значения в множествах $\mathscr{X}$ и $\mathscr{Y}$ соответственно. Доказать, что если пересечение $\mathscr{X} \cap \mathscr{Y}$ состоит ровно из двух элементов, то

$$H(Y/X) + H(X/Y) \geqslant h(\mathsf{P}(X \neq Y)).$$

В общем случае привести пример, когда это неравенство неверно.

1.36. Пусть X, Y, Z — независимые числовые сл. в. Доказать неравенство $I(X; X + Y + Z) \leqslant I(X; X + Y)$.

1.37. Для сл. в. X, Y, Z установить, верны ли следующие неравенства:

$$I((X, Y); Z) \geqslant I(X; Z),$$
$$H(X, Y/Z) \geqslant H(X/Z),$$
$$I(X; Z/Y) \geqslant I(Z; Y/X) - I(Z; Y) + I(Z; X),$$
$$H(X, Y, Z) - H(X, Y) \leqslant H(X, Z) - H(X).$$

Для верных неравенств описать условия их обращения в равенства.

1.38. Для сл. в. X, Y, Z доказать неравенства

$$H(X, Y/Z) \leqslant H(X/Z) + H(Y/Z), \qquad H(Z/X, Y) \leqslant H(Z/X)$$

и описать условия их обращения в равенства.

1.39. Привести примеры сл. в. X, Y, Z, для которых выполняются неравенства $I((X, Y); Z) < I(X; Z)$ или $I((X, Y); Z) > I(X; Z)$.

1.40. Пусть X, Y, Z — попарно независимые сл. в., каждая из которых принимает два значения 0 и 1 с равными вероятностями, так что $I(X; Y) = I(X; Z) = I(Y; Z) = 0$. Каково минимальное значение $H(X, Y, Z)$ при этом условии? Привести пример, когда это минимальное значение достигается.

1.41. Пусть сл. в. $X_1, \ldots, X_n$ независимы и одинаково распределены, $X_i \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$, $q = 1 - p$. Пусть $A = I(X_1, \ldots, X_n)$ — сл. в., принимающая значения $I(x_1, \ldots, x_n)$ с вероятностями $\prod_{i=1}^{n} \mathsf{P}_{X_i}(x_i)$. Найти математическое ожидание $\mathsf{E}\, A$ и дисперсию $\mathsf{D}\, A$ как функции от p, построить графики.

1.42. Пусть сл. в. $X_1, \ldots, X_n$ принимают два значения 0 и 1, причём вероятность $\mathsf{P}_{X_1, \ldots, X_n}(x_1, \ldots, x_n)$ равна 2^{-n+1}, если двоичная последовательность $(x_1, \ldots, x_n)$ имеет чётный вес, и равна 0 в противном случае. Вычислить

$$I(X_1; X_2), I(X_2; X_3/X_1), \ldots, I(X_{n-1}; X_n/X_1, \ldots, X_{n-1}).$$

1.43. Пусть $X = (X_1, \ldots, X_m)$ и $Y = (Y_1, \ldots, Y_n)$ — дискретные числовые сл. векторы, f и g — действительные функции от m и n переменных соответственно. Доказать неравенство $I(f(X); g(Y)) \leqslant I(X; Y)$.

1.44. Доказать правило цепочки для условной энтропии:

$$H(X_1, \ldots, X_n/Y) = \sum_{i=1}^{n} H(X_i/X_1, \ldots, X_{i-1}, Y).$$

1.45. Доказать правило цепочки для средней взаимной информации:

$$I(X_1, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y/X_1, \ldots, X_{i-1}).$$

Глава 2
Дискретные источники сообщений

§ 2.1. Вероятностные модели дискретных источников сообщений

Под дискретным источником сообщений, говоря нестрого, обычно подразумевают некоторое устройство, которое через определённые равные интервалы времени выдаёт (детерминированным или случайным образом) очередной символ (букву) сообщения, принадлежащий заданному конечному алфавиту.

Наша цель в этом параграфе — дать точное определение дискретного источника сообщений. Будем считать, что модель источника задана, если дан алфавит источника — конечное множество $\mathscr{A} = \{a_1, \ldots, a_m\}$, и для любой конечной последовательности $a^n = (a_{i_1}, \ldots, a_{i_n})$ букв из алфавита $\mathscr{A}$ определена вероятность $\mathsf{P}(a^n)$ появления этой последовательности на выходе источника. Мы построим вероятностную модель дискретного источника сообщений, следуя общей схеме А. Н. Колмогорова по определению вероятностной меры на бесконечномерном пространстве.

Пусть $\mathscr{A}^n$ обозначает n-ю декартову степень множества $\mathscr{A}$, т. е. множество всех слов длины n из символов алфавита $\mathscr{A}$ и

$$\mathscr{A}^\infty = \{\omega = (\omega_1, \omega_2, \ldots) \mid \omega_i \in \mathscr{A}\}$$

есть множество всех бесконечных последовательностей из символов данного алфавита. В соответствии с обозначениями § 1.2, построим вероятностное пространство $(\Omega, \mathscr{F}, \mathsf{P})$, для которого пространство элементарных событий $\Omega = \mathscr{A}^\infty$. Отметим, что множество $\mathscr{A}^\infty$ *несчётно*, и поэтому нужное вероятностное пространство не относится к дискретным, рассмотренным в § 1.2.

Для определения σ-алгебры $\mathscr{F}$ напомним понятия цилиндрических множеств. *Элементарным цилиндрическим множеством*

$$C(n_1, n_2, \ldots, n_k; a_{i_1}, \ldots, a_{i_k})$$

с параметрами $k \geqslant 1$, $1 \leqslant n_1 < n_2 < \ldots < n_k$, $a_{i_1}, \ldots, a_{i_k} \in \mathscr{A}$ называется множество всех последовательностей $\omega \in \mathscr{A}^\infty$, для которых

$\omega_{n_1} = a_{i_1}, \ldots, \omega_{n_k} = a_{i_k}$; другими словами, в последовательности ω на местах $n_1, \ldots, n_k$ записаны заданные символы $a_{i_1}, \ldots, a_{i_k}$, а на остальных местах — произвольные символы из алфавита $\mathscr{A}$.

Цилиндрическим множеством общего вида $C(n_1, n_2, \ldots, n_k; \mathscr{B})$ с параметрами k, $n_1, n_2, \ldots, n_k$ и $\mathscr{B} \subseteq \mathscr{A}^k$ называется множество всех последовательностей $\omega \in \mathscr{A}^\infty$, для которых $(\omega_{n_1}, \ldots, \omega_{n_k}) \in \mathscr{B}$; другими словами, в последовательности ω на местах $n_1, \ldots, n_k$ записано слово из множества $\mathscr{B}$.

Очевидно, что

$$C(n_1, n_2, \ldots, n_k; \mathscr{B}) = \bigcup_{(a_{i_1}, \ldots, a_{i_k}) \in \mathscr{B}} C(n_1, n_2, \ldots, n_k; a_{i_1}, \ldots, a_{i_k}).$$

В качестве σ-алгебры $\mathscr{F}$ будем рассматривать наименьшую по включению σ-алгебру подмножеств пространства Ω, содержащую все цилиндрические множества общего вида. Можно доказать, что такая наименьшая σ-алгебра состоит из множеств, полученных из цилиндрических множеств в результате всевозможных операций счётного объединения, счётного пересечения и разности.

Корректное определение вероятностной меры на σ-алгебре $\mathscr{F}$ является сложной задачей, и её решение обычно рассматривается в курсах теории меры и оснований теории вероятностей. Для изложения вероятностной модели дискретного источника сообщений мы ограничимся формулировкой теоремы о существовании и единственности искомой меры.

Определение 2.1. Предположим, что на σ-алгебре $\mathscr{F}$ задана вероятностная мера P. Тогда будем говорить, что задан дискретный случайный процесс (или бесконечная случайная последовательность)

$$A\colon \Omega \to \Omega, \quad A = (A_n)_{n=1}^{+\infty} = (A_1, A_2, \ldots),$$

принимающий значения в множестве $\Omega = \mathscr{A}^\infty$ и имеющий распределение $\mathsf{P}(A \in F) = \mathsf{P}(F)$, $F \in \mathscr{F}$.

Для начального отрезка сл. процесса A будем пользоваться также обозначением $A^k = (A_1, \ldots, A_k)$.

Определение 2.2. Функция $\mathsf{P}_{n_1, n_2, \ldots, n_k}$, заданная на классе всех подмножеств множества $\mathscr{A}^k$ формулой

$$\mathsf{P}_{n_1, n_2, \ldots, n_k}(\mathscr{B}) = \mathsf{P}(C(n_1, n_2, \ldots, n_k; \mathscr{B})), \quad \mathscr{B} \subseteq \mathscr{A}^k,$$

называется конечномерным распределением сл. процесса A на местах $n_1, \ldots, n_k$.

Замечание 2.1. 1. Нетрудно видеть, что дискретный сл. вектор

$$A_{n_1, n_2, \ldots, n_k} = (A_{n_1}, \ldots, A_{n_k}) \colon \Omega \to \mathscr{A}^k$$

имеет распределение

$$\mathsf{P}(A_{n_1, n_2, \ldots, n_k} \in \mathscr{B}) = \mathsf{P}_{n_1, n_2, \ldots, n_k}(\mathscr{B}), \quad \mathscr{B} \subseteq \mathscr{A}^k. \tag{2.1}$$

2. Для произвольного подмножества $\mathscr{B} \subseteq \mathscr{A}^k$ справедливо равенство

$$C(1, 2, \ldots, k, k+1; \mathscr{B} \times \mathscr{A}) = C(1, 2, \ldots, k; \mathscr{B}).$$

Вычисляя меру P от обеих частей этого равенства и принимая во внимание определение 2.2, приходим к равенству

$$\mathsf{P}_{1, 2, \ldots, k, k+1}(\mathscr{B} \times \mathscr{A}) = \mathsf{P}_{1, 2, \ldots, k}(\mathscr{B}). \tag{2.2}$$

Определение 2.3. Пусть для любого $k = 1, 2, \ldots$ на классе всех подмножеств множества $\mathscr{A}^k$ задана вероятностная мера $\mathsf{P}_{1, 2, \ldots, k}$. Последовательность мер $\mathsf{P}_{1, 2, \ldots, k}$, $k \geqslant 1$, называется согласованной, если для любых подмножеств $\mathscr{B} \subseteq \mathscr{A}^k$ справедливо равенство (2.2).

Таким образом, если на σ-алгебре $\mathscr{F}$, порождённой классом всех цилиндрических множеств, задана вероятностная мера P, то соответствующие конечномерные распределения $\mathsf{P}_{1, 2, \ldots, k}$ с необходимостью согласованы. Замечательный факт состоит в том, что свойство согласованности конечномерных распределений является также и достаточным для существования вероятностной меры P. Мы примем без доказательства следующую теорему.

Теорема 2.1. *Пусть для любого $k = 1, 2, \ldots$ на классе всех подмножеств множества $\mathscr{A}^k$ задана вероятностная мера $\mathsf{P}_{1, 2, \ldots, k}$ и последовательность мер $\mathsf{P}_{1, 2, \ldots, k}$, $k \geqslant 1$, является согласованной. Тогда на σ-алгебре $\mathscr{F}$, порождённой классом всех цилиндрических множеств, существует, и притом единственная, вероятностная мера P, продолжающая последовательность мер $\mathsf{P}_{1, 2, \ldots, k}$ в том смысле, что для любых подмножеств $\mathscr{B} \subseteq \mathscr{A}^k$ справедливо равенство*

$$\mathsf{P}(C(1, 2, \ldots, k; \mathscr{B})) = \mathsf{P}_{1, 2, \ldots, k}(\mathscr{B}).$$

Итак, согласно теореме 2.1, для задания вероятностной меры P на несчётном множестве $\mathscr{A}^\infty$ достаточно построить последовательность согласованных конечномерных дискретных распределений $\mathsf{P}_{1,2,\dots,k}$. Рассмотрим примеры таких построений.

Пример 2.1. *Дискретный источник без памяти* задаётся двумя параметрами $(\mathscr{A}, \vec{p})$, где $\mathscr{A} = \{a_1, \dots, a_m\}$ — алфавит источника и $\vec{p} = (p_1, \dots, p_m)$ — вероятностный вектор.

Пространство элементарных событий Ω и σ-алгебра $\mathscr{F}$ определяются стандартным образом, как описано выше. Конечномерные распределения $\mathsf{P}_{1,2,\dots,k}$ определяются на словах $(a_{i_1}, a_{i_2}, \dots, a_{i_k}) \in \mathscr{A}^k$ формулой

$$\mathsf{P}_{1,2,\dots,k}(a_{i_1}, a_{i_2}, \dots, a_{i_k}) = p_{i_1} p_{i_2} \cdots p_{i_k} \tag{2.3}$$

и продолжаются на подмножества $\mathscr{B} \subseteq \mathscr{A}^k$ по аддитивности.

Проверим, что определённые таким образом функции $\mathsf{P}_{1,2,\dots,k}$ действительно являются распределениями вероятностей и согласованы. Для этого достаточно установить, что $\mathsf{P}_{1,2,\dots,k}(\mathscr{A}^k) = 1$, и проверить равенство (2.2). Имеем

$$\begin{aligned}
\mathsf{P}_{1,2,\dots,k}(\mathscr{A}^k) &= \sum_{(a_{i_1},\dots,a_{i_k}) \in \mathscr{A}^k} p_{i_1} \cdots p_{i_k} = \\
&= \Bigl(\sum_{a_{i_1} \in \mathscr{A}} p_{i_1}\Bigr) \cdots \Bigl(\sum_{a_{i_k} \in \mathscr{A}} p_{i_k}\Bigr) = 1, \\
\mathsf{P}_{1,2,\dots,k,k+1}(\mathscr{B} \times \mathscr{A}) &= \sum_{(a_{i_1},\dots,a_{i_k}) \in \mathscr{B},\, a_{i_{k+1}} \in \mathscr{A}} p_{i_1} \cdots p_{i_k} p_{i_{k+1}} = \\
&= \sum_{(a_{i_1},\dots,a_{i_k}) \in \mathscr{B}} p_{i_1} \cdots p_{i_k} \sum_{a_{i_{k+1}} \in \mathscr{A}} p_{i_{k+1}} = \\
&= \sum_{(a_{i_1},\dots,a_{i_k}) \in \mathscr{B}} p_{i_1} \cdots p_{i_k} = \\
&= \mathsf{P}_{1,2,\dots,k}(\mathscr{B}).
\end{aligned}$$

Таким образом, в силу теоремы 2.1 искомая вероятностная мера P на σ-алгебре $\mathscr{F}$ существует, и притом единственная.

Отметим, что определённый здесь сл. процесс $A = (A_1, A_2, \dots)$ является последовательностью независимых и одинаково распределённых сл. величин.

Действительно,

$$\begin{aligned}\mathsf{P}(A_k = a_{i_k}) &= \sum_{(a_{i_1}, \ldots, a_{i_{k-1}}) \in \mathscr{A}^{k-1}} \mathsf{P}_{1, \ldots, k-1, k}(a_{i_1}, \ldots, a_{i_{k-1}}, a_{i_k}) = \\ &= \sum_{1 \leqslant i_1, \ldots, i_{k-1} \leqslant m} p_{i_1} \cdots p_{i_{k-1}} p_{i_k} = \\ &= \Big(\sum_{i_1=1}^{m} p_{i_1}\Big) \cdots \Big(\sum_{i_{k-1}=1}^{m} p_{i_{k-1}}\Big) p_{i_k} = \\ &= p_{i_k},\end{aligned}$$

так что согласно (2.3)

$$\mathsf{P}\big((A_1, \ldots, A_k) = (a_{i_1}, \ldots, a_{i_k})\big) = \prod_{s=1}^{k} \mathsf{P}(A_s = a_{i_s}).$$

Названия «источник без памяти» и «независимые сл. величины» по-разному выражают одну простую идею: очередная случайная буква A_k порождается источником независимо от предшествующих букв $A_1, A_2, \ldots, A_{k-1}$.

Пример 2.2. *Простой марковский источник* задаётся тремя параметрами $(\mathscr{A}, \vec{p}, Q)$, где $\mathscr{A} = \{a_1, \ldots, a_m\}$ — алфавит источника, $\vec{p} = (p_1, \ldots, p_m)$ — вероятностный вектор и $Q = (q_{ij})_{m \times m}$ — стохастическая матрица.

Пространство элементарных событий Ω и σ-алгебра $\mathscr{F}$ определяются стандартным образом, как описано выше. Конечномерные распределения $\mathsf{P}_{1, 2, \ldots, k}$ определяются условиями: если $k = 1$, то $\mathsf{P}_1(a_i) = p_i$, $1 \leqslant i \leqslant m$, а если $k > 1$, то для $1 \leqslant i_1, i_2, \ldots, i_k \leqslant m$

$$\mathsf{P}_{1, 2, \ldots, k}(a_{i_1}, a_{i_2}, a_{i_3}, \ldots, a_{i_{k-1}}, a_{i_k}) = p_{i_1} q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{k-1} i_k}, \tag{2.4}$$

и продолжаются на подмножества $\mathscr{B} \subseteq \mathscr{A}^k$ по аддитивности.

Проверим, что определённые таким образом функции $\mathsf{P}_{1, 2, \ldots, k}$ действительно являются распределениями вероятностей и согласованы. При $k = 1$ имеем

$$\mathsf{P}_1(\mathscr{A}) = \sum_{i=1}^{m} p_i = 1,$$

а при $k > 1$ равенство $\mathsf{P}_{1,2,\dots,k}(\mathscr{A}^k) = 1$ доказывается по индукции:

$$\begin{aligned}
\mathsf{P}_{1,2,\dots,k}(\mathscr{A}^k) &= \sum_{(a_{i_1},\dots,a_{i_k})\in\mathscr{A}^k} p_{i_1} q_{i_1 i_2} \cdots q_{i_{k-2} i_{k-1}} q_{i_{k-1} i_k} = \\
&= \sum_{(a_{i_1},\dots,a_{i_{k-1}})\in\mathscr{A}^{k-1}} p_{i_1} q_{i_1 i_2} \cdots q_{i_{k-2} i_{k-1}} \sum_{i_k=1}^{m} q_{i_{k-1} i_k} = \\
&= \sum_{(a_{i_1},\dots,a_{i_{k-1}})\in\mathscr{A}^{k-1}} p_{i_1} q_{i_1 i_2} \cdots q_{i_{k-2} i_{k-1}} = \\
&= \mathsf{P}_{1,2,\dots,k-1}(\mathscr{A}^{k-1}) = 1.
\end{aligned}$$

Условие согласованности также просто проверяется:

$$\begin{aligned}
\mathsf{P}_{1,2,\dots,k,k+1}(\mathscr{B}\times\mathscr{A}) &= \sum_{(a_{i_1},\dots,a_{i_k})\in\mathscr{B},\, a_{i_{k+1}}\in\mathscr{A}} p_{i_1} q_{i_1 i_2} \cdots q_{i_{k-1} i_k} q_{i_k i_{k+1}} = \\
&= \sum_{(a_{i_1},\dots,a_{i_k})\in\mathscr{B}} p_{i_1} q_{i_1 i_2} \cdots q_{i_{k-1} i_k} \sum_{i_{k+1}=1}^{m} q_{i_k i_{k+1}} = \\
&= \sum_{(a_{i_1},\dots,a_{i_k})\in\mathscr{B}} p_{i_1} q_{i_1 i_2} \cdots q_{i_{k-1} i_k} = \\
&= \mathsf{P}_{1,2,\dots,k}(\mathscr{B}).
\end{aligned}$$

Таким образом, и в этом примере в силу теоремы 2.1 искомая вероятностная мера P на σ-алгебре $\mathscr{F}$ существует, и притом единственная.

Нетрудно показать, что определённый в этом примере сл. процесс $A = (A_1, A_2, \dots)$ является простой однородной цепью Маркова. В отличие от источника без памяти, здесь очередная случайная буква A_k порождается в зависимости от одной предшествующей буквы A_{k-1}, и при фиксированном значении A_{k-1} буква A_k не зависит от $A_1, A_2, \dots, A_{k-2}$; другими словами, источник «имеет память на одну букву».

Отметим также, что простой марковский источник $(\mathscr{A}, \vec{p}, Q)$ является обобщением источника без памяти; действительно, если в матрице Q все строки одинаковы и равны $\vec{p}$, то возвращаемся к источнику без памяти $(\mathscr{A}, \vec{p})$.

Пример 2.3. *Марковский источник с глубиной зависимости* $s \geqslant 1$ задаётся параметрами $(\mathscr{A}, \mathsf{P}_{1,2,\dots,s}, Q^{(s)})$, где $\mathscr{A} = \{a_1, \dots, a_m\}$ — алфавит источника, $\mathsf{P}_{1,2,\dots,s}$ — произвольное распределение на множестве $\mathscr{A}^s$ и $Q^{(s)} = (q_{i_1,\dots,i_s;\,j})$, $1 \leqslant i_1, i_2, \dots, i_s, j \leqslant m$, — такая матрица

$m^s \times m$ из неотрицательных чисел, что для любых $i_1, i_2, \ldots, i_s$ имеет место равенство $\sum_{j=1}^{m} q_{i_1,\ldots,i_s;\,j} = 1$.

Конечномерные распределения $\mathsf{P}_{1,2,\ldots,k}$ при $k \neq s$ определяются условиями: если $k < s$, то

$$\mathsf{P}_{1,\ldots,k}(a_{i_1}, \ldots, a_{i_k}) = \sum_{(a_{i_{k+1}},\ldots,a_{i_s})\in\mathscr{A}^{s-k}} \mathsf{P}_{1,\ldots,s}(a_{i_1}, \ldots, a_{i_k}, a_{i_{k+1}}, \ldots, a_{i_s}),$$

а если $k > s$, то

$$\mathsf{P}_{1,\ldots,k}(a_{i_1}, \ldots, a_{i_s}, a_{i_{s+1}}, \ldots, a_{i_k}) = \\ = \mathsf{P}_{1,\ldots,s}(a_{i_1}, \ldots, a_{i_s}) q_{i_1,\ldots,i_s;\,i_{s+1}} q_{i_2,\ldots,i_{s+1};\,i_{s+2}} \cdots q_{i_{k-s},\ldots,i_{k-1};\,i_k}.$$

Оставим читателю в качестве несложного упражнения проверить, что определённые таким образом функции $\mathsf{P}_{1,2,\ldots,k}$ действительно являются распределениями вероятностей и согласованы.

Марковский источник с глубиной зависимости $s \geqslant 1$ является очевидным обобщением простого марковского источника. Можно сказать, что такой источник «имеет память на s предшествующих букв».

§ 2.2. Энтропия стационарных источников сообщений

Важным подклассом дискретных источников сообщений являются *стационарные* источники. Существуют различные эквивалентные определения стационарности; для наших целей будет удобно принять следующее определение.

Определение 2.4. Дискретный источник сообщений $(\Omega, \mathscr{F}, \mathsf{P})$ называется стационарным, если для любых $t \geqslant 1$, $1 \leqslant n_1 < n_2 < \ldots < n_k$, $\mathscr{B} \subseteq \mathscr{A}^k$ справедливо равенство

$$\mathsf{P}_{n_1+t,\,n_2+t,\ldots,\,n_k+t}(\mathscr{B}) = \mathsf{P}_{n_1,\,n_2,\ldots,\,n_k}(\mathscr{B}). \tag{2.5}$$

Замечание 2.2. 1. Можно сказать, что для стационарного источника конечномерные распределения инвариантны относительно сдвига по времени, или, что то же самое, сл. векторы $A_{n_1,\,n_2,\ldots,n_k}$ и $A_{n_1+t,\,n_2+t,\ldots,\,n_k+t}$ одинаково распределены.

2. Математическое понятие стационарного источника сообщений отражает свойство многих реальных источников сообщений, характеристики которых не меняются с течением времени или становятся практически неизменными, когда источник после некоторого начального периода функционирования входит в так называемый стационарный режим.

Утверждение 2.1. *Дискретный источник без памяти стационарен.*

Доказательство утверждения сводится к непосредственной проверке определения. Действительно, так как $A = (A_1, A_2, \ldots)$ является последовательностью независимых и одинаково распределённых сл. величин, для любого $(a_{i_1}, \ldots, a_{i_k}) \in \mathscr{A}^k$ вероятность

$$\mathsf{P}_{n_1, n_2, \ldots, n_k}(a_{i_1}, \ldots, a_{i_k}) = \mathsf{P}\big((A_{n_1}, \ldots, A_{n_k}) = (a_{i_1}, \ldots, a_{i_k})\big) = \\ = \prod_{s=1}^{k} \mathsf{P}(A_{n_s} = a_{i_s}) = p_{i_1} \ldots p_{i_k}$$

не зависит от $(n_1, n_2, \ldots, n_k)$, и в силу аддитивности это свойство сохраняется для любых подмножеств $\mathscr{B} \subseteq \mathscr{A}^k$. □

Утверждение 2.2. *Простой марковский источник $(\mathscr{A}, \vec{p}, Q)$ стационарен тогда и только тогда, когда $\vec{p} \cdot Q = \vec{p}$.*

Доказательство. Предположим, что источник стационарен. Формула (2.5) при $k = n_1 = t = 1$, $\mathscr{B} = \{a_j\}$, принимает вид $\mathsf{P}_1(a_j) = \mathsf{P}_2(a_j)$. При этом в силу (2.4)

$$\mathsf{P}_1(a_j) = p_j, \quad \mathsf{P}_2(a_j) = \sum_{a_i \in \mathscr{A}} \mathsf{P}_{1,2}(a_i, a_j) = \sum_{i=1}^{m} p_i q_{ij}.$$

Таким образом, для любого j, $1 \leqslant j \leqslant m$, справедливо равенство $p_j = \sum_{i=1}^{m} p_i q_{ij}$, что и требовалось.

С другой стороны, покажем, что равенство $\vec{p} \cdot Q = \vec{p}$ влечёт стационарность источника. Для этого достаточно установить, что для любых $1 \leqslant n_1 < n_2 < \ldots < n_k$, $\mathscr{B} \subseteq \mathscr{A}^k$ справедливо равенство

$$\mathsf{P}_{n_1, n_2, \ldots, n_k}(\mathscr{B}) = \mathsf{P}_{n_1+1, n_2+1, \ldots, n_k+1}(\mathscr{B}),$$

из которого равенство (2.5) вытекает по индукции. В силу аддитивности вероятностной меры, нам достаточно для любых $(a_{i_1}, \ldots, a_{i_k}) \in$

$\in \mathscr{A}^k$ проверить равенство

$$\mathsf{P}_{n_1, n_2, \ldots, n_k}(a_{i_1}, \ldots, a_{i_k}) = \mathsf{P}_{n_1+1, n_2+1, \ldots, n_k+1}(a_{i_1}, \ldots, a_{i_k}).$$

Имеем

$$\mathsf{P}_{n_1, n_2, \ldots, n_k}(a_{i_1}, \ldots, a_{i_k}) = \sum_{\substack{(y_1, y_2, \ldots, y_{n_k}) \in \mathscr{A}^{n_k} \\ y_{n_1} = a_{i_1}, \ldots, y_{n_k} = a_{i_k}}} \mathsf{P}_{1, 2, \ldots, n_1, \ldots, n_k}(y_1, y_2, \ldots, y_{n_1}, \ldots, y_{n_k}). \quad (2.6)$$

Рассмотрим одно слагаемое в (2.6) и предположим, что $y_1 = a_j$, $y_2 = a_s$. Тогда $\mathsf{P}_{1, 2, \ldots, n_1, \ldots, n_k}(y_1, y_2, \ldots, y_{n_1}, \ldots, y_{n_k})$ представляется в виде произведения, первые сомножители в котором $p_j\, q_{js}$. Подставляя вместо p_j сумму $\sum_{i=1}^{m} p_i q_{ij}$, получим

$$\mathsf{P}_{1, 2, \ldots, n_1, \ldots, n_k}(y_1, y_2, \ldots, y_{n_1}, \ldots, y_{n_k}) = \sum_{a_i \in \mathscr{A}} \mathsf{P}_{1, 2, \ldots, n_1+1, \ldots, n_k+1}(a_i, y_1, y_2, \ldots, y_{n_1}, \ldots, y_{n_k}).$$

Возвращаясь к (2.6), получим

$$\mathsf{P}_{n_1, n_2, \ldots, n_k}(a_{i_1}, \ldots, a_{i_k}) = \sum_{\substack{(a_i, y_1, y_2, \ldots, y_{n_k}) \in \mathscr{A}^{n_k+1} \\ y_{n_1} = a_{i_1}, \ldots, y_{n_k} = a_{i_k}}} \mathsf{P}_{1, 2, \ldots, n_1+1, \ldots, n_k+1}(a_i, y_1, y_2, \ldots, y_{n_1}, \ldots, y_{n_k}) =$$

$$= \mathsf{P}_{n_1+1, n_2+1, \ldots, n_k+1}(a_{i_1}, \ldots, a_{i_k}),$$

что и требовалось. □

Определение 2.5. Предположим, что задан некоторый дискретный источник сообщений $(\Omega, \mathscr{F}, \mathsf{P})$, или, что то же самое, определён дискретный сл. процесс $A = (A_1, A_2, \ldots)$.

Энтропией на одну букву k-буквенного сообщения A^k называется величина

$$H_k = \frac{1}{k} H(A^k).$$

Условной энтропией k-й буквы A_k $(k > 1)$ при условии предшествующих букв $A^{k-1} = (A_1, \ldots, A_{k-1})$ называется величина

$$H^{(k)} = H(A_k / A^{k-1}).$$

Кроме того, полагаем $H^{(1)} = H_1$.

Замечание 2.3. В соответствии с определениями 1.6, 2.2 и 2.5

$$H_k = -\frac{1}{k} \sum_{(a_{i_1}, \ldots, a_{i_k}) \in \mathscr{A}^k} \mathsf{P}_{1,\ldots,k}(a_{i_1}, \ldots, a_{i_k}) \log_2 \mathsf{P}_{1,\ldots,k}(a_{i_1}, \ldots, a_{i_k}),$$

$$H^{(k)} = - \sum_{(a_{i_1}, \ldots, a_{i_k}) \in \mathscr{A}^k} \mathsf{P}_{1,\ldots,k}(a_{i_1}, \ldots, a_{i_k}) \times$$

$$\times \log_2 \mathsf{P}\big(A_k = a_{i_k} \,/\, A^{k-1} = (a_{i_1}, \ldots, a_{i_{k-1}})\big).$$

Утверждение 2.3. *Справедливы соотношения*

$$H^{(k)} = kH_k - (k-1)H_{k-1}, \quad k > 1, \tag{2.7}$$

$$H_k = \frac{1}{k}(H^{(1)} + H^{(2)} + \ldots + H^{(k)}). \tag{2.8}$$

Доказательство. По свойству аддитивности энтропии (утверждение 1.4, с. 27)

$$H^{(k)} = H(A_k/A^{k-1}) = H(A^k) - H(A^{k-1}) = kH_k - (k-1)H_{k-1}.$$

По правилу цепочки (утверждение 1.5, с. 27)

$$kH_k = H(A^k) = H(A_1) + H(A_2/A^1) + \ldots + H(A_k/A^{k-1}) =$$
$$= H^{(1)} + H^{(2)} + \ldots + H^{(k)}. \qquad \square$$

Определение 2.6. Энтропией источника сообщений называется число

$$H_\infty = \lim_{k \to \infty} H_k,$$

если этот предел существует и конечен.

Замечание 2.4. 1. Величина H_∞ является мерой количества информации, приходящейся в среднем на один символ длинного сообщения, порождаемого источником, и иногда называется также скоростью порождения информации для данного источника сообщений.

2. Из свойства полуаддитивности энтропии (следствие 1.1, с. 28) и верхней оценки энтропии (п. 4 утверждения 1.2, с. 23) следует, что

$$H_k = \frac{1}{k} H(A^k) \leqslant \frac{1}{k} \sum_{i=1}^{k} H(A_i) \leqslant \log_2 m,$$

и поэтому если энтропия источника существует, то $H_\infty \leqslant \log_2 m$. При этом равенство $H_\infty = \log_2 m$ справедливо, например, когда

сл. величины $A_1, A_2, \ldots$ независимы и имеют одинаковое равномерное распределение.

Теорема 2.2. *Если дискретный источник сообщений стационарен, то обе последовательности H_k и $H^{(k)}$ не возрастают и сходятся к общему конечному пределу H_∞.*

Доказательство. Согласно утверждению 1.7 (с. 28), при $k > 1$

$$H^{(k)} = H(A_k/\, A_1, A_2, \ldots, A_{k-1}) \leqslant H(A_k/\, A_2, A_3, \ldots, A_{k-1}).$$

В силу стационарности источника

$$H(A_k/\, A_2, A_3, \ldots, A_{k-1}) = H(A_{k-1}/\, A_1, A_2, \ldots, A_{k-2}) = H^{(k-1)}.$$

Таким образом, $0 \leqslant H^{(k)} \leqslant H^{(k-1)}$. Так как неотрицательная последовательность $H^{(k)}$ не возрастает, она сходится к конечному пределу.

Применяя формулу (2.8), получим

$$\begin{aligned} H_k - H_{k+1} &= \frac{1}{k}\sum_{i=1}^{k} H^{(i)} - \frac{1}{k+1}\sum_{i=1}^{k+1} H^{(i)} = \\ &= \Big(\frac{1}{k} - \frac{1}{k+1}\Big)\sum_{i=1}^{k} H^{(i)} - \frac{1}{k+1}H^{(k+1)} = \\ &= \frac{1}{k(k+1)}\Big(\sum_{i=1}^{k} H^{(i)} - kH^{(k+1)}\Big) = \\ &= \frac{1}{k(k+1)}\sum_{i=1}^{k}(H^{(i)} - H^{(k+1)}). \end{aligned}$$

Так как последовательность $H^{(k)}$ не возрастает, все слагаемые в последней сумме неотрицательны, и поэтому $0 \leqslant H_{k+1} \leqslant H_k$. Поскольку последовательность H_k неотрицательна и не возрастает, она также сходится к конечному пределу.

Равенство $\lim\limits_{k\to\infty} H_k = \lim\limits_{k\to\infty} H^{(k)}$ вытекает из формулы (2.8) и приводимой ниже леммы 2.1. □

Лемма 2.1. *Если для числовой последовательности x_n, $n = 1, 2, \ldots$, существует конечный предел $\lim\limits_{n\to\infty} x_n = x$, то последовательность средних арифметических $\frac{1}{n}\sum_{i=1}^{n} x_i$ также сходится к числу x.*

Доказательство. Так как последовательность x_n сходится к конечному пределу, она ограничена, т. е. существует такое число $M > 0$, что $|x_n| \leqslant M$ для всех $n = 1, 2, \ldots$ По определению предела, для любого $\varepsilon > 0$ найдётся такое натуральное N, что при $n \geqslant N$ справедливо неравенство $|x_n - x| \leqslant \varepsilon$. Поэтому при $n \geqslant N$

$$\left|\frac{1}{n}\sum_{i=1}^{n} x_i - x\right| \leqslant \frac{1}{n}\sum_{i=1}^{N} |x_i - x| + \frac{1}{n}\sum_{i=N+1}^{n} |x_i - x| \leqslant \frac{N}{n}(M + |x|) + \frac{n-N}{n}\varepsilon.$$

Если дополнительно $n \geqslant N(M + |x|)/\varepsilon$, то $\left|\frac{1}{n}\sum_{i=1}^{n} x_i - x\right| \leqslant 2\varepsilon$, что и требовалось. □

В некоторых случаях значение предельной энтропии источника сообщений нетрудно вычислить, поскольку предельное значение достигается на конечном шаге.

Утверждение 2.4. *Для стационарного марковского источника с глубиной зависимости $s \geqslant 1$*

$$H^{(1)} \geqslant H^{(2)} \geqslant \ldots \geqslant H^{(s)} \geqslant H^{(s+1)} = H^{(s+2)} = \ldots = H_\infty.$$

Доказательство. Невозрастание последовательности $H^{(k)}$ и существование предельной энтропии вытекают из теоремы 2.2. Докажем равенство $H^{(s+1)} = H^{(s+t)}$, $t \geqslant 1$. Из определения марковского источника с глубиной зависимости $s \geqslant 1$ следует, что случайная буква A_{s+t} при условии фиксированных предшествующих s букв $A_t, A_{t+1}, \ldots, A_{t+s-1}$ не зависит от букв $A_1, A_2, \ldots, A_{t-1}$, поэтому в силу утверждения 1.7

$$\begin{aligned} H^{(s+t)} &= H(A_{s+t}/A_1, A_2, \ldots, A_{t-1}, A_t, A_{t+1}, \ldots, A_{t+s-1}) = \\ &= H(A_{s+t}/A_t, A_{t+1}, \ldots, A_{t+s-1}), \end{aligned}$$

а в силу стационарности источника

$$H(A_{s+t}/A_t, A_{t+1}, \ldots, A_{t+s-1}) = H(A_{s+1}/A_1, A_2, \ldots, A_s) = H^{(s+1)}. \quad \square$$

Для дискретного источника без памяти и простого марковского источника вычисление предельной энтропии не представляет труда.

Утверждение 2.5. *Для дискретного источника без памяти $(\mathscr{A}, \vec{p})$ предельная энтропия существует и равна $H_\infty = H(\vec{p})$.*

Доказательство. Так как сл. величины $A_1, A_2, \ldots$ независимы и одинаково распределены,

$$H_k = \frac{1}{k} H(A^k) = \frac{1}{k} \sum_{i=1}^{k} H(A_i) = H(A_1) = H(\vec{p}). \qquad \square$$

Утверждение 2.6. *Для стационарного простого марковского источника* $(\mathscr{A}, \vec{p}, Q)$ *предельная энтропия существует и равна*

$$H_\infty = \sum_{i=1}^{m} p_i H(\vec{Q}_i),$$

где $\vec{Q}_i$ *означает* i*-ю строку матрицы* Q.

Доказательство. В силу утверждения 2.4 $H_\infty = H^{(2)} = H(A_2 / A_1)$. Согласно определениям 1.6 и 2.2,

$$H(A_2 / A_1) = - \sum_{(a_i, a_j) \in \mathscr{A}^2} \mathsf{P}_{1,2}(a_i, a_j) \log_2 \mathsf{P}(A_2 = a_j / A_1 = a_i).$$

По формуле (2.4)

$$\mathsf{P}_{1,2}(a_i, a_j) = p_i q_{ij}, \quad \mathsf{P}(A_2 = a_j / A_1 = a_i) = \frac{\mathsf{P}_{1,2}(a_i, a_j)}{\mathsf{P}_1(a_i)} = q_{ij},$$

так что

$$H(A_2/A_1) = - \sum_{i=1}^{m} p_i \sum_{j=1}^{m} q_{ij} \log_2 q_{ij} = \sum_{i=1}^{m} p_i H(\vec{Q}_i). \qquad \square$$

§ 2.3. Свойство асимптотической равнораспределённости

Для дискретных источников сообщений, обладающих предельной энтропией, в ряде случаев имеют место некоторые важные асимптотические свойства, т. е. свойства множества всех сообщений заданной конечной длины при её неограниченном увеличении.

Определение 2.7. Условимся говорить, что дискретный источник сообщений $(\Omega, \mathscr{F}, \mathsf{P})$ (или, что то же самое, дискретный сл. процесс $A = (A_1, A_2, \ldots)$) обладает свойством *асимптотической равнораспределённости*, если для него существует предельная энтропия H_∞ и для любых сколь угодно малых $\varepsilon > 0$, $\delta > 0$ найдётся такое натуральное число N, зависящее от ε, δ и меры P, что для всех $n \geqslant N$

множество $\mathscr{A}^n$ допускает такое разбиение на два непересекающихся подмножества $\mathscr{A}^n = \mathscr{A}_{n,\varepsilon} \sqcup \mathscr{A}'_{n,\varepsilon}$, что выполнены следующие свойства:

1) для всех $a^n \in \mathscr{A}_{n,\varepsilon}$ справедливо неравенство

$$\left|\frac{1}{n} I(a^n) - H_\infty\right| \leqslant \varepsilon, \tag{2.9}$$

где $I(a^n) = -\log_2 \mathsf{P}(A^n = a^n) = -\log_2 \mathsf{P}_{1,2,\dots,n}(a^n)$ — собственная информация, заключённая в событии $\{A^n = a^n\}$;

2) справедливо неравенство

$$\mathsf{P}_{1,2,\dots,n}(\mathscr{A}'_{n,\varepsilon}) = \sum_{a^n \in \mathscr{A}'_{n,\varepsilon}} \mathsf{P}_{1,2,\dots,n}(a^n) \leqslant \delta. \tag{2.10}$$

Замечание 2.5. Напомним понятие *сходимости по вероятности* последовательности сл. величин. Говорят, что последовательность сл. величин ξ_n, $n \geqslant 1$, сходится по вероятности к сл. величине ξ, если для любых сколь угодно малых $\varepsilon > 0$, $\delta > 0$ найдётся такое натуральное число N, зависящее от ε и δ, что для всех $n \geqslant N$ справедливо неравенство

$$\mathsf{P}(|\xi_n - \xi| > \varepsilon) \leqslant \delta. \tag{2.11}$$

Сходимость по вероятности обозначают в виде $\xi_n \xrightarrow{\mathsf{P}} \xi$.

Так же как в п. 3 замечания 1.5 рассмотрим дискретную сл. в. $I(A^n)$, принимающую значения $I(a^n)$ с вероятностями $\mathsf{P}_{1,2,\dots,n}(a^n)$, $a^n \in \mathscr{A}^n$. Положим $\xi_n = \frac{1}{n} I(A^n)$, $\xi = H_\infty$. Тогда п. 1 определения 2.7 означает, что событие $\mathscr{A}_{n,\varepsilon}$ является подмножеством события $|\xi_n - \xi| \leqslant \varepsilon$, а п. 2 означает выполнение неравенств

$$\mathsf{P}(|\xi_n - \xi| > \varepsilon) \leqslant \mathsf{P}_{1,2,\dots,n}(\mathscr{A}'_{n,\varepsilon}) \leqslant \delta.$$

Поэтому если выполнено свойство асимптотической равнораспределённости, то $\frac{1}{n} I(A^n) \xrightarrow{\mathsf{P}} H_\infty$.

С другой стороны, пусть имеет место сходимость $\frac{1}{n} I(A^n) \xrightarrow{\mathsf{P}} H_\infty$. Для всех $n \geqslant 1$ к множеству $\mathscr{A}_{n,\varepsilon}$ отнесём те и только те последовательности a^n, для которых выполнено неравенство (2.9), т. е. положим

$$\mathscr{A}_{n,\varepsilon} = \left\{a^n \in \mathscr{A}^n : \left|\frac{1}{n} I(a^n) - H_\infty\right| \leqslant \varepsilon\right\}, \quad \mathscr{A}'_{n,\varepsilon} = \mathscr{A}^n \setminus \mathscr{A}_{n,\varepsilon}.$$

Тогда условие 1 определения 2.7 выполнено по построению множества $\mathscr{A}_{n,\varepsilon}$, и для всех достаточно больших n выполнено неравенство (2.11), которое равносильно условию 2 определения определения 2.7.

Таким образом, свойство асимптотической равнораспределённости равносильно сходимости $\frac{1}{n} I(A^n) \xrightarrow{\mathsf{P}} H_\infty$.

Замечание 2.6. Неравенство (2.9) можно переписать в виде

$$2^{-n(H_\infty+\varepsilon)} \leqslant \mathsf{P}_{1,2,\dots,n}(a^n) \leqslant 2^{-n(H_\infty-\varepsilon)}, \tag{2.12}$$

а неравенство (2.10) — в виде

$$\mathsf{P}_{1,2,\dots,n}(\mathscr{A}_{n,\varepsilon}) \geqslant 1-\delta, \tag{2.13}$$

т. е. все последовательности $a^n \in \mathscr{A}_{n,\varepsilon}$ имеют вероятности, которые отличаются от 2^{-nH_∞} не более чем в $2^{-n\varepsilon}$ раз, а их суммарная вероятность не меньше $1-\delta$. Поэтому говорят, что $\mathscr{A}_{n,\varepsilon}$ является $(1-\delta)$-высоковероятным множеством *ε-типичных* последовательностей.

Утверждение 2.7. *Пусть для дискретного источника сообщений существует предельная энтропия H_∞ и $\mathscr{A}_{n,\varepsilon}$ есть множество всех последовательностей $a^n \in \mathscr{A}^n$, для которых выполнено условие* (2.9), *где $\varepsilon > 0$. Тогда справедливо неравенство*

$$|\mathscr{A}_{n,\varepsilon}| \leqslant 2^{n(H_\infty+\varepsilon)}. \tag{2.14}$$

Кроме того, если источник обладает свойством асимптотической равнораспределённости, то для любого $\delta > 0$ при всех достаточно больших n справедливо неравенство

$$|\mathscr{A}_{n,\varepsilon}| \geqslant (1-\delta)2^{n(H_\infty-\varepsilon)}. \tag{2.15}$$

Доказательство. Согласно левому неравенству (2.12),

$$1 \leqslant 2^{n(H_\infty+\varepsilon)} \mathsf{P}_{1,2,\dots,n}(a^n).$$

Поэтому

$$|\mathscr{A}_{n,\varepsilon}| = \sum_{a^n \in \mathscr{A}_{n,\varepsilon}} 1 \leqslant 2^{n(H_\infty+\varepsilon)} \sum_{a^n \in \mathscr{A}_{n,\varepsilon}} \mathsf{P}_{1,2,\dots,n}(a^n) \leqslant 2^{n(H_\infty+\varepsilon)}.$$

Согласно правому неравенству (2.12),

$$1 \geqslant 2^{n(H_\infty-\varepsilon)} \mathsf{P}_{1,2,\dots,n}(a^n).$$

Поэтому если выполнено свойство асимптотической равнораспределённости, то при всех достаточно больших n

$$|\mathscr{A}_{n,\varepsilon}| = \sum_{a^n \in \mathscr{A}_{n,\varepsilon}} 1 \geqslant 2^{n(H_\infty - \varepsilon)} \sum_{a^n \in \mathscr{A}_{n,\varepsilon}} \mathsf{P}_{1,2,\ldots,n}(a^n) =$$

$$= 2^{n(H_\infty - \varepsilon)} \mathsf{P}_{1,2,\ldots,n}(\mathscr{A}_{n,\varepsilon}) \geqslant (1-\delta) 2^{n(H_\infty - \varepsilon)}. \qquad \square$$

Замечание 2.7. Если выполнено свойство асимптотической равнораспределённости, то, как следует из доказанного выше утверждения, множество $\mathscr{A}_{n,\varepsilon}$ всех *ε-типичных* последовательностей при достаточно больших n имеет мощность, близкую к 2^{nH_∞}. Для источника без памяти с равномерным распределением имеем $\mathscr{A}_{n,\varepsilon} = \mathscr{A}^n$, $H_\infty = \log_2 m$, $|\mathscr{A}_{n,\varepsilon}| = m^n = 2^{nH_\infty}$. Для любого источника с условием $H_\infty < \log_2 m$ получим

$$\frac{|\mathscr{A}_{n,\varepsilon}|}{|\mathscr{A}^n|} \leqslant \frac{2^{n(H_\infty + \varepsilon)}}{m^n} = 2^{-n(\log_2 m - H_\infty - \varepsilon)} \to 0,$$

если $0 < \varepsilon < \log_2 m - H_\infty$; таким образом, суммарная вероятность множества $\mathscr{A}_{n,\varepsilon}$ близка к единице, но его доля в множестве $\mathscr{A}^n$ всех последовательностей длины n близка к нулю.

Теорема 2.3. *Дискретный источник без памяти обладает свойством асимптотической равнораспределённости.*

Доказательство. Рассмотрим два способа доказательства теоремы, один из которых сводится к применению закона больших чисел, а второй связан с несложными оценками для биномиально распределённых сл. величин на основе неравенства Чебышёва. Второй способ более громоздкий, но даёт возможность явно оценить величину N из определения 2.7.

1. Согласно замечанию 2.5, достаточно доказать, что последовательность сл. величин $\frac{1}{n} I(A^n)$, $n \geqslant 1$, сходится по вероятности к числу H_∞.

Из определения количества собственной информации и формулы (2.3) следует равенство $I(A^n) = \sum_{i=1}^n I(A_i)$, и слагаемые $I(A_i)$, $1 \leqslant i \leqslant n$, представляют собой независимые одинаково распределённые сл. величины с конечным математическим ожиданием

$$\mathsf{E}\, I(A_1) = \sum_{j=1}^m p_j I(A_1 = a_j) = -\sum_{j=1}^m p_j \log_2 p_j = H_\infty$$

и конечным вторым моментом

$$\mathsf{E}\, I^2(A_1) = \sum_{j=1}^{m} p_j (\log_2 p_j)^2.$$

Поэтому в силу закона больших чисел последовательность сл. величин $\frac{1}{n} I(A^n)$, $n \geqslant 1$, сходится по вероятности к числу H_∞.

2. Положим по определению

$$\mathscr{A}_{n,\varepsilon} = \left\{ a^n \in \mathscr{A}^n : \left| \frac{1}{n} I(a^n) - H_\infty \right| \leqslant \varepsilon \right\}, \quad \mathscr{A}'_{n,\varepsilon} = \mathscr{A}^n \setminus \mathscr{A}_{n,\varepsilon}.$$

Пусть $h_{n,j} = h_j(a^n)$ обозначает число вхождений буквы $a_j \in \mathscr{A}$ в последовательность a^n. Тогда формулу (2.3) можно переписать в виде

$$\mathsf{P}(A^n = a^n) = \prod_{j=1}^{m} (p_j)^{h_{n,j}}. \tag{2.16}$$

Сл. в. $\xi_{n,j} = h_j(A^n)$ имеет биномиальное распределение с параметрами n, p_j, и её моменты равны $\mathsf{E}\,\xi_{n,j} = np_j$, $\mathsf{D}\,\xi_{n,j} = np_j(1-p_j)$.

Воспользуемся тем, что формулу (2.16) можно переписать в виде

$$I(A^n) = -\log_2 \prod_{j=1}^{m} (p_j)^{\xi_{n,j}} = -\sum_{j=1}^{m} \xi_{n,j} \log_2 p_j,$$

и оценим сверху вероятность $\mathsf{P}_{1,2,\ldots,n}(\mathscr{A}'_{n,\varepsilon})$. Имеем

$$\begin{aligned}
\mathsf{P}_{1,2,\ldots,n}(\mathscr{A}'_{n,\varepsilon}) &= \mathsf{P}\left(\left| \frac{1}{n} I(A^n) - H_\infty \right| > \varepsilon \right) = \\
&= \mathsf{P}\left(\left| -\frac{1}{n} \sum_{j=1}^{m} \xi_{n,j} \log_2 p_j + \sum_{j=1}^{m} p_j \log_2 p_j \right| > \varepsilon \right) = \\
&= \mathsf{P}\left(\left| \sum_{j=1}^{m} \log_2 p_j \left(p_j - \frac{\xi_{n,j}}{n} \right) \right| > \varepsilon \right) \leqslant \\
&\leqslant \mathsf{P}\left(\sum_{j=1}^{m} |\log_2 p_j| \left| \frac{\xi_{n,j}}{n} - p_j \right| > \varepsilon \right) \leqslant \\
&\leqslant \mathsf{P}\left(\bigcup_{j=1}^{m} \left\{ |\log_2 p_j| \left| \frac{\xi_{n,j}}{n} - p_j \right| > \frac{\varepsilon}{m} \right\} \right) \leqslant \\
&\leqslant \sum_{j=1}^{m} \mathsf{P}\left(\left| \frac{\xi_{n,j}}{n} - p_j \right| > \frac{\varepsilon}{m|\log_2 p_j|} \right).
\end{aligned}$$

Применив неравенство Чебышёва

$$\mathsf{P}\left(\left|\frac{\xi_{n,j}}{n} - p_j\right| \geqslant u\right) \leqslant \frac{p_j(1-p_j)}{nu^2}$$

с $u = \varepsilon/(m\,|\log_2 p_j|)$, получим

$$\mathsf{P}_{1,2,\ldots,n}(\mathscr{A}'_{n,\varepsilon}) \leqslant \frac{m^2}{n\varepsilon^2} \sum_{j=1}^{m} p_j(1-p_j)(\log_2 p_j)^2.$$

Пусть N — наименьшее целое число с условием

$$N \geqslant \frac{m^2}{\delta\varepsilon^2} \sum_{j=1}^{m} p_j(1-p_j)(\log_2 p_j)^2.$$

Тогда для всех $n \geqslant N$ справедливо неравенство $\mathsf{P}_{1,2,\ldots,n}(\mathscr{A}'_{n,\varepsilon}) \leqslant \delta$. Таким образом, показано, что выполняются все условия определения 2.7.

Отметим, что если $p_j = 0$, то $\xi_{n,j} = 0$ с вероятностью 1, поэтому в формуле (2.16) и ниже мы рассматриваем только такие индексы j, $1 \leqslant j \leqslant m$, для которых $p_j > 0$. Кроме того, мы исключаем в доказательстве случай вырожденного распределения $\vec{p}$, для которого утверждение теоремы, очевидно, выполняется. □

Перейдём к рассмотрению условий, при которых свойство асимптотической равнораспределённости выполняется для простого марковского источника $(\mathscr{A}, \vec{p}, Q)$. Для этого вначале напомним некоторые сведения о конечных цепях Маркова, известные из курса теории вероятностей. Вероятностью перехода из состояния a_i на шаге n в состояние a_j на шаге $n+k$ называется величина $q_{ij}^{(n,n+k)} = \mathsf{P}(A_{n+k} = a_j \,/\, A_n = a_i)$. Если источник стационарный, то $q_{ij}^{(n,n+k)} = q_{ij}^{(1,k+1)}$, $1 \leqslant i, j \leqslant m$. Матрица $Q^{(k)} = (q_{ij}^{(1,k+1)})_{m\times m}$ совпадает с k-й степенью матрицы Q, $Q^{(k)} = Q^k$. Источник называется *регулярным*, если для некоторого натурального k все элементы матрицы Q^k положительны. Для регулярного простого марковского источника справедлива так называемая *эргодическая теорема*, которую мы приводим без доказательства.

Теорема 2.4. *Если простой марковский источник $(\mathscr{A}, \vec{p}, Q)$ регулярен, то существует такое распределение $\vec{q} = (q_1, \ldots, q_m)$, $q_j > 0$, что*

для любого $1 \leqslant j \leqslant m$

$$\lim_{k\to\infty} q_{ij}^{(1,k+1)} = q_j.$$

При этом $\vec{q}$ *— единственное стационарное распределение, т. е. такое, что* $\vec{q}\cdot Q = \vec{q}$. *Кроме того, существуют такие числа* $0 \leqslant \rho < 1$ *и* $c > 0$, *что* $|q_{ij}^{(1,k+1)} - q_j| < c\cdot\rho^k$ *для любого* $k \geqslant 1$.

Нам понадобится закон больших чисел для сл. в. $\xi_{n,ij}$, равной числу появлений биграммы (i, j) в сл. последовательности $A^n = (A_1, \ldots, A_n)$. Точнее, пусть

$$\xi_{n,ij} = \sum_{k=1}^{n-1} \chi(A_k = i, A_{k+1} = j),$$

где $\chi(S)$ — индикатор события S.

Теорема 2.5. *Если простой марковский источник* $(\mathscr{A}, \vec{p}, Q)$ *стационарен и регулярен, то для любых* $1 \leqslant i, j \leqslant m$ *последовательность сл. величин* $\xi_{n,ij}/(n-1)$, $n \geqslant 2$, *сходится по вероятности к числу* $p_i q_{ij}$.

Доказательство. В силу стационарности для любого $k \geqslant 1$

$$\mathsf{E}\,\chi(A_k = i, A_{k+1} = j) = \mathsf{E}\,\chi(A_1 = i, A_2 = j) = \mathsf{P}\,(A_1 = i, A_2 = j) = p_i q_{ij},$$

поэтому $\mathsf{E}\,\xi_{n,ij} = (n-1)p_i q_{ij}$.

Дисперсия сл. в. $\xi_{n,ij}$ представляется в виде

$$\mathsf{E}\,\{\xi_{n,ij} - (n-1)p_i q_{ij}\}^2 = \sum_{1\leqslant k,l\leqslant n-1} \mu_{kl}, \tag{2.17}$$

где

$$\begin{aligned}\mu_{kl} &= \mathsf{E}\,\{\chi(A_k = i, A_{k+1} = j) - p_i q_{ij}\}\{\chi(A_l = i, A_{l+1} = j) - p_i q_{ij}\} = \\ &= \mathsf{E}\,\chi(A_k = i, A_{k+1} = j)\chi(A_l = i, A_{l+1} = j) - (p_i q_{ij})^2 = \\ &= \mathsf{P}\,(A_k = i, A_{k+1} = j, A_l = i, A_{l+1} = j) - (p_i q_{ij})^2.\end{aligned}$$

Разобьём область суммирования в формуле (2.17) на три части, в зависимости от соотношения между множествами $\{k, k+1\}$ и $\{l, l+1\}$.

Если множества $\{k, k+1\}$ и $\{l, l+1\}$ совпадают, т. е. $k = l$, то $\mu_{kk} = p_i q_{ij} - (p_i q_{ij})^2$, и вклад таких слагаемых в сумму равен

$$\Sigma' = (n-1)(p_i q_{ij} - (p_i q_{ij})^2) = O(n).$$

Если множества $\{k, k+1\}$ и $\{l, l+1\}$ имеют ровно один общий элемент, то $\mu_{kl} = 0$ при $i \neq j$ и $\mu_{kl} = p_i(q_{ii})^2 - (p_i q_{ii})^2$ при $i = j$. Вклад таких слагаемых в сумму равен

$$\Sigma'' = 2(n-2)(p_i(q_{ii})^2 - (p_i q_{ii})^2)\chi(i=j) = O(n).$$

Если множества $\{k, k+1\}$ и $\{l, l+1\}$ не пересекаются, то

$$\mu_{kl} = p_i q_{ij} q_{ji}^{(1,s)} q_{ij} - (p_i q_{ij})^2 = p_i (q_{ij})^2 (q_{ji}^{(1,s)} - p_i),$$

где $s = |l-k|$. Поскольку по условию источник стационарный и эргодический, по теореме 2.4 начальное распределение $\vec{p}$ и предельное распределение $\vec{q}$ совпадают, и $|q_{ji}^{(1,s)} - p_i| < c \cdot \rho^s$. Таким образом, существует такое $c_1 > 0$, что $|\mu_{kl}| < c_1 \cdot \rho^s$. Вклад таких слагаемых в сумму равен

$$\Sigma''' = \sum_{k=1}^{n-1} \sum_{l \neq k,\, k+1} \mu_{kl}$$

и по модулю не превосходит величины

$$2c_1 \sum_{k=1}^{n-1} \sum_{s=2}^{n-2} \rho^s < 2c_1 n \sum_{s=0}^{+\infty} \rho^s = \frac{2c_1 n}{1-\rho} = O(n).$$

Таким образом,

$$\mathsf{E}\{\xi_{n,ij} - (n-1)p_i q_{ij}\}^2 = \Sigma' + \Sigma'' + \Sigma''' = O(n),$$

и по неравенству Чебышёва

$$\mathsf{P}\left(\left|\frac{\xi_{n,ij}}{n-1} - p_i q_{ij}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{E}\{\xi_{n,ij} - (n-1)p_i q_{ij}\}^2}{(n-1)^2 \varepsilon^2} = O\left(\frac{1}{n}\right),$$

что и требовалось. □

Теорема 2.6. *Если простой марковский источник* $(\mathscr{A}, \vec{p}, Q)$ *стационарен и регулярен, то он обладает свойством асимптотической равнораспределённости.*

Доказательство. Как и в случае дискретного источника без памяти, в силу замечания 2.5, достаточно доказать, что последовательность сл. величин $\frac{1}{n} I(A^n)$, $n \geqslant 1$, сходится по вероятности к числу H_∞.

Из формулы (2.4) и определения сл. в. $\xi_{n,ij}$ вытекает равенство

$$I(A^n) = -\log_2 p_{A_1} - \sum_{k=1}^{n-1} \log_2 q_{A_k A_{k+1}} =$$
$$= -\log_2 p_{A_1} - \sum_{i,j=1}^{m} \xi_{n,ij} \log_2 q_{ij}.$$

Принимая во внимание утверждение 2.6, получим

$$\frac{1}{n} I(A^n) - H_\infty = -\frac{\log_2 p_{A_1}}{n} - \sum_{i,j=1}^{m} \frac{\xi_{n,ij}}{n} \log_2 q_{ij} + \sum_{i,j=1}^{m} p_i q_{ij} \log_2 q_{ij} =$$
$$= U_n + V_n,$$

где

$$U_n = \sum_{i,j=1}^{m} \Big(p_i q_{ij} - \frac{\xi_{n,ij}}{n-1} \Big) \log_2 q_{ij},$$
$$V_n = -\frac{\log_2 p_{A_1}}{n} + \sum_{i,j=1}^{m} \frac{\xi_{n,ij}}{n(n-1)} \log_2 q_{ij}.$$

Отметим, что если $q_{ij} = 0$, то $\xi_{n,ij} = 0$ с вероятностью 1, поэтому в выписанных выше суммах мы рассматриваем только такие индексы i, j, $1 \leqslant i, j \leqslant m$, для которых $q_{ij} > 0$.

Из теоремы 2.5 следует, что последовательность сл. величин U_n сходится по вероятности к нулю. Из доказательства теоремы 2.3 видно, что сл. в. $I(A_1) = -\log_2 p_{A_1}$ имеет конечную дисперсию, а из доказательства теоремы 2.5 — что дисперсия сл. в. $\xi_{n,ij}/(n-1)$ стремится к нулю; поэтому последовательность V_n также сходится по вероятности к нулю. Таким образом,

$$\frac{1}{n} I(A^n) - H_\infty = U_n + V_n \xrightarrow{\mathsf{P}} 0,$$

что и требовалось. □

Замечание 2.8. Теоремы 2.3 и 2.6 могут быть существенно обобщены. Свойством асимптотической равнораспределённости обладают не только источники без памяти и регулярные простые марковские источники, но и все так называемые эргодические дискретные источники; более того, для таких источников последовательность

$I(A^n)/n$ сходится к H_∞ не только по вероятности, но и почти наверное (с вероятностью 1). Однако подробное изложение этих результатов (теорем Мак-Миллана—Хинчина и Бреймана) потребовало бы существенно увеличить объём учебного пособия.

В качестве одного из полезных применений свойства асимптотической равнораспределённости рассмотрим оценку количества *высоковероятных* последовательностей, порождаемых источником. Пусть задан дискретный источник сообщений, $\delta > 0$, и $s(n, \delta)$ обозначает мощность наименьшего подмножества $\mathscr{C}_n \subseteq \mathscr{A}^n$, для которого $\mathsf{P}_{1,2,\ldots,n}(\mathscr{C}_n) \geqslant 1 - \delta$. Очевидно, что подмножество $\mathscr{C}_n$ можно построить следующим образом: начать с пустого множества и добавлять к нему поочерёдно последовательности $a^n \in \mathscr{A}^n$, начиная с наиболее вероятной, затем наиболее вероятную из оставшихся, и так далее до тех пор, пока суммарная вероятность отобранных последовательностей не превысит величины $1 - \delta$. Что можно сказать о величине $s(n, \delta)$?

Теорема 2.7. *Если дискретный источник сообщений обладает свойством асимптотической равнораспределённости, то для произвольного δ, $0 < \delta < 1/2$, существует предел*

$$\lim_{n\to\infty} \frac{1}{n} \log_2 s(n, \delta) = H_\infty.$$

Доказательство. Пусть $\varepsilon > 0$; в соответствии с определением 2.7 выбирается такое подмножество $\mathscr{A}_{n,\varepsilon}$, что

$$\mathsf{P}_{1,2,\ldots,n}(\mathscr{A}_{n,\varepsilon}) \geqslant 1 - \delta$$

для всех достаточно больших n. По определению множества $\mathscr{C}_n$,

$$s(n, \delta) = |\mathscr{C}_n| \leqslant |\mathscr{A}_{n,\varepsilon}|,$$

и, в силу (2.14),

$$s(n, \delta) \leqslant 2^{n(H_\infty + \varepsilon)}.$$

С другой стороны, рассуждая так же, как в доказательстве утверждения 2.7, получим

$$
\begin{aligned}
s(n,\delta) = |\mathscr{C}_n| \geqslant |\mathscr{C}_n \cap \mathscr{A}_{n,\varepsilon}| &= \sum_{a^n \in \mathscr{C}_n \cap \mathscr{A}_{n,\varepsilon}} 1 \geqslant \\
&\geqslant 2^{n(H_\infty - \varepsilon)} \sum_{a^n \in \mathscr{C}_n \cap \mathscr{A}_{n,\varepsilon}} \mathsf{P}_{1,2,\ldots,n}(a^n) = \\
&= 2^{n(H_\infty - \varepsilon)} \mathsf{P}_{1,2,\ldots,n}(\mathscr{C}_n \cap \mathscr{A}_{n,\varepsilon}) \geqslant \\
&\geqslant (1 - 2\delta) 2^{n(H_\infty - \varepsilon)}.
\end{aligned}
$$

Здесь мы воспользовались тем, что если для некоторых событий A и B верны неравенства $\mathsf{P}(A) \geqslant 1 - \delta$ и $\mathsf{P}(B) \geqslant 1 - \delta$, то $\mathsf{P}(AB) \geqslant 1 - 2\delta$.

Таким образом,

$$
H_\infty - \varepsilon + \frac{1}{n}\log_2(1 - 2\delta) \leqslant \frac{1}{n}\log_2 s(n,\delta) \leqslant H_\infty + \varepsilon.
$$

Отсюда следует, что при $n \to \infty$ нижний и верхний пределы последовательности $\frac{1}{n}\log_2 s(n,\delta)$ заключены в границах

$$
H_\infty - \varepsilon \leqslant \underline{\lim_{n\to\infty}}\ \frac{1}{n}\log_2 s(n,\delta) \leqslant \overline{\lim_{n\to\infty}}\ \frac{1}{n}\log_2 s(n,\delta) \leqslant H_\infty + \varepsilon.
$$

Так как эти неравенства справедливы для любого $\varepsilon > 0$, предел последовательности $\frac{1}{n}\log_2 s(n,\delta)$ существует и равен H_∞. □

В качестве ещё одного применения свойства асимптотической равнораспределённости рассмотрим *сжимающее кодирование* сообщений дискретного источника. Построим такое взаимно однозначное отображение (кодирование) множества $\mathscr{A}^n$ сообщений длины n в некоторое множество двоичных *кодовых слов*, вообще говоря, различной длины, чтобы средняя длина кодового слова была как можно меньше. Теория такого кодирования будет подробно рассмотрена в следующей главе, а сейчас наша цель — построить пример кодирования на основе свойства асимптотической равнораспределённости.

Пример 2.4. Пусть $m = 2$, $\mathscr{A} = \{0, 1\}$, $\varepsilon > 0$, $\delta > 0$, и выбрано такое подмножество $\mathscr{A}_{n,\varepsilon}$, что $\mathsf{P}_{1,2,\ldots,n}(\mathscr{A}_{n,\varepsilon}) \geqslant 1 - \delta$.

Обозначим через d наименьшее целое число с условием $2^d \geqslant |\mathscr{A}_{n,\varepsilon}|$. Очевидно, что все последовательности из множества $\mathscr{A}_{n,\varepsilon}$ можно занумеровать d-разрядными двоичными числами. В качестве кодового слова $\varphi(a^n)$ для последовательности $a^n \in \mathscr{A}_{n,\varepsilon}$ возьмём d-

разрядный номер этой последовательности, к которому слева приписан символ 0. Для последовательности $a^n \in \mathscr{A}^n \setminus \mathscr{A}_{n,\varepsilon}$ в качестве кодового слова $\varphi(a^n)$ возьмём саму последовательность a^n, к которой слева приписан символ 1. Очевидно, что по первому символу кодового слова $\varphi(a^n)$ можно однозначно определить, каким способом оно получено, и восстановить a^n.

Длину кодового слова $\varphi(a^n)$ обозначим через $\mathrm{len}(\varphi(a^n))$. Оценим *коэффициент сжатия* μ_n, который по определению полагаем равным отношению средней длины кодового слова для случайной последовательности A^n к n — длине исходной последовательности,

$$\mu_n = \frac{1}{n}\mathsf{E}\ \mathrm{len}(\varphi(A^n)).$$

Нетрудно подсчитать, что

$$\mu_n = \frac{1}{n}\{(d+1)\mathsf{P}_{1,2,\ldots,n}(\mathscr{A}_{n,\varepsilon}) + (n+1)(1-\mathsf{P}_{1,2,\ldots,n}(\mathscr{A}_{n,\varepsilon}))\}.$$

Из (2.14) и (2.15) следует, что

$$n(H_\infty - \varepsilon) + \log_2(1-\delta) \leqslant d \leqslant n(H_\infty + \varepsilon) + 1,$$

поэтому

$$(1-\delta)\left\{H_\infty - \varepsilon + \frac{\log_2(1-\delta)}{n}\right\} + \frac{1}{n} \leqslant \mu_n \leqslant H_\infty + \varepsilon + \delta + \frac{2}{n}.$$

Отсюда видно, что при достаточно малых ε, δ и достаточно больших n коэффициент сжатия μ_n можно сделать сколь угодно близким к предельной энтропии H_∞.

§ 2.4. Задачи

2.1. Для источника без памяти с параметрами $\mathscr{A} = \{0, 1\}$, $\vec{p} = (q, p)$ рассмотрим следующую схему кодирования. Для $0 < k < n$ обозначим через $\mathscr{A}_k^n$ множество всех двоичных слов длины n, в которых число единиц не превосходит k. Пусть d — наименьшее целое с условием $2^d \geqslant |\mathscr{A}_k^n|$. Слова из $\mathscr{A}_k^n$ занумеруем d-разрядными двоичными числами. В качестве кодового слова $\varphi(a^n)$ для $a^n \in \mathscr{A}_k^n$ возьмём d-разрядный номер слова a^n, к которому слева приписан символ 0, а для $a^n \in \mathscr{A}^n \setminus \mathscr{A}_k^n$ в качестве $\varphi(a^n)$ возьмём само слово a^n, к которому слева приписан символ 1.
Найти выражение для коэффициента сжатия $\mu_n = \frac{1}{n}\mathsf{E}\ \mathrm{len}(\varphi(A^n))$. При каком k значение μ_n минимально? Вычислить μ_n для $n = 100$, $k = 3$, $p = 0{,}005$.

2.2. Для источника без памяти с параметрами $\mathscr{A} = \{0, 1\}$, $\vec{p} = \left(\frac{1}{4}, \frac{3}{4}\right)$ указать такое N, что при $n \geqslant N$ выполняются условия асимптотической равнораспределённости для $\varepsilon = 0{,}1$, $\delta = 0{,}1$; $\varepsilon = 0{,}01$, $\delta = 0{,}05$.

2.3. Для источника без памяти с параметрами $\mathscr{A} = \{0, 1\}$, $\vec{p} = (q, p)$ предположим, что символы 0 и 1 имеют «длительности передачи» $t(0) = 1$ и $t(1) = 2$. Найти p, для которого величина $\dfrac{H(A^n)}{\mathsf{E}\, t(A^n)}$ максимальна.

2.4. Для простого марковского источника $(\mathscr{A}, \vec{p}, Q)$ с заданной матрицей Q найти вектор $\vec{p}$, для которого источник будет стационарным, и вычислить H_k, $H^{(k)}$ и H_∞, если матрица Q равна

а) $\begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix}$; б) $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix}$; в) $\begin{pmatrix} 0 & \frac{1}{5} & \frac{4}{5} \\ \frac{1}{2} & \frac{1}{10} & \frac{2}{5} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$;

г) $\begin{pmatrix} 1-p & \frac{p}{2} & \frac{p}{2} \\ \frac{p}{2} & 1-p & \frac{p}{2} \\ \frac{p}{2} & \frac{p}{2} & 1-p \end{pmatrix}$; д) $\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$;

е) $\begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{2} \end{pmatrix}$.

2.5. Для марковского источника с алфавитом $\mathscr{A} = \{0, 1\}$ и глубиной зависимости $s = 2$, с заданной матрицей $Q^{(2)}$ переходных вероятностей найти распределение $\mathsf{P}_{1,2}$, для которого источник будет стационарным, и вычислить H_k, $H^{(k)}$ и H_∞, если матрица $Q^{(2)}$ равна

а) $\begin{pmatrix} & 0 & 1 \\ 00 & 0{,}8 & 0{,}2 \\ 01 & 0{,}5 & 0{,}5 \\ 10 & 0{,}5 & 0{,}5 \\ 11 & 0{,}2 & 0{,}8 \end{pmatrix}$; б) $\begin{pmatrix} & 0 & 1 \\ 00 & 0{,}25 & 0{,}75 \\ 01 & 0{,}25 & 0{,}75 \\ 10 & 0{,}75 & 0{,}25 \\ 11 & 0{,}75 & 0{,}25 \end{pmatrix}$.

2.6. На шахматной доске размером 3×3 случайно ставится король, и следующая позиция короля выбирается случайно и равновероятно по всем возможным ходам из текущей позиции. Для такого случайного процесса вычислить H_∞. Тот же вопрос для ладьи, ферзя, слона (чёрного и белого).

2.7. Колода, состоящая из 26 красных карт и 26 чёрных, хорошо перетасована. Поочерёдно извлекают карту за картой без возвращения. Пусть A_k — цвет k-й извлечённой карты.

а) Вычислить $H(A_1)$, $H(A_2)$, $H(A_1, \ldots, A_{52})$.

б) Выяснить, как изменяется величина $H(A_k/A_1, \ldots, A_{k-1})$ с ростом k — возрастает или убывает.

2.8. Для простого марковского источника доказать неравенства:

а) $I(A_1; A_3) + I(A_2; A_4) \leqslant I(A_1; A_4) + I(A_2; A_3)$;

б) $H(A_1/A_{k+1}) \geqslant H(A_1/A_k)$ («чем дальше в будущее, тем труднее определить начальное состояние»).

2.9. Для простого марковского источника $(\mathscr{A}, \vec{p}, Q)$ найти выражение для величины $I(A_1; A_2, \ldots, A_k)$.

2.10. Для стационарного источника сообщений $A = (A_1, A_2, \ldots)$ доказать, что $\lim_{n\to\infty} \frac{1}{n} I(A_1, \ldots, A_n; A_{n+1}, \ldots, A_{2n}) = 0$.

2.11. Пусть $A = (A_1, A_2, \dots)$ — стационарный источник сообщений, f и g — функции от одного и двух переменных соответственно, $Y_n = f(A_n)$, $Z_n = f(A_n, A_{n+1})$. Обозначим через $H_\infty(A)$, $H_\infty(Y)$ и $H_\infty(Z)$ предельные энтропии источников A, $Y = (Y_1, Y_2, \dots)$ и $Z = (Z_1, Z_2, \dots)$ соответственно.

а) Доказать неравенство $H_\infty(Y) \leqslant H_\infty(A)$.

б) Каково соотношение между $H_\infty(Z)$ и $H_\infty(A)$?

2.12. Для источника без памяти с алфавитом $\mathscr{A} = \{1, 2, 3\}$ и распределением вероятностей $\vec{p} = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$ описать предельное распределение для сл. в. $Y_n = \left(\prod_{i=1}^{n} A_i\right)^{1/n}$.

Глава 3
Кодирование дискретных источников

§ 3.1. Коды и кодовые деревья

В этой главе будем рассматривать методы преобразования (кодирования) последовательностей, порождаемых дискретным источником, с целью их оптимального представления. При этом оптимальность будем понимать в смысле наименьшей средней длины последовательностей, получаемых в результате такого кодирования.

Пусть даны два конечных множества, $\mathscr{A} = \{a_1, \ldots, a_m\}$ — алфавит источника сообщений, и $\mathscr{B} = \{b_1, \ldots, b_D\}$ — *кодовый* алфавит. Обозначим через $\mathscr{B}^* = \bigcup_{n \geqslant 1} \mathscr{B}^n$ множество всех конечных последовательностей в алфавите $\mathscr{B}$.

Определение 3.1. Алфавитным D-ичным кодированием будем называть произвольное отображение $\varphi\colon \mathscr{A} \to \mathscr{B}^*$. При этом образ $\varphi(a_i) \in \mathscr{B}^*$ буквы $a_i \in \mathscr{A}$ назовём кодовым словом, или результатом кодирования буквы a_i, а длину этого кодового слова обозначим $\ell_i = \mathrm{len}(\varphi(a_i))$.

Набор кодовых слов $\varphi(\mathscr{A}) = (\varphi(a_1), \ldots, \varphi(a_m))$ будем называть D-ичным кодом для алфавита $\mathscr{A}$. Если при этом все кодовые слова имеют одинаковую длину, кодирование φ назовём равномерным, в противном случае — неравномерным.

Определение 3.2. Продолжением φ^* алфавитного кодирования φ на множество $\mathscr{A}^* = \bigcup_{n \geqslant 1} \mathscr{A}^n$ всех конечных последовательностей в алфавите $\mathscr{A}$ назовём отображение $\varphi^*\colon \mathscr{A}^* \to \mathscr{B}^*$, которое получается по правилу сцепления (приписывания) кодовых слов: для любого $a^n = (a_{i_1}, \ldots, a_{i_n})$ положим $\varphi^*(a_{i_1}, \ldots, a_{i_n}) = \varphi(a_{i_1}) \ldots \varphi(a_{i_n})$.

Определение 3.3. Алфавитное кодирование φ (и набор кодовых слов $\varphi(\mathscr{A})$) называем

1) префиксным (суффиксным), если никакое кодовое слово $\varphi(a_i)$ не является началом (окончанием) какого-либо другого кодового слова $\varphi(a_j)$, $i \neq j$;
2) однозначно декодируемым, если отображение φ^* инъективно.

Замечание 3.1. Если кодирование φ однозначно декодируемое, то для любой последовательности $a^n = (a_{i_1}, \ldots, a_{i_n})$ соответствующая *кодовая последовательность* $\varphi^*(a_{i_1}, \ldots, a_{i_n})$ единственным образом разбивается на кодовые слова $\varphi(a_{i_1}), \ldots, \varphi(a_{i_n})$ из кода $\varphi(\mathscr{A})$.

Пример 3.1. Пусть $m = D = 2$, $\mathscr{A} = \mathscr{B} = \{0, 1\}$. В следующей таблице приведены пять вариантов алфавитного кодирования и указано, являются они или нет префиксными, суффиксными и однозначно декодируемыми.

	φ_1	φ_2	φ_3	φ_4	φ_5
0	0	0	0	00	0
1	01	10	010	11	00
префиксное	–	+	–	+	–
суффиксное	+	–	–	+	–
однозначно декодируемое	+	+	+	+	–

Утверждение 3.1. *Если алфавитное кодирование φ является префиксным или суффиксным, то оно — однозначно декодируемое. Обратное неверно.*

Доказательство. Предположим, что алфавитное кодирование φ префиксное. Укажем алгоритм, называемый «чтением слева направо», который для произвольной последовательности $a^n = (a_{i_1}, \ldots, a_{i_n})$ выполняет разбиение кодовой последовательности $\varphi^*(a_{i_1}, \ldots, a_{i_n})$ на кодовые слова $\varphi(a_{i_1}), \ldots, \varphi(a_{i_n})$ из кода $\varphi(\mathscr{A})$. Алгоритм состоит в чтении последовательности $\varphi^*(a_{i_1}, \ldots, a_{i_1})$ по одному символу слева направо, пока прочитанные символы впервые не составят некоторое кодовое слово $\varphi(a_{i_1})$. Это кодовое слово декодируется в символ a_{i_1} и отбрасывается. Если остались ещё непрочитанные символы кодовой последовательности, процесс чтения слева направо продолжается. Очевидно, что алгоритм выполняет единственно возможное разбиение кодовой последовательности на кодовые слова — достаточно заметить, что, в силу свойства префиксности, первое кодовое слово определяется однозначно, и применить индукцию по длине кодовой последовательности.

Для суффиксного кодирования доказательство аналогично, с очевидной заменой алгоритма на «чтение справа налево». Последняя часть утверждения следует из приведённого выше примера φ_3. □

Напомним некоторые понятия и утверждения теории графов, которые предполагаются известными читателю. Пусть $\mathcal{V}$ — конечное множество и $\mathcal{E} \subseteq \mathcal{V}^2$. Пару $G = (\mathcal{V}, \mathcal{E})$ называют (неориентированным) конечным *графом* с множеством вершин $\mathcal{V}$ и множеством рёбер $\mathcal{E}$. Если $u, v \in \mathcal{V}$ и $(u, v) \in \mathcal{E}$, то говорят, что вершины u и v связаны *ребром* (u, v) в графе G. По определению, вершины u и v связаны в графе G *путём* $(u, w_1), (w_1, w_2), \ldots, (w_{k-1}, w_k), (w_k, v)$, если множество $\mathcal{E}$ содержит указанные рёбра; если при этом $u = v$, то такой путь называется *циклом*. Граф G называется *связным*, если в нём любые две различные вершины связаны некоторым путём.

Граф G называется *деревом*, если он связен и не содержит циклов. Любые две вершины в дереве связаны единственным путём. Дерево G называется *корневым*, если в нём выделена некоторая вершина $v_0 \in \mathcal{V}$, называемая *корнем*. Очевидно, одному дереву соответствует столько различных корневых деревьев, сколько в нём вершин. Для корневого дерева индуктивно определяется понятие *уровня*. Нулевой уровень $\mathcal{V}_0$ содержит единственную вершину — корень v_0. Если уже определены уровни $\mathcal{V}_l$, $1 \leqslant l \leqslant k-1$, то к уровню $\mathcal{V}_k$ отнесём все те вершины, которые не входят в перечисленные уровни $\mathcal{V}_l$ с номерами $l < k$ и связаны ребром с вершинами уровня $\mathcal{V}_{k-1}$. *Высотой* вершины называется номер уровня, в котором лежит эта вершина. Высотой корневого дерева называется максимальный номер непустого уровня этого дерева.

Если вершины u и v связаны ребром и при этом $u \in \mathcal{V}_{k-1}$, $v \in \mathcal{V}_k$, то вершина u называется *предком* вершины v, а вершина v — *потомком* вершины u; при этом ребро (u, v) называется исходящим из вершины u. *Степенью* вершины называется число её потомков, или, что то же самое, число исходящих из неё рёбер. Если вершина v не имеет потомков, то она называется *концевой* вершиной, или *листом*. Если две вершины u и v имеют общего предка, то u и v называются *братьями*.

Если G — корневое дерево, $v \in \mathcal{V}_k$, то *поддеревом* G_v с корнем v называется такой граф, у которого множество вершин состоит из вершины v и всех вершин дерева G, связанных в нём некоторым путём

с вершиной v и принадлежащих уровням $\mathcal{V}_l$ с номерами $l > k$; две вершины дерева G_v связаны ребром тогда и только тогда, когда они связаны ребром в исходном дереве G. Если вершины u и v связаны некоторым путём в корневом дереве G, то либо u принадлежит G_v, либо v принадлежит G_u.

Корневое дерево G называется *D-ичным*, если любая его вершина имеет не более D потомков. Для такого дерева индуктивно определим D-ичную *разметку*, то есть сопоставим каждой вершине u её метку $\mu(u)$, представляющую собой некоторое слово в алфавите $\mathscr{B} = \{b_1, \ldots, b_D\}$. Вначале для каждой неконцевой вершины u произвольным образом пометим все исходящие из u рёбра различными символами алфавита $\mathscr{B}$; это можно сделать, поскольку число потомков не превосходит D. По определению, метка $\mu(v_0)$ корня v_0 есть пустое слово. Если вершина u является предком вершины v, метка $\mu(u)$ уже определена, а ребро (u, v) помечено символом $b_j \in \mathscr{B}$, то метка $\mu(v)$ вершины v получается из метки $\mu(u)$ приписыванием справа символа b_j, $\mu(v) = \mu(u)b_j$. Очевидно, для вершин k-го уровня метка имеет длину k. Полученное в результате дерево с метками будем назвать *размеченным*.

Пример 3.2. На рисунке 3.1 приведены размеченные корневые деревья: а) двоичное, высоты два, с пятью вершинами, в том числе тремя листьями; б) троичное, высоты три, с девятью вершинами, в том числе шестью листьями.

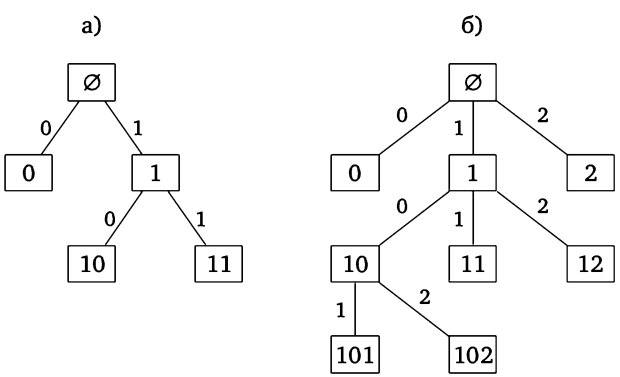

Рис. 3.1. Размеченные корневые деревья

Теорема 3.1.

1. *Если* G — *D-ичное размеченное корневое дерево, то множество меток всех его листьев образует префиксный D-ичный код.*
2. *Если* $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ — *D-ичное префиксное кодирование, то существует такое D-ичное размеченное корневое дерево* G, *для которого множество меток всех листьев совпадает с кодом* $\varphi(\mathscr{A})$.

Доказательство. 1. Пусть G — D-ичное размеченное корневое дерево, имеющее хотя бы два листа u и v. Докажем от противного, что метка $\mu(v)$ не является префиксом метки $\mu(u)$. Действительно, если метка $\mu(v)$ является префиксом метки $\mu(u)$, то из определения размеченного дерева следует, что вершина u принадлежит поддереву G_v, поэтому вершина v — не лист, получили противоречие.

2. Пусть $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ — D-ичное префиксное кодирование с длинами кодовых слов $\operatorname{len}(\varphi(a_i)) = \ell_i$, $1 \leqslant i \leqslant m$, и $L = \max(\ell_1, \ldots, \ell_m)$. Рассмотрим *полное* D-ичное размеченное корневое дерево G' высоты L, то есть такое дерево, в котором каждая вершина на уровнях $\mathscr{V}_k$, $1 \leqslant k \leqslant L-1$, имеет ровно D потомков, а все D^L вершин уровня $\mathscr{V}_L$ являются листьями. В качестве меток вершин дерева G' встречаются все возможные слова в алфавите $\mathscr{B}$, имеющие длину не более L, в том числе все кодовые слова кода $\varphi(\mathscr{A})$.

Применим к дереву G' алгоритм *удаления листьев*, состоящий в следующем. На первом шаге алгоритма последовательно просматриваем листья дерева G' на уровне $\mathscr{V}_L$. Если метка $\mu(v)$ рассматриваемого листа v принадлежит коду $\varphi(\mathscr{A})$, то этот лист оставляем без изменений; если же не принадлежит, то удаляем из дерева G' лист v и ребро, связывающее v с его предком на предшествующем уровне дерева. При этом хотя бы один лист на уровне $\mathscr{V}_L$ останется, так как в коде $\varphi(\mathscr{A})$ имеется хотя бы одно кодовое слово длины L. Если для некоторой вершины u на уровне $\mathscr{V}_{L-1}$ все её потомки были удалены, то вершина u становится листом.

На втором шаге алгоритма последовательно просматриваем новые листья, образовавшиеся на уровне $\mathscr{V}_{L-1}$, и проделываем с ними описанную выше операцию.

Перемещаясь таким образом уровень за уровнем ближе к корню, завершим алгоритм тогда, когда на очередном шаге невозможно удалить ни один лист. В итоге получим искомое дерево G. □

Замечание 3.2. Теорема 3.1 устанавливает взаимно однозначное соответствие между множествами слов в алфавите $\mathscr{B}$, в которых никакое слово не является префиксом другого слова, с одной стороны, и размеченными D-ичными деревьями, с другой стороны. Однако множество слов с указанным свойством — ещё не префиксный код, поскольку множество неупорядочено. Чтобы из данного множества слов получить алфавитное префиксное кодирование, или префиксный код, необходимо ещё задать взаимно однозначное соответствие между символами алфавита $\mathscr{A}$ и кодовыми словами, т. е. пронумеровать кодовые слова. Такое соответствие можно задать и в терминах размеченного D-ичного дерева — как соответствие между символами алфавита $\mathscr{A}$ и листьями дерева. Именно в таком смысле понимается термин «кодовые деревья» в заголовке этого параграфа и далее по тексту главы. Очевидно, если дано размеченное дерево с m листьями, то ему отвечает $m!$ различных кодовых деревьев, или, что эквивалентно, $m!$ различных префиксных кодов.

Далее в этой главе мы будем рассматривать префиксное кодирование φ и соответствующее кодовое дерево G как эквивалентные описания одного и того же объекта и переходить с языка префиксного кодирования на язык кодовых деревьев и обратно, когда это позволяет сделать рассуждения более простыми и наглядными.

Теорема 3.2 (неравенство Крафта).

1. *Если* $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ *— D-ичное префиксное алфавитное кодирование с длинами кодовых слов* $\operatorname{len}(\varphi(a_i)) = \ell_i$, $1 \leqslant i \leqslant m$, *то справедливо неравенство*
$$\sum_{i=1}^{m} D^{-\ell_i} \leqslant 1. \tag{3.1}$$
2. *Если натуральные числа D, m, $\ell_1, \ldots, \ell_m$ удовлетворяют неравенству* (3.1), *то существует D-ичное префиксное алфавитное кодирование* $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ *с длинами кодовых слов* $\operatorname{len}(\varphi(a_i)) = \ell_i$, $1 \leqslant i \leqslant m$.

Доказательство. Как и выше, обозначим $L = \max(\ell_1, \ldots, \ell_m)$.

1. Рассмотрим полное D-ичное размеченное корневое дерево G высоты L. Число листьев дерева G равно D^L. Выделим в дереве G те вершины v_i, $1 \leqslant i \leqslant m$, для которых метки $\mu(v_i)$ являются кодовыми словами кода $\varphi(\mathscr{A})$, и рассмотрим поддеревья G_{v_i} дерева G. Так как

вершина v_i в дереве G принадлежит уровню с номером ℓ_i, высота поддерева G_{v_i} равна $L - \ell_i$, а число листьев этого поддерева равно $D^{L-\ell_i}$. Каждый лист поддерева G_{v_i} является и листом дерева G, и при $i \neq j$ множества листьев поддеревьев G_{v_i} и G_{v_j} не пересекаются. Следовательно,

$$D^{L-\ell_1} + \ldots + D^{L-\ell_m} \leqslant D^L.$$

Сокращая на D^L, приходим к неравенству (3.1).

2. Для $\ell \in \{1, \ldots, L\}$ обозначим через w_ℓ количество чисел набора $\ell_1, \ldots, \ell_m$, равных ℓ. Неравенство (3.1) можно переписать в виде

$$\sum_{\ell=1}^{L} w_\ell D^{-\ell} \leqslant 1. \tag{3.2}$$

По определению числа L, $w_L > 0$, следовательно, последнее слагаемое в сумме (3.2) положительно. Поэтому частичные суммы первых $k < L$ слагаемых строго меньше единицы:

$$\sum_{\ell=1}^{k} w_\ell D^{-\ell} < 1, \quad k = 1, \ldots, L-1,$$

или в другой форме

$$D^k - \sum_{\ell=1}^{k} w_\ell D^{k-\ell} > 0, \quad k = 1, \ldots, L-1. \tag{3.3}$$

Чтобы доказать существование D-ичного префиксного кодирования $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ с длинами кодовых слов $\operatorname{len}(\varphi(a_i)) = \ell_i$, $1 \leqslant i \leqslant m$, достаточно, в силу теоремы 3.1, построить D-ичное размеченное корневое дерево G с m листьями, метки которых имеют длины $\ell_1, \ldots, \ell_m$, или, что то же самое, дерево G, в котором на первом уровне w_1 листьев, на втором уровне w_2 листьев, …, на L-м уровне w_L листьев. Будем строить дерево G, начиная с корня, последовательно уровень за уровнем, от первого уровня до уровня с номером L. Для наглядности опишем построение первых двух уровней.

Возьмём в качестве корня вершину v_0 и добавим к ней D потомков $v_1, \ldots, v_D$. Объявим $v_1, \ldots, v_{w_1}$ листьями, а $v_{w_1+1}, \ldots, v_D$ — промежуточными вершинами, из которых будут исходить рёбра к вершинам второго уровня. Неравенство (3.3) при $k = 1$ имеет вид $D - w_1 > 0$; это означает, что хотя бы одна промежуточная вершина

на первом уровне есть, и мы можем продолжить процесс и перейти к построению второго уровня дерева.

К каждой из промежуточных вершин $v_{w_1+1}, \ldots, v_D$ добавим по D потомков, и общее число вершин на втором уровне будет равно $D(D-w_1)$. Первые w_2 из этих вершин второго уровня объявим листьями, а остальные $D(D-w_1)-w_2$ вершин — промежуточными вершинами, из которых будут исходить рёбра к вершинам третьего уровня. Неравенство (3.3) при $k=2$ имеет вид $D(D-w_1)-w_2>0$; это означает, что хотя бы одна промежуточная вершина на втором уровне есть, и мы можем продолжить процесс и перейти к построению третьего уровня.

Продолжая таким же образом, предположим, что уже построены первые $k<L$ уровней дерева G, причём на каждом уровне i, $1\leqslant i\leqslant k$, имеется ровно w_i листьев и $D^i-\sum_{\ell=1}^{i} w_\ell D^{i-\ell}$ промежуточных вершин. Тогда неравенство (3.3) гарантирует, что хотя бы одна промежуточная вершина на k-м уровне есть, и мы можем к каждой такой промежуточной вершине добавить по D потомков и построить $(k+1)$-й уровень. На шаге $k=L-1$ построение искомого дерева завершается. □

Теорема 3.3 (неравенство Мак-Миллана). *Если алфавитное кодирование $\varphi\colon \mathscr{A}\to\mathscr{B}^*$ с длинами кодовых слов $\operatorname{len}(\varphi(a_i))=\ell_i$, $1\leqslant i\leqslant m$, является однозначно декодируемым, то справедливо неравенство* (3.1).

Доказательство. Для натуральных n и ℓ обозначим через $C_{n,\ell}$ множество таких последовательностей $a^n=(a_{i_1},\ldots,a_{i_n})\in\mathscr{A}^n$, для которых кодовая последовательность $\varphi^*(a_{i_1},\ldots,a_{i_n})$ имеет длину $\ell_{i_1}+\ldots+\ell_{i_n}$, равную ℓ, и пусть $c_{n,\ell}=|C_{n,\ell}|$. По условию, отображение $\varphi^*\colon\mathscr{A}^*\to\mathscr{B}^*$ инъективно, и при этом $\varphi^*(C_{n,\ell})\subseteq\mathscr{B}^\ell$. Следовательно, $c_{n,\ell}\leqslant|\mathscr{B}^\ell|=D^\ell$.

Рассмотрим выражение

$$\begin{aligned}\Big(\sum_{i=1}^{m} D^{-\ell_i}\Big)^n &= \sum_{1\leqslant i_1,\ldots,i_n\leqslant m} D^{-(\ell_{i_1}+\ldots+\ell_{i_n})} = \\ &= \sum_{(a_{i_1},\ldots,a_{i_n})\in\mathscr{A}^n} D^{-\operatorname{len}(\varphi^*(a_{i_1},\ldots,a_{i_n}))} = \\ &= \sum_{\ell=1}^{nL} c_{n,\ell} D^{-\ell},\end{aligned}$$

где $L = \max(\ell_1, \ldots, \ell_m)$. Так как $c_{n,\ell} D^{-\ell} \leqslant 1$, получаем неравенство

$$\Big(\sum_{i=1}^{m} D^{-\ell_i}\Big)^n \leqslant \sum_{\ell=1}^{nL} 1 = nL,$$

справедливое для всех $n \geqslant 1$. Отсюда следует, что $\sum_{i=1}^{m} D^{-\ell_i} \leqslant 1$, поскольку в противном случае $\lim_{n\to+\infty} \big(\sum_{i=1}^{m} D^{-\ell_i}\big)^n/(nL) = +\infty$. □

Из теоремы 3.3 и второй части теоремы 3.2 вытекает

Следствие 3.1. *Если алфавитное кодирование $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ является однозначно декодируемым, то существует префиксное алфавитное кодирование $\varphi_1\colon \mathscr{A} \to \mathscr{B}^*$ с таким же набором длин кодовых слов, т. е.* $\operatorname{len}(\varphi_1(a_i)) = \operatorname{len}(\varphi(a_i))$, $1 \leqslant i \leqslant m$.

В дальнейшем при рассмотрении оптимальных кодов нам понадобится один важный класс деревьев, называемых насыщенными.

Определение 3.4. Говорят, что D-ичное корневое дерево G является насыщенным, если любая неконцевая вершина имеет ровно D потомков, за исключением, может быть, одной вершины на предпоследнем уровне, называемой *особой* вершиной, которая имеет D_0 потомков, $2 \leqslant D_0 < D$.

Сразу отметим, что при $D = 2$ в насыщенном дереве G нет особых вершин.

Пример 3.3. На рисунке 3.2 приведены примеры деревьев: (а) насыщенное, $D = 2$; (б) ненасыщенное, $D = 2$; (в) насыщенное, $D = 3$; (г) насыщенное, $D = 4$.

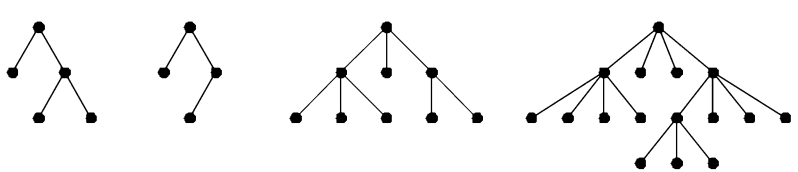

Рис. 3.2. Насыщенные и ненасыщенные корневые деревья

Лемма 3.1. *Для любых целых чисел $D \geqslant 2$, $m \geqslant 1$ существует насыщенное D-ичное корневое дерево с m листьями.*

Доказательство. При фиксированном $D \geqslant 2$ докажем лемму индукцией по $m \geqslant 1$. При $m \leqslant D$ для построения искомого дерева достаточно взять корень v_0 и добавить к нему m потомков $v_1, \ldots, v_m$.

Пусть G — насыщенное D-ичное корневое дерево с m листьями. Покажем, как из него получить насыщенное D-ичное корневое дерево G′ с $m+1$ листьями. Если в дереве G есть особая вершина, то достаточно добавить к ней ещё одного потомка. Если же особой вершины в дереве G нет, то можно взять любой лист на последнем уровне дерева и добавить к нему два потомка. В любом из этих двух случаев число листьев увеличится на 1. □

По заданному m можно судить о том, имеется ли в насыщенном D-ичном корневом дереве с m листьями особая вершина (и если имеется, то сколько у неё потомков), а также о том, достигается ли равенство в неравенстве Крафта.

Лемма 3.2. *Для $D \geqslant 2$, $m \geqslant 1$ обозначим через m_0 такое наименьшее целое число вида $m_0 = 1 + k(D-1)$, что $m_0 \geqslant m$. Пусть* G *— насыщенное D-ичное корневое дерево с m листьями, высоты которых равны $\ell_1, \ldots, \ell_m$, и $K = \sum_{i=1}^{m} D^{-\ell_i}$.*

Если $m_0 = m$, то в дереве G *особой вершины нет и $K = 1$, а если $m_0 > m$, то особая вершина в дереве* G *есть, число её потомков равно $D_0 = m - m_0 + D$, и $K < 1$.*

Доказательство. Докажем лемму индукцией по $m \geqslant 1$.

Если $m \leqslant D$, то $m_0 = D$, $k = 1$, и насыщенное дерево G представляет собой корень с m листьями. Если $m = D$, то корень — не особая вершина, а если $m < D$, то особая. При этом $K = mD^{-1}$, поэтому $K = 1$ тогда и только тогда, когда $m = D$.

Пусть $m > D$, $k > 1$, тогда насыщенное дерево G имеет не менее двух уровней, а корень дерева имеет ровно D потомков. Обозначим через $L = \max(\ell_1, \ldots, \ell_m)$ высоту дерева G.

Если в дереве G есть особая вершина u на уровне $L-1$, удалим её D_0 потомков, $2 \leqslant D_0 < D$, а саму особую вершину u сделаем листом. Получим новое насыщенное дерево G′ с $m - D_0 + 1 < m$ листьями, не имеющее особой вершины; соответствующая K ве-

личина K' для дерева G' равна $K' = K - D^{-L}D_0 + D^{-(L-1)}$. По предположению индукции, число листьев в дереве G' без особой вершины равно $1 + (k-1)(D-1)$, и $K' = 1$. Приравнивая $m - D_0 + 1$ и $1 + (k-1)(D-1)$, получаем $D_0 = m - m_0 + D$, так что $m = m_0 + D_0 - D < m_0$. Из условия $K' = 1$ получаем

$$K = 1 + D^{-L}D_0 - D^{-(L-1)} = 1 - D^{-(L-1)}(D - D_0) < 1.$$

Если особой вершины в дереве G нет, возьмём произвольную неконцевую вершину u на уровне $L-1$; эта вершина имеет ровно D потомков. Удалим из дерева D потомков вершины u, а саму вершину u сделаем листом. Получим новое насыщенное дерево G' с $m - D + 1 < m$ листьями, не имеющее особой вершины; соответствующая K величина K' для дерева G' равна

$$K' = K - D^{-L}D + D^{-(L-1)} = K.$$

По предположению индукции, число листьев в дереве G' без особой вершины равно $1 + (k-1)(D-1)$, и $K' = 1$, так что $K = 1$. Приравнивая $m - D + 1$ и $1 + (k-1)(D-1)$, получаем $m = m_0$. □

Пример 3.4. Если $m = 8$, $D = 4$, то из условия $m_0 = 1 + 3k \geqslant 8$ получаем $k = 3$, $m_0 = 10 > 8$, поэтому особая вершина есть, и число её потомков равно $D_0 = m - m_0 + D = 2$. Если $m = 7$, $D = 3$, то из условия $m_0 = 1 + 2k \geqslant 7$ получаем $k = 3$, $m_0 = 7$, поэтому особой вершины нет. Возможные насыщенные деревья с данными параметрами приведены на рисунке 3.3.

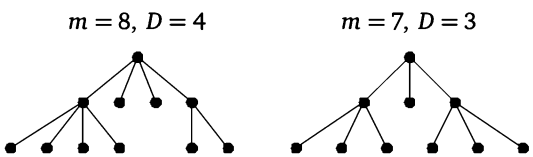

Рис. 3.3. Насыщенные корневые деревья

§ 3.2. Средняя длина оптимального кода

Предположим, что дано алфавитное кодирование $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ с длинами кодовых слов $\mathrm{len}(\varphi(a_i)) = \ell_i$, $1 \leqslant i \leqslant m$, и правило кодирования φ^* применяется к последовательностям, порождаемым

дискретным источником без памяти $(\mathscr{A}, \vec{p})$, $\vec{p} = (p_1, \ldots, p_m)$. Далее в этом параграфе считаем, что распределение $\vec{p} = (p_1, \ldots, p_m)$ задано и фиксировано.

Нас будет интересовать такое кодирование φ, для которого выполняется условие однозначного декодирования, и при этом средняя длина случайной кодовой последовательности $\varphi^*(A^n)$ минимальна при каждом $n \geqslant 1$. Поскольку для источника без памяти сл. величины $A_1, A_2, \ldots$ одинаково распределены,

$$\mathsf{E}\ \mathrm{len}(\varphi^*(A^n)) = n\mathsf{E}\ \mathrm{len}(\varphi(A_1)) = n \sum_{i=1}^{m} p_i \ell_i \,. \tag{3.4}$$

Эти соображения приводят нас к следующим определениям.

Определение 3.5. Пусть заданы алфавит источника $\mathscr{A} = \{a_1, \ldots, a_m\}$, кодовый алфавит $\mathscr{B} = \{b_1, \ldots, b_D\}$ и распределение $\vec{p} = (p_1, \ldots, p_m)$. Тогда средней длиной кодового слова при алфавитном кодировании $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ (или, коротко, средней длиной кода $\varphi(\mathscr{A})$) называется величина

$$\ell^{\varphi} = \sum_{i=1}^{m} p_i \ell_i \,.$$

Определение 3.6. Алфавитное кодирование $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ (и код $\varphi(\mathscr{A})$) называется оптимальным, если φ однозначно декодируемо и при этом средняя длина ℓ^{φ} минимальна.

Замечание 3.3. 1. Из утверждения 3.1 (с. 74) и следствия 3.1 (с. 81) вытекает важный вывод: если оптимальное кодирование существует, то существует и префиксное кодирование с таким же набором длин кодовых слов и это префиксное кодирование также оптимальное. Поэтому далее при описании оптимальных кодов мы можем ограничиться только префиксными кодами.

2. Так как длины кодовых слов — натуральные числа, $\ell_i \geqslant 1$, для любого кодирования φ верно неравенство $\ell^{\varphi} \geqslant 1$. Поэтому при $m \leqslant D$ задача о нахождении оптимального кодирования тривиальна. Действительно, положим $\varphi(a_i) = b_i$, $1 \leqslant i \leqslant m$. Очевидно, $\ell^{\varphi} = 1$, поэтому такое кодирование оптимально.

Отметим, что из определения оптимального кодирования ещё не следует, что такое кодирование существует. Действительно, класс всевозможных кодирований бесконечен, и, вообще говоря, неясно,

достигается ли точная нижняя грань $\inf_\varphi \ell^\varphi$ на каком-то кодировании φ. Докажем, что эта точная нижняя грань достигается.

Утверждение 3.2. *Оптимальное кодирование φ существует.*

Доказательство. Пусть L — такое наименьшее натуральное число, что $m \leqslant D^L$. Тогда существует равномерное кодирование φ, для которого длины всех кодовых слов равны L. Следовательно, $\inf_\varphi \ell^\varphi \leqslant L$, и поэтому величина $\inf_\varphi \ell^\varphi$ не изменится, если мы исключим из рассмотрения все кодирования φ, для которых $\ell^\varphi > L$.

Для кодирований φ с условием $\ell^\varphi = \sum_{i=1}^m p_i \ell_i \leqslant L$ заметим, что если $p_i = 0$, то значение ℓ_i несущественно при вычислении ℓ^φ, а если $p_i > 0$, то $\ell_i \leqslant L/p_i \leqslant L/p$, где p — наименьшее из чисел $p_1, \ldots, p_m$, не равных нулю. Таким образом, значение $\inf_\varphi \ell^\varphi$ определяется классом кодирований φ, для которых все длины ℓ_i кодовых слов с $p_i > 0$ не превосходят L/p, а длины остальных кодовых слов не учитываются. При таких ограничениях средняя длина ℓ^φ может принимать лишь конечное число различных значений и поэтому достигает минимума при некотором φ. □

Теорема 3.4. *Если алфавитное кодирование φ однозначно декодируемо, то справедливо неравенство*

$$\ell^\varphi \geqslant \frac{H(\vec{p})}{\log_2 D},$$

причём равенство имеет место тогда и только тогда, когда все положительные координаты вектора $\vec{p}$ имеют вид $p_i = D^{-\ell_i}$.

Доказательство. Рассмотрим разность

$$\begin{aligned} H(\vec{p}) - \ell^\varphi \log_2 D &= -\sum_{i=1}^m p_i \log_2 p_i - \log_2 D \sum_{i=1}^m p_i l_i = \\ &= -\sum_{i=1}^m p_i \log_2 p_i + \sum_{i=1}^m p_i \log_2 D^{-\ell_i} = \\ &= \sum_{i=1}^m p_i \log_2 \frac{D^{-\ell_i}}{p_i}. \end{aligned}$$

При этом мы можем считать, что суммирование распространяется только на индексы i с условием $p_i > 0$, поскольку при $p_i = 0$ соответствующие слагаемые в выражениях для $H(\vec{p})$ и ℓ^φ равны нулю.

Применяя неравенство (1.3) (с. 24) и теорему 3.3 (неравенство Мак-Миллана), получим

$$H(\vec{p}) - \ell^{\varphi} \log_2 D \leqslant \log_2 e \sum_{i=1}^{m} p_i \Big(\frac{D^{-\ell_i}}{p_i} - 1 \Big) =$$
$$= \log_2 e \Big(\sum_{i=1}^{m} D^{-\ell_i} - \sum_{i=1}^{m} p_i \Big) \leqslant 0.$$

Вторая часть теоремы получается из условия обращения неравенства (1.3) в равенство. □

Теорема 3.5. *Если распределение $\vec{p}$ невырожденное, то существует такое префиксное кодирование φ, для которого справедливо неравенство*

$$\ell^{\varphi} < 1 + \frac{H(\vec{p})}{\log_2 D}. \tag{3.5}$$

Доказательство. Если $p_i > 0$, то обозначим через ℓ_i наименьшее натуральное число с условием $\ell_i \geqslant -\log_D p_i$. Так как $p_i < 1$,

$$0 \leqslant \ell_i - 1 < -\log_D p_i \leqslant \ell_i, \tag{3.6}$$

или, что то же самое,

$$D^{-\ell_i} \leqslant p_i < D^{-\ell_i + 1}. \tag{3.7}$$

Из левого неравенства (3.7) получаем

$$\sum_{i:\, p_i > 0} D^{-\ell_i} \leqslant \sum_{i=1}^{m} p_i = 1. \tag{3.8}$$

Если все вероятности $p_1, \ldots, p_m$ положительны, то нетрудно завершить доказательство. Действительно, по теореме 3.2 (неравенство Крафта) из (3.8) следует существование префиксного кодирования φ с длинами кодовых слов $\operatorname{len}(\varphi(a_i)) = \ell_i$, $1 \leqslant i \leqslant m$. Из левого неравенства (3.6) получаем

$$\ell^{\varphi} = \sum_{i:\, p_i > 0} p_i \ell_i < \sum_{i:\, p_i > 0} p_i (1 - \log_D p_i) =$$
$$= 1 - \sum_{i=1}^{m} p_i \log_D p_i = 1 + \frac{H(\vec{p})}{\log_2 D}.$$

Если же среди вероятностей $p_1, \ldots, p_m$ есть равные нулю, то нужны дополнительные рассуждения. Обозначим через m_1 число таких

индексов i, $1 \leqslant i \leqslant m$, для которых $p_i = 0$. Рассмотрим два возможных случая относительно суммы $K_1 = \sum_{i:\, p_i > 0} D^{-\ell_i}$.

Если $K_1 < 1$, то для указанных выше m_1 индексов i с условием $p_i = 0$ положим $\ell_i = \ell$, где натуральное ℓ выберем столь большим, чтобы выполнялось неравенство $K_1 + m_1 D^{-\ell} \leqslant 1$. Для набора $\ell_1, \ldots, \ell_m$ можем применить теорему 3.2, обеспечивающую существование префиксного кодирования φ с заданными длинами кодовых слов, и сделанная выше оценка для ℓ^{φ} сохраняется.

Если $K_1 = 1$, то из неравенства (3.8) следует, что все положительные числа p_i имеют вид $p_i = D^{-\ell_i}$. Рассуждая так же, как в доказательстве теоремы 3.4, получаем равенство

$$\sum_{i:\, p_i > 0} p_i \ell_i = \frac{H(\vec{p})}{\log_2 D}.$$

Не ограничивая общности, предположим, что $p_1 > 0$. Заменив в сумме

$$K_1 = \sum_{i:\, p_i > 0} D^{-\ell_i}$$

определённое выше число ℓ_1 на $\ell_1 + 1$, вместо значения $K_1 = 1$ получим значение

$$D^{-(\ell_1 + 1)} + \sum_{i > 1:\, p_i > 0} D^{-\ell_i} < 1.$$

Для указанных выше m_1 индексов i с условием $p_i = 0$ положим $\ell_i = \ell$, где натуральное ℓ выберем столь большим, чтобы выполнялось неравенство

$$D^{-(\ell_1 + 1)} + \sum_{i > 1:\, p_i > 0} D^{-\ell_i} + m_1 D^{-\ell} \leqslant 1.$$

Для набора $\ell_1 + 1, \ell_2, \ldots, \ell_m$ мы можем снова применить теорему 3.2, обеспечивающую существование префиксного кодирования φ с заданными длинами кодовых слов, при этом

$$\ell^{\varphi} = p_1(\ell_1 + 1) + \sum_{i > 1:\, p_i > 0} p_i \ell_i = p_1 + \sum_{i:\, p_i > 0} p_i \ell_i = p_1 + \frac{H(\vec{p})}{\log_2 D}.$$

Так как распределение $\vec{p}$ невырожденное, $p_1 < 1$. Следовательно, и в случае $K_1 = 1$ получаем неравенство

$$\ell^{\varphi} < 1 + \frac{H(\vec{p})}{\log_2 D}. \qquad \square$$

Замечание 3.4. Если распределение $\vec{p}$ вырожденное, то утверждение теоремы 3.5 неверно. В этом случае $H(\vec{p}) = 0$, так что неравенство (3.5) принимает вид $\ell^{\varphi} < 1$, но $\ell^{\varphi} \geqslant 1$ для любого кодирования φ.

Для вырожденного распределения существует префиксное кодирование φ с условием $\ell^{\varphi} = 1$, так что неравенство (3.5) превращается в равенство. Действительно, символу a с вероятностью 1 сопоставим кодовое слово 0, а всем остальным символам алфавита $\mathscr{A}$ так сопоставим кодовые слова, начинающиеся с 1, чтобы получился префиксный код.

Следствие 3.2. *Средняя длина оптимального алфавитного кодирования φ удовлетворяет неравенствам*

$$\frac{H(\vec{p})}{\log_2 D} \leqslant \ell^{\varphi} \leqslant 1 + \frac{H(\vec{p})}{\log_2 D}.$$

Доказательство получается применением теорем 3.4 и 3.5 и замечания 3.4. Действительно, левое неравенство верно для любого алфавитного кодирования φ, а правое — для некоторого, а потому и для оптимального. □

Мы начали этот параграф с рассуждений относительно дискретного источника без памяти. Предположим теперь, что имеется произвольный дискретный источник сообщений $(\Omega, \mathscr{F}, \mathsf{P})$ (или, что то же самое, дискретный сл. процесс $A = (A_1, A_2, \ldots)$).

Откажемся от продолжения алфавитного кодирования $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ на последовательности символов алфавита $\mathscr{A}$ по правилу сцепления, а вместо этого будем кодировать целые n-символьные блоки. Другими словами, для любого $n \geqslant 1$ рассмотрим алфавит $\mathscr{A}^n$ с заданным на нём распределением $\mathsf{P}_{1,2,\ldots,n}$, и пусть φ_n — соответствующее этому распределению *оптимальное* D-ичное кодирование $\varphi_n\colon \mathscr{A}^n \to \mathscr{B}^*$ со средней длиной кодового слова, равной

$$\ell^{\varphi_n} = \mathsf{E}\ \mathrm{len}(\varphi_n(A^n)) = \sum_{a^n \in \mathscr{A}^n} \mathsf{P}_{1,2,\ldots,n}(a^n)\, \mathrm{len}(\varphi_n(a^n)),$$

где A^n — случайная n-символьная последовательность, порождаемая источником. Что можно сказать о среднем числе символов кодового алфавита $\mathscr{B}$, приходящемся на один символ алфавита $\mathscr{A}$,

т. е. о величине

$$\ell^{(n)} = \frac{\ell^{\varphi_n}}{n},$$

при большой длине блока n? Следующий важный результат известен как *теорема кодирования при отсутствии шума*.

Теорема 3.6. *Если для дискретного источника сообщений существует предельная энтропия H_∞, то существует предел*

$$\lim_{n\to\infty} \ell^{(n)} = \frac{H_\infty}{\log_2 D}.$$

Доказательство. Применяя к нашему случаю следствие 3.2, мы должны заменить ℓ^{φ} на ℓ^{φ_n}, а энтропию $H(\vec{p})$ — на $H(A^n)$. Получим неравенства

$$\frac{H(A^n)}{\log_2 D} \leqslant \ell^{\varphi_n} \leqslant 1 + \frac{H(A^n)}{\log_2 D}.$$

Разделим почленно на n и, согласно определению 2.5 (с. 53), подставим $H_n = H(A^n)/n$. Получим неравенства

$$\frac{H_n}{\log_2 D} \leqslant \ell^{(n)} \leqslant \frac{1}{n} + \frac{H_n}{\log_2 D}.$$

Так как по условию теоремы существует предел $\lim_{n\to\infty} H_n = H_\infty$, левая и правая части неравенства сходятся к $H_\infty/\log_2 D$, а значит, и заключённая между ними величина $\ell^{(n)}$ сходится к тому же пределу. □

Замечание 3.5. Интересно сопоставить доказанную теорему и пример 2.4 (с. 67) — при $D = 2$ получается, что простой способ кодирования на основе свойства асимптотической равнораспределённости даёт такой же предельный результат для средней длины кодового слова, как и оптимальное префиксное кодирование длинных блоков. Теорема 3.6 доказана в более общих условиях, чем пример 2.4: вместо свойства асимптотической равнораспределённости требуется только существование предельной энтропии. Как было установлено выше, предельная энтропия существует для широкого класса дискретных источников сообщений, например, для всех стационарных источников.

§ 3.3. Алгоритм оптимального кодирования

Вначале мы рассмотрим два алгоритма построения двоичных кодов ($D = 2$), алгоритмы Фано и Хаффмана, поскольку, во-первых, двоичные коды наиболее важны для практических приложений и, во-вторых, на примере двоичного кодового алфавита легче понять идеи, положенные в основу этих алгоритмов.

Алгоритм Фано мы приводим по той причине, что это исторически один из первых алгоритмов построения префиксных двоичных кодов, он весьма простой и при этом позволяет получать коды с достаточно малой средней длиной — в большинстве случаев применения алгоритма Фано получается либо оптимальный код, либо код, близкий к оптимальному.

Алгоритм Хаффмана, в отличие от алгоритма Фано, всегда строит оптимальный префиксный код. Обычно в учебной литературе оптимальность кода Хаффмана доказывается только в двоичном случае, а для $D > 2$ лишь указывается на возможность «аналогичного» доказательства. Однако строгое рассмотрение недвоичного случая связано с преодолением некоторых трудностей; мы приведём полное доказательство в общем случае.

Всюду в этом параграфе предполагаем, что на алфавите источника сообщений $\mathscr{A} = \{a_1, \ldots, a_m\}$ $(m \geqslant 2)$ задано распределение вероятностей $\vec{p} = (p_1, \ldots, p_m)$, причём вероятности символов упорядочены в невозрастающем порядке, т. е. $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$.

Алгоритм Фано ($D = 2$, $\mathscr{B} = \{0, 1\}$)

Выберем число k, $1 \leqslant k < m$, так, чтобы величина

$$\left| \sum_{i=1}^{k} p_i - \sum_{i=k+1}^{m} p_i \right|$$

была минимальной. Разобьём множество $\mathscr{A} = \{a_1, \ldots, a_m\}$ на подмножества: $\mathscr{A} = \mathscr{A}_0 \sqcup \mathscr{A}_1$, где $\mathscr{A}_0 = \{a_1, \ldots, a_k\}$, $\mathscr{A}_1 = \{a_{k+1}, \ldots, a_m\}$.

Следующие шаги алгоритма определим индуктивно. Предположим, что уже задано подмножество $\mathscr{A}_{i_1, \ldots, i_t} \subseteq \mathscr{A}$, где $i_1, \ldots, i_t \in \mathscr{B}$.

Если подмножество $\mathscr{A}_{i_1, \ldots, i_t}$ состоит из единственного символа a_j, то для этого символа определяем кодирование $\varphi(a_j) = i_1 \ldots i_t$. Если же в подмножестве $\mathscr{A}_{i_1, \ldots, i_t} = \{a_j, \ldots, a_s\}$ не менее двух символов, то

выберем число k, $j \leqslant k < s$, так, чтобы минимизировать величину

$$\Big|\sum_{i=j}^{k} p_i - \sum_{i=k+1}^{s} p_i\Big|,$$

и разобьём множество $\mathscr{A}_{i_1,\ldots,i_t}$ на подмножества $\mathscr{A}_{i_1,\ldots,i_t,0} = \{a_j, \ldots$ $\ldots, a_k\}$ и $\mathscr{A}_{i_1,\ldots,i_t,1} = \{a_{k+1}, \ldots, a_s\}$.

Процесс разбиения на подмножества продолжается, пока не получим все одноэлементные подмножества и тем самым не определим кодирование φ всюду на алфавите $\mathscr{A}$.

Замечание 3.6. В описанном алгоритме имеется неоднозначность в выборе значения k, поскольку указанное выше минимальное значение модуля разности сумм вероятностей может достигаться при различных k. Можно принять соглашение о том, что в случае такой неоднозначности выбирается, например, наименьшее возможное значение k.

Замечание 3.7. Очевидно, что в результате применения алгоритма Фано получается префиксное кодирование, так как фактически происходит построение кодового дерева от корня $\mathscr{A}$ к листьям — одноэлементным подмножествам $\mathscr{A}_{i_1,\ldots,i_t}$ (см. рисунок 3.4).

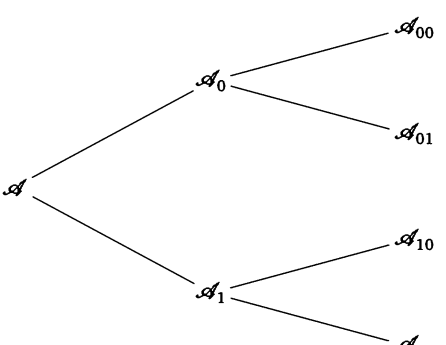

Рис. 3.4. Схема алгоритма Фано

Пример 3.5. Рассмотрим алфавит $\mathscr{A} = \{a, b, c, d, e\}$ и распределение вероятностей $\vec{p} = (0{,}3;\ 0{,}2;\ 0{,}2;\ 0{,}2;\ 0{,}1)$. Последовательные разбие-

ния подмножеств в алгоритме Фано можно наглядно представить следующей диаграммой.

a	0,3	0	0	
b	0,2	0	1	
c	0,2	1	0	
d	0,2	1	1	0
e	0,1	1	1	1

Рис. 3.5. Пример алгоритма Фано

В результате получается кодирование $\varphi(a) = 00$, $\varphi(b) = 01$, $\varphi(c) = 10$, $\varphi(d) = 110$, $\varphi(e) = 111$ со средней длиной $\ell^{\varphi} = (0{,}3 + 0{,}2 + 0{,}2) \cdot 2 + (0{,}2 + 0{,}1) \cdot 3 = 2{,}3$ бита.

Алгоритм Хаффмана ($D = 2$, $\mathscr{B} = \{0, 1\}$)

1-й этап — построение двоичного дерева. Будем строить двоичное дерево с m листьями, начиная с листьев и продвигаясь к корню. Возьмём в качестве листьев дерева символы $a_1, \ldots, a_m$, которым приписаны вероятности $p_1, \ldots, p_m$.

Основная операция алгоритма — *редукция* — состоит в следующем. Возьмём две вершины a_{m-1} и a_m с *наименьшими* вероятностями p_{m-1} и p_m и добавим в дерево новую вершину $a_{m-1} \cup a_m$, которой припишем вероятность $p_{m-1} + p_m$. Вершину $a_{m-1} \cup a_m$ соединим рёбрами с вершинами a_{m-1} и a_m и объявим общим предком для этих вершин. Ребро от $a_{m-1} \cup a_m$ к a_{m-1} пометим символом 0, а ребро от $a_{m-1} \cup a_m$ к a_m — символом 1. Получаем новый (редуцированный) алфавит $\mathscr{A}^{(1)} = \{a_1, \ldots, a_{m-2}, a_{m-1} \cup a_m\}$ с набором вероятностей $\vec{p}^{\,(1)} = (p_1, \ldots, p_{m-2}, p_{m-1} + p_m)$.

Если в полученном алфавите $\mathscr{A}^{(1)}$ не менее двух символов, то упорядочим эти символы в порядке невозрастания вероятностей набора $\vec{p}^{\,(1)}$ и снова применим операцию редукции.

Если же алфавит $\mathscr{A}^{(1)}$ состоит из одного символа с приписанной ему вероятностью 1, то этот один символ объявим корнем и завершим этап построения дерева.

2-й этап — кодирование. Чтобы получить кодовое слово $\varphi(a_j)$ для символа a_j, последовательно считываем метки рёбер на пути от корня дерева к листу a_j.

Замечание 3.8. В описанном алгоритме имеется неоднозначность в операции редукции — как выбирать две вершины с наименьшими вероятностями, если имеется больше двух вершин с одинаковыми наименьшими вероятностями? Примем соглашение о том, что в случае такой неоднозначности выбираются две вершины с наибольшими номерами в списке вершин. Можно было бы принять и другое соглашение о правиле выбора, и тогда, вообще говоря, получалось бы другое дерево и другое кодирование. Можно доказать, что при любом таком соглашении о выборе пары вершин для редукции будет получаться оптимальное кодирование, т. е. с точки зрения средней длины кода несущественно, какое именно соглашение будет принято. Мы отложим доказательство до рассмотрения общего случая алгоритма Хаффмана.

Замечание 3.9. Очевидно, что алгоритм Хаффмана приводит к префиксному кодированию, так как фактически происходит построение размеченного кодового дерева (см. п. 1 теоремы 3.1).

Пример 3.6. Применим алгоритм Хаффмана к данным, рассмотренным в примере 3.5. Шаги алгоритма представлены на следующем рисунке.

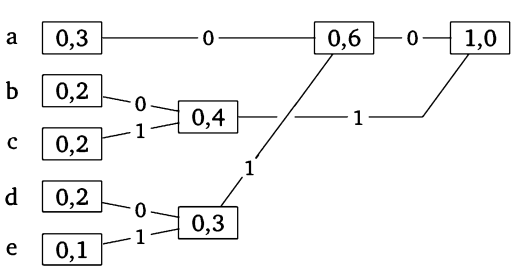

Рис. 3.6. Пример алгоритма Хаффмана

В результате получается кодирование $\varphi(a) = 00$, $\varphi(b) = 10$, $\varphi(c) = 11$, $\varphi(d) = 010$, $\varphi(e) = 011$ со средней длиной $\ell^{\varphi} = (0{,}3 + 0{,}2 + 0{,}2) \cdot 2 + (0{,}2 + 0{,}1) \cdot 3 = 2{,}3$ бита.

Для рассмотрения алгоритма Хаффмана в общем случае (не обязательно двоичном, $D \geqslant 2$) нам понадобится несколько лемм.

Лемма 3.3. *Пусть* $\varphi : \mathscr{A} \to \mathscr{B}^*$ *— оптимальное* D*-ичное алфавитное кодирование с длинами кодовых слов* $\operatorname{len}(\varphi(a_i)) = \ell_i$, $1 \leqslant i \leqslant m$, *соответствующее распределению* $\vec{p} = (p_1, \ldots, p_m)$. *Если* $p_i > p_j$, *то* $\ell_i \leqslant \ell_j$.

Доказательство. Предположим противное, т. е. $p_i > p_j$, но $\ell_i > \ell_j$. Рассмотрим кодирование φ_1, которое отличается от φ только перестановкой двух кодовых слов, т. е. $\varphi_1(a_i) = \varphi(a_j)$, $\varphi_1(a_j) = \varphi(a_i)$. Тогда разность средних длин ℓ^{φ} и ℓ^{φ_1} представится в виде

$$\ell^{\varphi} - \ell^{\varphi_1} = \ell_i p_i + \ell_j p_j - \ell_i p_j - \ell_j p_i = (\ell_i - \ell_j)(p_i - p_j) > 0.$$

Отсюда следует, что $\ell^{\varphi} > \ell^{\varphi_1}$, что противоречит предположению об оптимальности кодирования φ. □

Лемма 3.4. *Если* $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$, *то существует оптимальное* D*-ичное алфавитное кодирование* φ, *для которого длины кодовых слов удовлетворяют неравенствам* $\ell_1 \leqslant \ell_2 \leqslant \ldots \leqslant \ell_m$.

Доказательство. Вначале заметим, что если $p_i = p_j$, то описанная в доказательстве леммы 3.3 перестановка кодовых слов не изменяет средней длины кода. Поэтому если φ — оптимальное кодирование, то любая перестановка кодовых слов между символами алфавита $\mathscr{A}$, имеющими равные вероятности, приводит к кодированию φ_1, которое также является оптимальным.

Пусть φ — некоторое оптимальное кодирование. Предположим, что $m = m_1 + m_2 + \ldots + m_s$, и упорядочение $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$ можно записать в виде

$$p_1 = \ldots = p_{m_1} > p_{m_1+1} = \ldots = p_{m_1+m_2} > \ldots > p_{m_1+\ldots+m_{s-1}+1} = \ldots = p_m,$$

другими словами, набор вероятностей $(p_1, \ldots, p_m)$ разбивается на s поднаборов длин $m_1, \ldots, m_s$, внутри которых вероятности равны, а на границах поднаборов — строгие неравенства. Согласно сделанному выше замечанию, перестановкой кодовых слов внутри каждого поднабора можно получить оптимальное кодирование φ_1, для которого длины кодовых слов удовлетворяют неравенствам

$$\ell_1 \leqslant \ldots \leqslant \ell_{m_1}, \quad \ell_{m_1+1} \leqslant \ldots \leqslant \ell_{m_1+m_2}, \ldots, \quad \ell_{m_1+\ldots+m_{s-1}+1} \leqslant \ldots \leqslant \ell_m,$$

а в силу леммы 3.3 на границах между поднаборами справедливы неравенства

$$\ell_{m_1} \leqslant \ell_{m_1+1}, \quad \ell_{m_1+m_2} \leqslant \ell_{m_1+m_2+1}, \ldots, \quad \ell_{m_1+\ldots+m_{s-1}} \leqslant \ell_{m_1+\ldots+m_{s-1}+1}. \quad \square$$

Пусть $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ — некоторое D-ичное префиксное алфавитное кодирование и G — соответствующее кодовое дерево, т. е. размеченное D-ичное корневое дерево с m листьями, которым сопоставлены кодовые слова $\varphi(a_1), \ldots, \varphi(a_m)$ для символов $a_1, \ldots, a_m$ с приписанными им вероятностями $p_1, \ldots, p_m$. Рассмотрим две операции, позволяющие получить новое D-ичное префиксное алфавитное кодирование φ_1 и соответствующее кодовое дерево G_1.

Пусть u — лист дерева G, находящийся на уровне $\ell \geqslant 2$, и метка $\mu(u) = b_{i_1} \ldots b_{i_{\ell-1}} b_{i_\ell}$ листа u является кодовым словом $\varphi(a_j)$ для символа a_j с приписанной ему вероятностью p_j.

Операция 1 (удаление одиночного листа). Предположим, что лист u является единственным потомком некоторой вершины v, имеющей метку $\mu(v) = b_{i_1} \ldots b_{i_{\ell-1}}$. Рассмотрим новое кодирование φ_1, которое отличается от φ только тем, что кодовое слово $\varphi(a_j) = \mu(u) = b_{i_1} \ldots b_{i_{\ell-1}} b_{i_\ell}$ заменяется на кодовое слово $\varphi_1(a_j) = \mu(v) = b_{i_1} \ldots b_{i_{\ell-1}}$. Другими словами, из дерева G удаляем лист u вместе с ребром (v, u), и в полученном дереве G_1 вершина v становится листом.

Замечание 3.10. Так как в результате операции 1 длина одного кодового слова уменьшилась на один символ, $\ell^{\varphi} - \ell^{\varphi_1} = p_j \geqslant 0$, причём равенство $\ell^{\varphi_1} = \ell^{\varphi}$ имеет место тогда и только тогда, когда $p_j = 0$.

Операция 2 (перемещение листа к ненасыщенной вершине). Предположим, что v — неконцевая вершина на уровне $t \leqslant \ell - 1$, имеющая метку $\mu(v) = b_{k_1} \ldots b_{k_t}$ и являющаяся *ненасыщенной* вершиной, т. е. v имеет менее D потомков. Пусть b — символ кодового алфавита $\mathscr{B}$, не встретившийся в качестве метки ребра от вершины v к её потомкам. Рассмотрим новое кодирование φ_1, которое отличается от φ только тем, что кодовое слово $\varphi(a_j) = \mu(u) = b_{i_1} \ldots b_{i_\ell}$ заменяется на кодовое слово $\varphi_1(a_j) = \mu(v)b = b_{k_1} \ldots b_{k_t} b$. Другими словами, из дерева G удаляем лист u вместе с ведущим к нему ребром, но добавляем вершину u в качестве нового листа — потомка вершины v, и ребро (v, u) в новом дереве G_1 помечаем символом b.

Замечание 3.11. Так как в результате операции 2 длина одного кодового слова уменьшилась на $\ell - t - 1 \geqslant 0$ символов, разность $\ell^{\varphi} - \ell^{\varphi_1} = (\ell - t - 1) \times p_j \geqslant 0$, причём равенство $\ell^{\varphi_1} = \ell^{\varphi}$ имеет место тогда и только тогда, когда $t = \ell - 1$ или $p_j = 0$.

Лемма 3.5. *Существует такое оптимальное D-ичное префиксное алфавитное кодирование φ, для которого кодовое дерево* G *является насыщенным.*

Доказательство. Пусть длины кодовых слов равны $\ell_1, \ldots, \ell_m$, тогда $L = \max(\ell_1, \ldots, \ell_m)$ есть высота дерева G. Индукцией по $L \geqslant 1$ докажем, что существует алгоритм, который за конечное число операций 1 и 2 из оптимального кодового дерева G получает насыщенное оптимальное кодовое дерево.

Если $L = 1$, то с необходимостью $m \leqslant D$, и дерево G представляет собой корень с m потомками; очевидно, такое дерево насыщено. Предположим, что искомый алгоритм существует для любых оптимальных кодовых деревьев, высота которых строго меньше L, и построим такой алгоритм для оптимального кодового дерева G высоты $L > 1$. Пусть s есть число листьев на уровне L. Проведём индукцию по $s \geqslant 1$.

Пусть $s = 1$. Это значит, что на последнем уровне G имеется ровно один лист u, и к этому листу применима операция 1, дающая новое кодирование φ_1 и соответствующее кодовое дерево G_1. Если вероятность p_j, соответствующая листу u, положительна, то, в силу замечания 3.10, справедливо неравенство $\ell^{\varphi_1} < \ell^{\varphi}$, что противоречит оптимальности кодирования φ. Следовательно, $p_j = 0$, $\ell^{\varphi_1} = \ell^{\varphi}$, и кодирование φ_1 также оптимально. По предположению индукции, для оптимального кодового дерева G_1 высоты $L - 1$ существует алгоритм, позволяющий за конечное число операций 1 и 2 получить насыщенное оптимальное кодовое дерево. Добавив к этому алгоритму одну операцию 1, получим искомый алгоритм для дерева G.

Пусть $s \geqslant 2$. Рассмотрим два возможных случая.

Предположим, что в дереве G на некотором уровне $t < L - 1$ имеется ненасыщенная неконцевая вершина v. Возьмём произвольный лист u на уровне L и применим операцию 2 перемещения листа u к вершине v. Получим новое кодирование φ_1 и кодовое дерево G_1. Если вероятность p_j, соответствующая листу u, положительна, то,

в силу замечания 3.11, справедливо неравенство $\ell^{\varphi_1} < \ell^{\varphi}$, что противоречит оптимальности кодирования φ. Следовательно, $p_j = 0$, $\ell^{\varphi_1} = \ell^{\varphi}$, и кодирование φ_1 также оптимально. По предположению индукции, для оптимального кодового дерева G_1, имеющего $s - 1$ вершин на уровне L, существует алгоритм, позволяющий за конечное число операций 1 и 2 получить насыщенное оптимальное кодовое дерево. Добавив к этому алгоритму одну операцию 2, получим искомый алгоритм для дерева G.

Предположим теперь, что в дереве G ни на одном из уровней $t < L - 1$ нет ненасыщенных неконцевых вершин, т. е. невозможно операцией 2 уменьшить число листьев на уровне L. Это значит, что в дереве G все неконцевые вершины имеют D потомков, за исключением, может быть, нескольких вершин на уровне $L - 1$. Если таких вершин нет, то дерево G насыщенное и никаких операций делать не нужно; в противном случае рассмотрим $v_1, \ldots, v_k$ — все ненасыщенные неконцевые вершины на уровне $L - 1$, имеющие соответственно $s_1, \ldots, s_k$ потомков, $2 \leqslant s_j < D$, $s_1 + \ldots + s_k = s_0 \leqslant s$. Если при помощи операции 2 перемещать потомков от одной из ненасыщенных вершин $v_1, \ldots, v_k$ к другой, то средняя длина кода не изменяется (поскольку перемещение происходит на одном уровне), и снова получается оптимальный код. Будем проводить такие перемещения листьев в следующем порядке: сначала все возможные перемещения от вершин $v_2, \ldots, v_k$ к вершине v_1, затем от вершин $v_3, \ldots, v_k$ к вершине v_2 и т. д.

Результат будет зависеть от того, каковы неполное частное t и остаток r при делении s_0 на D, $s_0 = tD + r$, $0 \leqslant r \leqslant D - 1$. Если $t = 0$, то вершина v_1 будет иметь s_0 потомков, $2 \leqslant s_0 = r < D$, а вершины $v_2, \ldots, v_k$ станут листьями, т. е. получим насыщенное дерево с особой вершиной v_1. Если $t > 0$ и $r = 0$, то $v_1, \ldots, v_t$ станут насыщенными неконцевыми вершинами, а вершины $v_{t+1}, \ldots, v_k$ станут листьями, т. е. получим насыщенное дерево без особой вершины.

Наконец, если $t > 0$ и $r > 0$, то $v_1, \ldots, v_t$ станут насыщенными неконцевыми вершинами, вершина v_{t+1} будет иметь $r < D$ потомков, а вершины $v_{t+2}, \ldots, v_k$ станут листьями. В этом последнем случае остаётся рассмотреть значение r. Если $r > 1$, то v_{t+1} — единственная особая вершина, а полученное дерево насыщенное. Пусть $r = 1$, т. е. вершина v_{t+1} имеет единственного потомка u. Тогда

с необходимостью этому потомку приписана нулевая вероятность, и при помощи операции 1 получаем насыщенное кодовое дерево. Таким образом, и в этом случае искомый алгоритм насыщения кодового дерева существует и сводится к конечному числу операций 2 и, возможно, одной операции 1. □

Определение 3.7. Префиксное алфавитное кодирование φ называется *приведённым*, если соответствующее кодовое дерево G насыщено и, кроме того,

1) если в дереве G на предпоследнем уровне есть особая вершина, то её D_0 потомкам соответствуют символы алфавита $\mathscr{A}$ с D_0 наименьшими вероятностями из набора $p_1, \ldots, p_m$;
2) если в дереве G нет особой вершины, то имеется неконцевая вершина на предпоследнем уровне, D потомкам которой соответствуют символы алфавита $\mathscr{A}$ с D наименьшими вероятностями из набора $p_1, \ldots, p_m$.

Лемма 3.6. *Существует приведённое оптимальное D-ичное префиксное алфавитное кодирование φ.*

Доказательство. Рассмотрим оптимальное кодирование φ, для которого кодовое дерево G насыщено; такое кодирование существует согласно лемме 3.5. Обозначим длины кодовых слов ℓ_i, $1 \leqslant i \leqslant m$.

Лемму 3.3 можно сформулировать иначе: если $\ell_i > \ell_j$, то $p_i \leqslant p_j$. Это значит, что листьям оптимального кодового дерева, находящимся на уровне $L = \max(\ell_1, \ldots, \ell_m)$, приписаны *наименьшие* вероятности из набора $p_1, \ldots, p_m$. Если переставлять между данными листьями символы алфавита $\mathscr{A}$ с приписанными им вероятностями, то средняя длина кода не изменится, а дерево останется насыщенным. Очевидно, что подходящей перестановкой можно добиться выполнения условий определения 3.7. □

Замечание 3.12. Напомним, что m есть число символов алфавита $\mathscr{A}$, а m_0 — такое наименьшее целое число вида $m_0 = 1 + k(D-1)$, что $m_0 \geqslant m$. Положим

$$s = \begin{cases} D, & \text{если } m_0 = m, \\ D_0 = m - m_0 + D, & \text{если } m_0 > m. \end{cases}$$

Принимая во внимание лемму 3.2 (с. 82), условия 1 и 2 определения 3.7 можно записать так: в приведённом кодировании наименее вероятным s символам алфавита $\mathscr{A}$ соответствуют кодовые слова одинаковой максимальной длины, различающиеся между собой только в последнем символе.

Операция 3 (редукция приведённого кодирования). Пусть $m > D$, $\varphi\colon \mathscr{A} \to \mathscr{B}^*$ — приведённое D-ичное префиксное алфавитное кодирование, $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$. Предположим, что кодовые слова $\varphi(a_{m-s+1}), \ldots, \varphi(a_m)$ для s символов $a_{m-s+1}, \ldots, a_m$ с наименьшими вероятностями $p_{m-s+1}, \ldots, p_m$ имеют одинаковую длину L, общий префикс $b_{j_1}, \ldots, b_{j_{L-1}}$, и различаются между собой только в последнем символе.

Рассмотрим алфавит $\mathscr{A}^{(1)} = \{a_1, \ldots, a_{m-s}, \sigma\}$, отличающийся от алфавита $\mathscr{A}$ удалением s символов $a_{m-s+1}, \ldots, a_m$ и добавлением одного символа $\sigma = a_{m-s+1} \cup \ldots \cup a_m$, а также заданное на алфавите $\mathscr{A}^{(1)}$ распределение вероятностей $\vec{p}^{(1)} = (p_1, \ldots, p_{m-s}, \gamma)$, где $\gamma = p_{m-s+1} + \ldots + p_m$. Зададим кодирование $\varphi_1\colon \mathscr{A}^{(1)} \to \mathscr{B}^*$ соотношениями

$$\varphi_1(a_i) = \varphi(a_i), \quad 1 \leqslant i \leqslant m-s,$$
$$\varphi_1(\sigma) = b_{j_1} \ldots b_{j_{L-1}}.$$

Кодирование φ_1 будем называть редуцированным по отношению к кодированию φ.

Лемма 3.7. *Пусть φ — приведённое D-ичное префиксное алфавитное кодирование и φ_1 — редуцированное кодирование. Кодирование φ оптимально для распределения $\vec{p} = (p_1, \ldots, p_m)$ тогда и только тогда, когда кодирование φ_1 оптимально для распределения $\vec{p}^{(1)} = (p_1, \ldots, p_{m-s}, \gamma)$.*

Доказательство. Из определения операции редукции вытекает соотношение $\ell^{\varphi} - \ell^{\varphi_1} = \gamma L - \gamma(L-1) = \gamma$, или

$$\ell^{\varphi} = \ell^{\varphi_1} + \gamma. \tag{3.9}$$

Предположим, что кодирование φ оптимально, и докажем, что кодирование φ_1 оптимально. Допустим, что φ_1 не оптимально. Тогда существует такое кодирование $\psi_1\colon \mathscr{A}^{(1)} \to \mathscr{B}^*$, что

$$\ell^{\psi_1} < \ell^{\varphi_1}. \tag{3.10}$$

Зададим кодирование $\psi\colon \mathscr{A} \to \mathscr{B}^*$ соотношениями

$$\begin{aligned}
\psi(a_i) &= \psi_1(a_i), \qquad 1 \leqslant i \leqslant m-s,\\
\psi(a_{m-s+1}) &= \psi_1(\sigma)b_1,\\
\psi(a_{m-s+2}) &= \psi_1(\sigma)b_2,\\
&\ldots\ldots\ldots\ldots\ldots\\
\psi(a_m) &= \psi_1(\sigma)b_s.
\end{aligned}$$

Тогда

$$\ell^{\psi} = \sum_{i=1}^{m-s} p_i \operatorname{len}(\psi_1(a_i)) + \gamma(\operatorname{len}(\psi_1(\sigma)) + 1) = \ell^{\psi_1} + \gamma. \tag{3.11}$$

Из (3.9), (3.10) и (3.11) следует неравенство $\ell^{\psi} < \ell^{\varphi}$, что противоречит оптимальности φ.

Докажем теперь, что из оптимальности кодирования φ_1 вытекает оптимальность кодирования φ. Допустим, что φ не оптимально. Согласно лемме 3.6 существует приведённое оптимальное кодирование ψ, $\ell^{\psi} < \ell^{\varphi}$. Поскольку кодирование ψ приведённое, к нему можем применить операцию редукции, в результате получим новое кодирование ψ_1, для которого $\ell^{\psi} = \ell^{\psi_1} + \gamma$. Принимая во внимание (3.9), получаем неравенство $\ell^{\psi_1} < \ell^{\varphi_1}$, что противоречит оптимальности φ_1. □

Леммы 3.3, 3.4, 3.5 и 3.6 постепенно сужали класс, в котором мы ищем оптимальное кодирование. Операция редукции и лемма 3.7 позволяют завершить этот поиск и сформулировать алгоритм построения оптимального D-ичного кодирования. Принимая во внимание п. 2 замечания 3.3 (с. 84), полагаем $m \geqslant D$.

Алгоритм Хаффмана ($D \geqslant 2$)

1-й этап — слияние. Пусть задан алфавит $\mathscr{A} = \{a_1, \ldots, a_m\}$ с распределением вероятностей $\vec{p} = (p_1, \ldots, p_m)$, причём $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$.

Если $m = D$, переходим ко второму этапу.

Если же $m > D$, выполняем *слияние* наименее вероятных s символов:

$$\begin{aligned}
\mathscr{A} &\to \mathscr{A}^{(1)} = \{a_1, \ldots, a_{m-s}, \sigma\},\\
\vec{p} &\to \vec{p}^{\,(1)} = (p_1, \ldots, p_{m-s}, \gamma),
\end{aligned}$$

где $\sigma = a_{m-s+1} \cup \ldots \cup a_m$, $\gamma = p_{m-s+1} + \ldots + p_m$. Упорядочиваем символы алфавита $\mathscr{A}^{(1)}$ в порядке невозрастания вероятностей распределения $\vec{p}^{(1)}$ и возвращаемся к началу первого этапа.

Замечание 3.13. В описании первого этапа используются введённые выше обозначения m_0, D_0, s и k. При первом слиянии мощность алфавита уменьшается на $s-1$ символов, при каждом последующем — на $D-1$ символов. Так как $m_0 = 1 + k(D-1)$, $m = D + (k-2)(D-1) + s - 1$, первый этап завершается построением цепочки из $k-1$ слияний

$$\mathscr{A} = \mathscr{A}^{(0)} \to \mathscr{A}^{(1)} \to \ldots \to \mathscr{A}^{(k-1)}, \tag{3.12}$$

приводящей к алфавиту из D символов, $|\mathscr{A}^{(k-1)}| = D$.

2-й этап — кодирование. Пусть на первом этапе алгоритма построена цепочка слияний (3.12). В качестве кодирования $\varphi_{k-1}\colon \mathscr{A}^{(k-1)} \to \mathscr{B}^*$ возьмём произвольное взаимно однозначное соответствие между алфавитами $\mathscr{A}^{(k-1)}$ и $\mathscr{B}$.

Предположим, что уже заданы кодирования $\varphi_{k-1}, \varphi_{k-2}, \ldots, \varphi_{i+1}$, и зададим кодирование $\varphi_i\colon \mathscr{A}^{(i)} \to \mathscr{B}^*$ следующим образом. Если алфавит $\mathscr{A}^{(i+1)}$ был получен из алфавита $\mathscr{A}^{(i)}$ слиянием некоторых s символов $a'_1, \ldots, a'_s$ в один новый символ σ', то кодирование φ_i получается из кодирования φ_{i+1} заменой одного равенства $\varphi_{i+1}(\sigma') = w \in \mathscr{B}^*$ на s равенств $\varphi_i(a'_j) = wb_j$, $1 \leqslant j \leqslant s$.

Второй этап алгоритма завершается построением искомого кодирования $\varphi = \varphi_0$.

Замечание 3.14. Несложным упражнением для читателя будет переформулировать алгоритм Хаффмана для $D \geqslant 2$ в терминах кодовых деревьев и проверить, что этот алгоритм является обобщением алгоритма Хаффмана для $D = 2$.

Пример 3.7. Построим оптимальное троичное кодирование для алфавита $\mathscr{A} = \{a, b, c, d, e, f, g, h\}$ с распределением вероятностей

$$\vec{p} = (0{,}2;\ 0{,}14;\ 0{,}13;\ 0{,}13;\ 0{,}12;\ 0{,}12;\ 0{,}08;\ 0{,}08).$$

Здесь $m = 8$, наименьшее m_0 с условием $m_0 = 1 + 2k \geqslant 8$ равно 9, поэтому в насыщенном троичном дереве есть особая вершина с $D_0 = m - m_0 + D = 2$ потомками. Шаги алгоритма Хаффмана представлены на следующем рисунке.

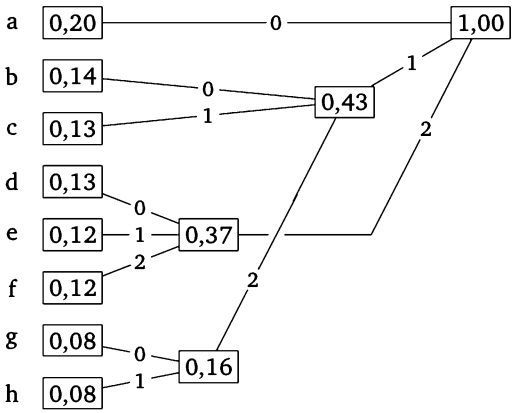

Рис. 3.7. Пример алгоритма Хаффмана ($D = 3$)

В результате получается кодирование $\varphi(a) = 0$, $\varphi(b) = 10$, $\varphi(c) = 11$, $\varphi(d) = 20$, $\varphi(e) = 21$, $\varphi(f) = 22$, $\varphi(g) = 120$, $\varphi(e) = 121$ со средней длиной $\ell^{\varphi} = 0{,}2 \cdot 1 + (0{,}14 + 0{,}13 + 0{,}13 + 0{,}12 + 0{,}12) \cdot 2 + (0{,}08 + 0{,}08) \cdot 3 = 1{,}96$ троичных символов.

Теорема 3.7. *Для любых m, $D \geqslant 2$ алгоритм Хаффмана строит оптимальное приведённое D-ичное префиксное алфавитное кодирование.*

Доказательство. Зафиксируем $D \geqslant 2$ и применим индукцию по $m \geqslant 2$. При $m \leqslant D$ утверждение теоремы очевидно.

Пусть $m > D$. Алгоритм Хаффмана на первом этапе строит цепочку слияний алфавитов (3.12) с заданными на них распределениями $\vec{p} = \vec{p}^{(0)} \to \vec{p}^{(1)} \to \ldots \to \vec{p}^{(k-1)}$, а на втором этапе, уже в обратном порядке, — цепочку кодирований $\varphi = \varphi_0 \leftarrow \varphi_1 \leftarrow \ldots \leftarrow \varphi_{k-1}$, или, что то же самое, цепочку кодовых деревьев $\mathrm{G} = \mathrm{G}_0 \leftarrow \mathrm{G}_1 \leftarrow \ldots \leftarrow \mathrm{G}_{k-1}$.

При этом из описания алгоритма ясно, что все кодовые деревья G_i в цепочке являются насыщенными, все кодирования φ_i — приведёнными, и каждое кодирование φ_i является редуцированным по отношению к кодированию φ_{i-1}, $1 \leqslant i \leqslant k-1$.

По предположению индукции, кодирование φ_1 является оптимальным для алфавита $\mathscr{A}^{(1)}$ с распределением $\vec{p}^{(1)}$, поэтому в силу леммы 3.7 кодирование $\varphi = \varphi_0$ является оптимальным для алфавита

$\mathscr{A} = \mathscr{A}^{(0)}$ с распределением $\vec{p} = \vec{p}^{(0)}$. □

Замечание 3.15. До сих пор наше внимание было сосредоточено на оптимальном посимвольном кодировании случайных последовательностей $A^n = (A_1, \ldots, A_n)$ символов, порождаемых простейшим дискретным источником сообщений — источником без памяти. Согласно определению 3.2 (с. 73), кодирование φ^* последовательности A^n сводится к применению *одного и того же* кодирования φ ко всем символам $A_1, \ldots, A_n$. Из равенства (3.4) (с. 84) видно, что при таком кодировании последовательностей для источника без памяти минимальная средняя длина кодового слова для всей последовательности достигается, если φ — оптимальное алфавитное кодирование.

Если же возникает задача оптимального алфавитного (посимвольного) кодирования последовательностей, порождаемых источником сообщений более сложной природы, рассмотренная выше схема продолжения кодирования уже не годится. Очевидно, что если распределение очередного случайного символа A_n, порождаемого источником, зависит от предшествующих символов $A_1, \ldots, A_{n-1}$, то и кодирование φ_n, применяемое к символу A_n, также должно зависеть от $A_1, \ldots, A_{n-1}$.

В качестве самостоятельного исследования предлагаем читателю рассмотреть задачу оптимального кодирования случайных последовательностей, порождаемых простым марковским источником $(\mathscr{A}, \vec{p}, Q)$. Пусть $\varphi_0 \colon \mathscr{A} \to \mathscr{B}^*$ — оптимальное алфавитное кодирование, соответствующее начальному распределению $\vec{p}$, и $\varphi_j \colon \mathscr{A} \to \mathscr{B}^*$, $1 \leqslant j \leqslant m$, — оптимальное алфавитное кодирование, соответствующее распределению $\vec{Q}_j$ (напомним, что $\vec{Q}_j$ означает j-ю строку матрицы Q). Эти кодирования могут быть построены, например, при помощи алгоритма Хаффмана. Определим кодирование $\varphi^* \colon \mathscr{A}^* \to \mathscr{B}^*$ последовательностей символов формулой

$$\varphi^*(a_{i_1}, \ldots, a_{i_n}) = \varphi_0(a_{i_1})\varphi_{i_1}(a_{i_2})\varphi_{i_2}(a_{i_3}) \ldots \varphi_{i_{n-1}}(a_{i_n}).$$

Что можно утверждать о величине $\mathsf{E}\ \mathrm{len}(\varphi^*(A^n))$? Будет ли такое кодирование последовательностей оптимальным (при всех $n \geqslant 1$) в классе всех посимвольных кодирований φ^*?

Заинтересованный читатель может сравнить свои ответы на поставленные здесь вопросы с теоремой 3.6 и самостоятельно изучить замечательную теорему А. Я. Хинчина, устанавливающую точную

нижнюю грань коэффициента сжатия сообщений марковского источника в классе всевозможных кодирований (например, по сборнику трудов [32]).

§ 3.4. Задачи

3.1. Выяснить, являются ли однозначно декодируемыми следующие коды:
а) 010, 0100, 0010; б) 0, 01, 011, 111;
в) 0, 01, 001, 0010, 0011; г) 110, 11, 100, 00, 10;
д) 100, 001, 101, 1101, 11011;
е) 010, 0001, 0110, 1100, 00011, 00110, 11110, 10101;
ж) 00, 012, 0110, 0112, 100, 201, 212, 22;
з) 123, 1234, 5, 421, 2135, 3513, 3512, 5124.

3.2. Выяснить, какие из следующих кодов не могут быть получены по алгоритму Хаффмана при любом распределении $\vec{p}$:
а) 0, 10, 11; б) 00, 01, 10, 110; в) 01, 10.

3.3. Построить двоичный код из четырёх кодовых слов, не префиксный, не суффиксный, но однозначно декодируемый.

3.4. Предположим, что двоичный код состоит из более чем 2^n кодовых слов, длина каждого из которых не превосходит n. Может ли такой код быть однозначно декодируемым; префиксным?

3.5. Построить двоичный префиксный код с заданной последовательностью длин кодовых слов:
а) 1, 2, 3, 3; б) 1, 2, 4, 4, 4; в) 2, 2, 3, 4, 4;
г) 2, 2, 2, 4, 4, 4; д) 2, 3, 3, 3, 4, 4;
е) 1, 2, 3, 3, 4, 4, 4, 4; ж) 2, 2, 3, 3, 4, 4, 4, 4.

3.6. Выяснить, может ли заданный набор чисел быть набором длин кодовых слов однозначно декодируемого D-ичного кода:
а) 1, 2, 2, 2, 3, 3, 3, 3 ($D=3$); б) 1, 1, 2, 2, 3, 3, 3 ($D=3$);
в) 1, 1, 1, 2, 2, 2, 2, 3 ($D=4$); г) 1, 2, 2, 2, 2 ($D=3$);
д) 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3 ($D=3$).

3.7. Выяснить, при каком наименьшем D заданный набор чисел может быть набором длин кодовых слов однозначно декодируемого D-ичного кода:
а) 1, 2, 2, 3; б) 2, 2, 2, 4, 4, 4; в) 1, 1, 2, 2, 3, 3.

3.8. При $D=4$ построить или доказать невозможность построения однозначно декодируемого D-ичного кода, в котором длины кодовых слов не превосходят 6, причём количество кодовых слов длины ℓ равно k_ℓ, $1 \leqslant \ell \leqslant 6$, если набор $(k_1, \ldots, k_6)$ равен:
а) (3, 3, 3, 3, 4, 0); б) (2, 7, 3, 3, 5, 0); в) (1, 7, 3, 7, 4, 0);
г) (0, 7, 3, 11, 3, 4).

3.9. Доказать (индукцией по m), что сумма длин кодовых слов оптимального двоичного кода для m-буквенного алфавита не превосходит $\frac{(m+2)(m-1)}{2}$.

3.10. Доказать, что если для D-ичного префиксного кода с длинами кодовых слов ℓ_i, $1 \leqslant i \leqslant m$, справедливо неравенство $\sum_{i=1}^{m} D^{-\ell_i} < 1$, то существует сколь угодно длинная последовательность символов кодового алфавита, которая не может быть представлена как последовательность кодовых слов.

3.11. Построить двоичные коды по алгоритмам Фано и Хаффмана для указанных ниже распределений, сравнить средние длины кодовых слов:

а) (0,4, 0,2, 0,2, 0,2); б) (0,7, 0,1, 0,1, 0,1);

в) $\left(\frac{27}{40}, \frac{9}{40}, \frac{3}{40}, \frac{1}{40}\right)$; г) $\left(\frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{20}\right)$;

д) $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right)$;е) (0,2, 0,2, 0,2, 0,2, 0,2);

ж) $\left(\frac{28}{72}, \frac{15}{72}, \frac{12}{72}, \frac{11}{72}, \frac{6}{72}\right)$; з) (0,5, 0,2, 0,1, 0,09, 0,08, 0,03);

и) (0,5, 0,2, 0,1, 0,09, 0,08, 0,03);

к) (0,4, 0,2, 0,1, 0,1, 0,1, 0,1);

л) (0,4, 0,3, 0,1, 0,07, 0,06, 0,04, 0,03);

м) (0,4, 0,3, 0,08, 0,06, 0,04, 0,04, 0,04, 0,04);

н) (0,32, 0,24, 0,20, 0,09, 0,05, 0,04, 0,04, 0,02).

3.12. Построить коды по алгоритму Хаффмана для указанных ниже распределений при $D \in \{2, 3, 4\}$, вычислить средние длины кодовых слов:

а) $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right)$; б) (0,3, 0,2, 0,2, 0,2, 0,1);

в) $\left(\frac{3}{8}, \frac{1}{6}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{12}\right)$; г) $\left(\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{4}{21}, \frac{5}{21}, \frac{6}{21}\right)$;

д) (0,2, 0,15, 0,05, 0,2, 0,25, 0,15);

е) (0,4, 0,2, 0,1, 0,1, 0,1, 0,1);

ж) (0,3, 0,3, 0,1, 0,1, 0,1, 0,1);

з) (0,3, 0,2, 0,15, 0,15, 0,1, 0,1);

и) (0,24, 0,24, 0,16, 0,16, 0,12, 0,8);

к) (0,3, 0,2, 0,1, 0,1, 0,1, 0,1, 0,1);

л) (0,2, 0,12, 0,08, 0,15, 0,25, 0,1, 0,1);

м) (0,3, 0,25, 0,15, 0,1, 0,1, 0,05, 0,05);

н) (0,49, 0,26, 0,12, 0,04, 0,04, 0,03, 0,02);

о) (0,25, 0,05, 0,1, 0,13, 0,2, 0,12, 0,08, 0,07);
п) (0,21, 0,2, 0,17, 0,16, 0,12, 0,08, 0,04, 0,02);
р) (0,2, 0,15, 0,15, 0,13, 0,12, 0,11, 0,11, 0,03);
с) (0,3, 0,2, 0,15, 0,1, 0,1, 0,08, 0,05, 0,02).

3.13. Для распределения $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{12}\right)$ построить два различных оптимальных двоичных кода с наборами длин кодовых слов (1, 2, 3, 3) и (2, 2, 2, 2).

3.14. Для распределения (0,3, 0,2, 0,2, 0,1, 0,1, 0,05, 0,05) построить три оптимальных двоичных кода с наборами длин кодовых слов (2, 2, 2, 3, 4, 5, 5), (2, 2, 3, 3, 3, 4, 4) и (2, 2, 2, 4, 4, 4, 4).

3.15. Для равномерного распределения на алфавите из m символов найти длины кодовых слов двоичного кода Хаффмана и сравнить среднюю длину этого кода с энтропией, если $m = 100$; в общем случае.

3.16. Построить двоичный код Хаффмана для распределения $\left(\frac{1}{3}, \frac{1}{5}, \frac{1}{5}, \frac{2}{15}, \frac{2}{15}\right)$. Доказать, что этот код является оптимальным и для распределения $\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$.

3.17. Построить пример такого оптимального двоичного префиксного кодирования φ для алфавита $\mathscr{A} = \{a_1, \ldots, a_m, a_{m+1}, \ldots, a_{m+k}\}$ с распределением $(p_1, \ldots, p_m, 0, \ldots, 0)$, что ограничение $\varphi_{\mathscr{B}}$ кодирования φ на алфавит $\mathscr{B} = \{a_1, \ldots, a_m\}$ с распределением $(p_1, \ldots, p_m)$ не является оптимальным.

3.18. Пусть дано распределение (0,3, 0,3, 0,2, 0,1, 0,1), и пусть ℓ_i — длины кодовых слов двоичного кода Хаффмана, соответствующего этому распределению , $i = 1, \ldots, 5$. Найти такое распределение $\vec{q} = (q_1, q_2, q_3, q_4, q_5)$, что $\sum_{i=1}^{5} \ell_i q_i = H(\vec{q})$.

3.19. Для распределения $\left(\frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^{k-2}}, \frac{1}{2^{k-1}}, \frac{1}{2^{k-1}}\right)$ на множестве из k символов построить оптимальное двоичное кодирование φ_k, вычислить его среднюю длину ℓ^{φ_k} и предел $\lim_{k\to\infty} \ell^{\varphi_k}$.

3.20. Вычислить среднюю длину ℓ^{φ_n} оптимального двоичного кодрования φ_n для случайных последовательностей длины n, порождаемых источником без памяти $(\mathscr{A}, \vec{p})$, и сравнить величины $\ell^{(n)} = \frac{\ell^{\varphi_n}}{n}$ и $H(\vec{p})$, если:

$n = 2, 3, 4$, $\mathscr{A} = \{0, 1\}$, $\vec{p} = (0{,}9, 0{,}1)$ или $\vec{p} = (0{,}6, 0{,}4)$;

$n = 2, 3$, $\mathscr{A} = \{0, 1, 2\}$, $\vec{p} = (0{,}7, 0{,}2, 0{,}1)$.

3.21. Построить кодирование φ^*, описанное в замечании 3.15, для простого марковского источника с матрицей $Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

3.22. Пусть $(\ell_1, \ldots, \ell_m)$ — длины кодовых слов двоичного кода Хаффмана, соответствующего распределению $(p_1, \ldots, p_m)$, $p_1 \geqslant p_2 \geqslant \geqslant \ldots \geqslant p_m$. Доказать, что если $p_1 > \frac{2}{5}$, то $\ell_1 = 1$; если $p_1 > \frac{1}{3}$, то $\ell_1 \geqslant 2$.

3.23. Обозначим через $L_D(\vec{p})$ среднюю длину оптимального D-ичного кода, соответствующего распределению $\vec{p}$.

1. Доказать, что $L_D(\vec{p})$ является непрерывной функцией аргумента $\vec{p}$ в симплексе S_{m-1}.
2. Пусть $m \geqslant 2$. Доказать, что для любого $t \in [0, 1]$ найдётся такое распределение $\vec{p} \in S_{m-1}$, для которого $L_D(\vec{p}) - \frac{H(\vec{p})}{\log_2 D} = t$.
3. Для любого $\delta > 0$ построить распределение $\vec{p}$, для которого справедливо неравенство $L_2(\vec{p}) > H(\vec{p}) + 1 - \delta$.

Глава 4
Дискретные каналы связи и их пропускная способность

§ 4.1. Основные понятия

Реальный канал связи представляет собой систему устройств для преобразования сигналов и физическую среду, в которой эти сигналы передаются (хранятся). Сигналы, передаваемые по реальным каналам связи, подвергаются искажениям вследствие различных шумов и помех. Для нас физическая природа сигналов и механизм возникновения искажений будут несущественны, поскольку предметом нашего рассмотрения будет лишь математическая модель канала связи.

Мы будем рассматривать только дискретные каналы, т. е. такие каналы, для которых множества входных и выходных символов дискретны.

Определение 4.1. Говорят, что задан дискретный канал связи, если даны два конечных множества $\mathscr{X} = \{x_1, \ldots, x_q\}$ (входной алфавит) и $\mathscr{Y} = \{y_1, \ldots, y_s\}$ (выходной алфавит) и для любого $n = 1, 2, \ldots$ определены переходные вероятности

$$\pi^{(n)}(y_{j_1}, \ldots, y_{j_n}/x_{i_1}, \ldots, x_{i_n}) \geqslant 0$$

так, что для любых $(x_{i_1}, \ldots, x_{i_n})$ справедливо равенство

$$\sum_{(y_{j_1}, \ldots, y_{j_n})} \pi^{(n)}(y_{j_1}, \ldots, y_{j_n}/x_{i_1}, \ldots, x_{i_n}) = 1.$$

Далее для краткости будем пользоваться обозначениями

$$x^n = (x_{i_1}, \ldots, x_{i_n}) \in \mathscr{X}^n, \qquad y^n = (y_{j_1}, \ldots, y_{j_n}) \in \mathscr{Y}^n.$$

Замечание 4.1. Если задан канал связи, то произвольному распределению $U_{1, \ldots, n}$ на множестве $\mathscr{X}^n$ ставится в соответствие распределение $V_{1, \ldots, n}$ на множестве $\mathscr{Y}^n$, определяемое формулой

$$V_{1, \ldots, n}(y^n) = \sum_{x^n} U_{1, \ldots, n}(x^n)\, \pi^{(n)}(y^n/x^n). \tag{4.1}$$

Если задано некоторое распределение $U_{1,\dots,n}$ на множестве $\mathscr{X}^n$, то на множестве $\mathscr{X}^n \times \mathscr{Y}^n$ определим пару сл. векторов (X^n, Y^n) с совместным распределением

$$\mathsf{P}(X^n = x^n, Y^n = y^n) = U_{1,\dots,n}(x^n)\pi^{(n)}(y^n/x^n). \tag{4.2}$$

Будем называть X^n случайным входным вектором канала, а Y^n — случайным выходным вектором. Из определения 4.1 и формул (4.1), (4.2) следует, что

$$\mathsf{P}(X^n = x^n) = U_{1,\dots,n}(x^n),$$
$$\mathsf{P}(Y^n = y^n) = V_{1,\dots,n}(y^n),$$
$$\mathsf{P}(Y^n = y^n/X^n = x^n) = \pi^{(n)}(y^n/x^n).$$

Таким образом, переходная вероятность $\pi^{(n)}(y^n/x^n)$ есть условная вероятность появления на выходе канала вектора y^n при условии, что на вход канала поступил вектор x^n.

Определение 4.2. Дискретный канал связи называется каналом без памяти, если дана последовательность $\pi_1, \pi_2, \dots$ стохастических матриц размера $q \times s$ и для любых $n = 1, 2, \dots$ справедливо равенство

$$\pi^{(n)}(y^n/x^n) = \pi_1(y_{j_1}/x_{i_1}) \cdot \pi_2(y_{j_2}/x_{i_2}) \cdot \dots \cdot \pi_n(y_{j_n}/x_{i_n}).$$

Здесь $\pi_k(y_j/x_i)$ есть элемент матрицы π_k, находящийся в строке с номером x_i и столбце с номером y_j (мы полагаем, что строки матрицы π_k пронумерованы символами $x_1, \dots, x_q$ входного алфавита, а столбцы — символами $y_1, \dots, y_s$ выходного алфавита).

Замечание 4.2. Для дискретного канала без памяти вероятность перехода на шаге k входного символа x_{i_k} в выходной символ y_{j_k} не зависит от того, каковы были переходы до шага k и после шага k.

Определение 4.3. Дискретный канал без памяти называется стационарным, если $\pi_1 = \pi_2 = \dots$, т. е. переходные вероятности символов не зависят от номера шага.

Далее мы будем рассматривать только дискретные стационарные каналы связи без памяти, называя их просто каналами связи. Таким образом, далее мы будем пользоваться следующим определением.

Определение 4.4. Каналом связи назовём тройку объектов $(\mathscr{X}, \mathscr{Y}, \pi)$, где $\mathscr{X} = \{x_1, \ldots, x_q\}$ — входной алфавит, $\mathscr{Y} = \{y_1, \ldots, y_s\}$ — выходной алфавит и $\pi = (\pi_{ij}) = (\pi(y_j/x_i))$ — стохастическая матрица размера $q \times s$.

Замечание 4.3. В условиях определения 4.4 формула (4.1) принимает вид

$$V_{1,\ldots,n}(y^n) = \sum_{x^n} U_{1,\ldots,n}(x^n)\pi_{i_1 j_1}\ldots\pi_{i_n j_n}.$$

Если $n = 1$, то распределения U_1 и V_1 на множествах $\mathscr{X}$ и $\mathscr{Y}$ соответственно можно записать в виде вероятностных векторов $\vec{u} = (u_1, u_2, \ldots, u_q)$ и $\vec{v} = (v_1, v_2, \ldots, v_s)$, а формулу (4.1) в виде $v_j = \sum_{i=1}^{q} u_i \pi_{ij}$, $1 \leqslant j \leqslant s$, или в матричном виде $\vec{v} = \vec{u} \cdot \pi$.

Канал связи $(\mathscr{X}, \mathscr{Y}, \pi)$ удобно изображать графически в виде следующей диаграммы: элементы входного алфавита $x_i \in \mathscr{X}$ и элементы выходного алфавита $y_j \in \mathscr{Y}$ представляем в виде двух вертикальных рядов точек, и если $\pi_{ij} > 0$, то пару точек x_i и y_j соединяем стрелкой с пометкой π_{ij}.

Пример 4.1. *Двоичный симметричный канал* с параметром p, или сокращенно ДСК(p), задаётся входным алфавитом $\mathscr{X} = \{0, 1\}$, выходным алфавитом $\mathscr{Y} = \{0, 1\}$ и матрицей $\pi = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$, где $0 \leqslant p \leqslant 1$. При передаче символа по ДСК может происходить правильная передача или искажение (замена на противоположный символ), а параметр p представляет собой вероятность искажения символа при передаче. ДСК(p) представляется следующей диаграммой.

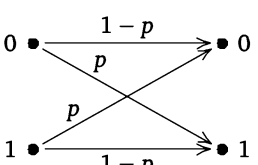

Рис. 4.1. Двоичный симметричный канал с параметром p

Пример 4.2. *Двоичный канал со стиранием* с параметром p задаётся входным алфавитом $\mathscr{X} = \{0, 1\}$, выходным алфавитом $\mathscr{Y} = \{0, E, 1\}$

и матрицей $\pi = \begin{pmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{pmatrix}$, где $0 \leqslant p \leqslant 1$. При передаче символа по данному каналу может происходить правильная передача или искажение (стирание), выходной символ E обозначает результат стирания входного символа, а параметр p представляет собой вероятность стирания. Канал со стиранием представляется следующей диаграммой.

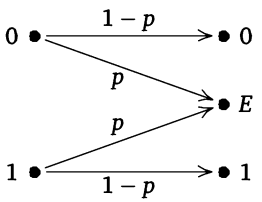

Рис. 4.2. Двоичный канал со стиранием с параметром p

Пусть дан канал связи $(\mathscr{X}, \mathscr{Y}, \pi)$, и пусть для некоторого $n \geqslant 1$ задано распределение $U_{1,\ldots,n}$ на множестве $\mathscr{X}^n$. Для пары сл. векторов X^n, Y^n определены следующие величины: энтропия сл. вектора X^n

$$H(X^n) = -\sum_{x^n} U_{1,\ldots,n}(x^n) \log_2 U_{1,\ldots,n}(x^n),$$

энтропия сл. вектора Y^n

$$H(Y^n) = -\sum_{y^n} V_{1,\ldots,n}(y^n) \log_2 V_{1,\ldots,n}(y^n),$$

совместная энтропия сл. векторов X^n, Y^n

$$H(X^n, Y^n) = -\sum_{x^n,\, y^n} \mathsf{P}(X^n = x^n, Y^n = y^n) \log_2 \mathsf{P}(X^n = x^n, Y^n = y^n),$$

условная энтропия сл. вектора Y^n при условии сл. вектора X^n

$$H(Y^n/X^n) = -\sum_{x^n,\, y^n} \mathsf{P}(X^n = x^n, Y^n = y^n) \log_2 \pi^{(n)}(y^n/x^n) =$$

$$= -\sum_{x^n,\, y^n} \mathsf{P}(X^n = x^n, Y^n = y^n) \sum_{k=1}^{n} \log_2 \pi(y_{j_k}/x_{i_k}),$$

средняя взаимная информация между сл. векторами X^n и Y^n

$$I(X^n; Y^n) = H(Y^n) - H(Y^n/X^n).$$

Определение 4.5. Пропускной способностью канала связи называется величина

$$C^* = \sup_{n \geqslant 1} \Big(\frac{1}{n} \sup_{U_{1,\dots,n}} I(X^n; Y^n) \Big),$$

где внутренний супремум берётся по всем возможным распределениям $U_{1,\dots,n}$ сл. вектора X^n.

Замечание 4.4. 1. Пропускная способность C^* есть точная верхняя грань (по всем длинам $n = 1, 2, \dots$ векторов на входе и выходе канала связи и по всем распределениям $U_{1,\dots,n}$ на множестве $\mathscr{X}^n$ входных векторов) средней взаимной информации между входом и выходом, приходящейся на один переданный символ. Можно сказать, что пропускная способность определяет наибольшее допустимое количество полезной информации, которая может быть передана по каналу связи от отправителя получателю.

2. Из приведённых выше формул для средней взаимной информации $I(X^n; Y^n)$ видно, что эту величину можно рассматривать как действительную функцию, аргументом которой является распределение $U_{1,\dots,n}$ на множестве $\mathscr{X}^n$. Поскольку эта функция непрерывна, а класс всех распределений $U_{1,\dots,n}$ на $\mathscr{X}^n$ представляет собой замкнутое и ограниченное множество в евклидовом пространстве $\mathbb{R}^{q^n}$, для каждого $n \geqslant 1$ по теореме Вейерштрасса точная верхняя грань $\sup_{U_{1,\dots,n}} I(X^n; Y^n)$ *достигается* на некотором распределении $U_{1,\dots,n}$. Такое распределение $U_{1,\dots,n}$ (при заданном n) называют оптимальным входным распределением и говорят, что при данном входном распределении канал связи используется наилучшим образом, так что в сообщениях длины n по этому каналу передаётся максимально возможное количество информации.

Следующая теорема показывает, что в определении 4.5 достаточно рассмотреть случай, когда $n = 1$.

Теорема 4.1. *Для стационарного дискретного канала связи без памяти справедливо равенство*

$$C^* = \max_{\vec{u}} I(X; Y),$$

где максимум берётся по всем распределениям $\vec{u}$ на множестве $\mathscr{X}$.

Доказательство. Для любого распределения $U_{1,\dots,n}$ на $\mathscr{X}^n$ условную энтропию $H(Y^n/X^n)$ можно представить в виде

$$H(Y^n/X^n) = -\sum_{x^n,\,y^n} \mathsf{P}(X^n = x^n,\, Y^n = y^n)\log_2 \pi^{(n)}(y^n/x^n) =$$

$$= -\sum_{k=1}^{n}\sum_{x^n,\,y^n} U_{1,\dots,n}(x^n)\pi^{(n)}(y^n/x^n)\log_2 \pi(y_{j_k}/x_{i_k}) =$$

$$= -\sum_{k=1}^{n}\sum_{x_{i_k}\in\mathscr{X}}\sum_{y_{j_k}\in\mathscr{Y}} \mathsf{P}^{(k)}(x_{i_k})\pi(y_{j_k}/x_{i_k})\log_2 \pi(y_{j_k}/x_{i_k}),$$

где введено новое обозначение $\mathsf{P}^{(k)}$ для распределения k-го символа X_k сл. вектора $X^n = (X_1, \dots, X_n)$, т. е.

$$\mathsf{P}^{(k)}(x_{i_k}) = \mathsf{P}(X_k = x_{i_k}) = \sum_{x_{i_1},\dots,x_{i_{k-1}},x_{i_{k+1}},\dots,x_{i_n}} U_{1,\dots,n}(x^n).$$

Условная энтропия $H(Y_k/X_k)$ k-го выходного сл. символа Y_k при условии X_k равна

$$H(Y_k/X_k) = -\sum_{x_{i_k}\in\mathscr{X}}\sum_{y_{j_k}\in\mathscr{Y}} \mathsf{P}^{(k)}(x_{i_k})\pi(y_{j_k}/x_{i_k})\log_2 \pi(y_{j_k}/x_{i_k}),$$

поэтому условная энтропия $H(Y^n/X^n)$ представляется в виде

$$H(Y^n/X^n) = \sum_{k=1}^{n} H(Y_k/X_k).$$

По свойству полуаддитивности энтропии (следствие 1.1 на с. 28)

$$H(Y^n) \leqslant \sum_{k=1}^{n} H(Y_k),$$

следовательно,

$$\boldsymbol{I}(X^n; Y^n) \leqslant \sum_{k=1}^{n} \boldsymbol{I}(X_k; Y_k). \tag{4.3}$$

Если распределение $U_{1,\dots,n}$ таково, что координаты сл. вектора $X^n = (X_1, \dots, X_n)$ независимы, то координаты сл. вектора $Y^n = (Y_1, \dots, Y_n)$ также будут независимы. Действительно, если

$$U_{1,\dots,n}(x^n) = \prod_{k=1}^{n} \mathsf{P}^{(k)}(x_{i_k}),$$

то

$$
\begin{aligned}
V_{1,\dots,n}(y^n) &= \mathsf{P}(Y^n = y^n) = \\
&= \sum_{x^n} U_{1,\dots,n}(x^n)\pi^{(n)}(y^n/x^n) = \\
&= \sum_{x^n} \prod_{k=1}^{n} \mathsf{P}^{(k)}(x_{i_k})\pi(y_{j_k}/x_{i_k}) = \\
&= \prod_{k=1}^{n} \sum_{x_{i_k}\in\mathscr{X}} \mathsf{P}^{(k)}(x_{i_k})\pi(y_{j_k}/x_{i_k}) = \\
&= \prod_{k=1}^{n} \mathsf{P}(Y_k = y_{j_k}),
\end{aligned}
$$

и при этом в неравенстве (4.3) достигается равенство. Следовательно,

$$
I(X^n; Y^n) \leqslant \sum_{k=1}^{n} \max_{\mathsf{P}^{(k)}} I(X_k; Y_k) = n \max_{\vec{u}} I(X; Y),
$$

и при этом в неравенстве

$$
I(X^n; Y^n) \leqslant \sum_{k=1}^{n} \max_{\mathsf{P}^{(k)}} I(X_k; Y_k)
$$

имеет место равенство, если координаты случайного входного вектора $X^n = (X_1, \dots, X_n)$ независимы и одинаково распределены. Отсюда ясно, что

$$
\frac{1}{n} \sup_{U_{1,\dots,n}} I(X^n; Y^n) = \max_{\vec{u}} I(X; Y). \qquad \square
$$

Утверждение 4.1. *Пусть заданы канал связи* $(\mathscr{X}, \mathscr{Y}, \pi)$ *и входное распределение* $\vec{u}$. *Тогда средняя взаимная информация* $I(X; Y)$ *представляется в виде*

$$
I(X; Y) = H\Big(\sum_{i=1}^{q} u_i \vec{\pi}_i\Big) - \sum_{i=1}^{q} u_i H(\vec{\pi}_i), \tag{4.4}
$$

где через $\vec{\pi}_i$ *обозначена* i*-я строка матрицы* π.

Доказательство. Строка $\vec{\pi}_i$ матрицы π задаёт условное распределение выходного символа Y при условии, что дано значение входного

символа $X = x_i$. Поэтому условная энтропия $H(Y/X)$ представляется в виде

$$H(Y/X) = \sum_{i=1}^{q} u_i H(Y/X = x_i) = \sum_{i=1}^{q} u_i H(\vec{\pi}_i).$$

Согласно замечанию 4.3, распределение $\vec{v}$ выходного символа Y представляется в виде

$$\vec{v} = \vec{u} \cdot \pi = \sum_{i=1}^{q} u_i \vec{\pi}_i.$$

Подставляя полученные выражения в формулу

$$I(X; Y) = H(Y) - H(Y/X),$$

завершаем доказательство. □

В общем случае вычисление пропускной способности C^* для произвольного канала связи $(\mathscr{X}, \mathscr{Y}, \pi)$ представляет собой сложную задачу.

Эту задачу можно сформулировать как задачу поиска условного максимума неотрицательной функции $I(X; Y)$, зависящей от q действительных переменных $u_1, \ldots, u_q$, при условии совокупности ограничений, состоящей из одного равенства $u_1 + \ldots + u_q = 1$ и q неравенств $u_i \geqslant 0$, $i = 1, \ldots, q$. Применение известного метода множителей Лагранжа здесь затрудняется тем, что некоторые ограничения представляют собой неравенства. Кроме того, максимум может достигаться не во внутренней точке симплекса S_{q-1}, а на границе этого множества.

Приведём без доказательства теорему о необходимых и достаточных условиях, при которых набор $\vec{u} = (u_1, \ldots, u_q)$ обращает в максимум функцию $I(X; Y)$. Нам понадобится обозначение

$$I(x_i; Y) = \sum_{j=1}^{s} \pi_{ij} \log \frac{\pi_{ij}}{\sum_{t=1}^{q} u_t \pi_{tj}}, \quad x_i \in \mathscr{X}.$$

Теорема 4.2. *Пусть задан канал связи $(\mathscr{X}, \mathscr{Y}, \pi)$. Распределение $\vec{u} = (u_1, \ldots, u_q)$ обращает в максимум величину $I(X; Y)$ в том и только том случае, когда существует такое число C, что*

$$\text{если } u_i > 0, \text{ то } I(x_i; Y) = C,$$
$$\text{если } u_i = 0, \text{ то } I(x_i; Y) \leqslant C.$$

При этом пропускная способность C^* равна числу C.

Если дано входное распределение $\vec{u} = (u_1, \ldots, u_q)$ и предполагается, что оно является оптимальным входным распределением для рассматриваемого канала связи, то условия теоремы 4.2 позволяют легко проверить это предположение и тем самым найти C^*. Если же распределение $\vec{u} = (u_1, \ldots, u_q)$ неизвестно, то условия теоремы дают систему уравнений относительно распределения $\vec{u}$. Эта система уравнений является трансцендентной, и в общем случае неизвестен способ её аналитического решения. В некоторых случаях система упрощается, и её удаётся решить аналитически. В других случаях для нахождения пропускной способности C^* и оптимального входного распределения $\vec{u}$ применяют численные итерационные алгоритмы.

§ 4.2. Симметричные каналы

Рассмотрим некоторые частные случаи, когда вычисление пропускной способности C^* для канала связи $(\mathscr{X}, \mathscr{Y}, \pi)$ существенно упрощается.

Определение 4.6. Канал связи $(\mathscr{X}, \mathscr{Y}, \pi)$ называется

1) *симметричным по входу*, если все строки матрицы $\pi_{q\times s}$ являются перестановками одного и того же набора чисел $\pi'_1, \ldots, \pi'_s$;
2) *симметричным по выходу*, если все столбцы матрицы $\pi_{q\times s}$ являются перестановками одного и того же набора чисел $\pi''_1, \ldots, \pi''_q$;
3) просто *симметричным*, если он симметричен по входу и по выходу.

Утверждение 4.2. *Если канал связи симметричен по входу, то*

$$C^* \leqslant \log_2 s + \sum_{i=1}^{s} \pi'_i \log_2 \pi'_i.$$

Доказательство. Так как $|\mathscr{Y}| = s$, верно неравенство $H(Y) \leqslant \log_2 s$, причём равенство достигается только для равномерного выходного распределения $\vec{v} = (1/s, \ldots, 1/s)$.

Так как канал связи симметричен по входу,

$$H(Y/X) = -\sum_{i=1}^{q}\sum_{j=1}^{s} u_i \pi_{ij} \log_2 \pi_{ij} = -\sum_{i=1}^{q} u_i \sum_{j=1}^{s} \pi_{ij} \log_2 \pi_{ij} =$$
$$= -\sum_{i=1}^{q} u_i \sum_{j=1}^{s} \pi'_j \log_2 \pi'_j = -\sum_{j=1}^{s} \pi'_j \log_2 \pi'_j \underbrace{\sum_{i=1}^{q} u_i}_{=1}.$$

Применяя утверждение 4.1 и теорему 4.1, завершаем доказательство. □

Утверждение 4.3. *Если канал связи симметричен по выходу, то равномерному распределению* $\vec{u} = (1/q, \ldots, 1/q)$ *на входе соответствует равномерное распределение* $\vec{v} = (1/s, \ldots, 1/s)$ *на выходе.*

Доказательство. Для $1 \leqslant j \leqslant s$

$$v_j = \sum_{i=1}^{q} u_i \pi_{ij} = \frac{1}{q} \sum_{i=1}^{q} \pi_{ij} = \frac{1}{q} \sum_{i=1}^{q} \pi''_i,$$

т. е. v_j не зависит от j. Следовательно, $v_j = 1/s$. □

Следствие 4.1. *Если канал связи симметричен, то*

$$C^* = \log_2 s + \sum_{i=1}^{s} \pi'_i \log_2 \pi'_i.$$

Доказательство вытекает из утверждений 4.2 и 4.3. □

Пример 4.3 (продолжение к примеру 4.1). Для ДСК с параметром p имеем $q = s = 2$, и в силу следствия 4.1

$$C^* = \log_2 s + \sum_{i=1}^{s} \pi'_i \log_2 \pi'_i =$$
$$= \log_2 2 + p \log_2 p + (1-p) \log_2 (1-p) =$$
$$= 1 - h(p),$$

где $h(p) = -p \log_2 p - (1-p) \log_2 (1-p)$. График зависимости C^* от p представлен на следующем рисунке.

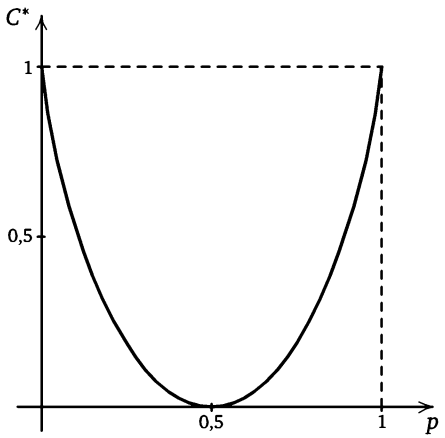

Рис. 4.3. Зависимость пропускной способности ДСК от p

График показывает, что полученный результат вполне согласуется с интуицией. Действительно, если $p = 0$, то искажения в ДСК не происходят и каждый переданный символ безошибочно воспроизводится на выходе канала, при этом пропускная способность равна $C^* = 1$. Если $p = 1$, то каждый переданный символ наверняка искажается, и поэтому по выходу можно однозначно восстановить вход; в этом случае также $C^* = 1$. Если же $p = 1/2$, то любому переданному символу будет соответствовать случайный выходной символ с равномерным распределением, статистически не зависящий от переданного символа; разумеется, такой канал не передаёт никакой полезной информации, и его пропускная способность равна нулю.

Пример 4.4 (продолжение к примеру 4.2). Для двоичного канала со стиранием с параметром p имеем $q = 2$, $s = 3$. Этот канал симметричен по входу, но не симметричен по выходу. Пусть задано некоторое входное распределение $\vec{u} = (\alpha, \beta)$. Повторяя рассуждения из доказательства утверждения 4.2, получим, что энтропия $H(Y/X) = H(1-p, 0, p) = h(p)$ не зависит от $\vec{u}$. Далее, выходное распределение равно $\vec{v} = \vec{u} \cdot \pi = ((1-p)\alpha, p, (1-p)\beta)$. Энтропия $H(\vec{v})$ достигает своего максимального значения при $\alpha = \beta = 1/2$, и это максимальное значение равно $H((1-p)/2, p, (1-p)/2)$. Та-

ким образом,

$$C^* = H\left(\frac{1-p}{2}, p, \frac{1-p}{2}\right) - h(p) =$$
$$= -2 \cdot \frac{1-p}{2} \log_2 \frac{1-p}{2} - p \log_2 p + p \log_2 p + (1-p) \log_2 (1-p) =$$
$$= 1 - p.$$

И в этом примере полученный результат вполне согласуется с интуицией. Если $p = 0$, то стирания никогда не происходят, каждый переданный символ безошибочно воспроизводится на выходе канала, и пропускная способность равна $C^* = 1$. Если $p = 1$, то каждый переданный символ наверняка стирается и по выходу нельзя получить никакой информации о входе; при этом $C^* = 0$. Однако до вычислений трудно было предположить, что пропускная способность линейно зависит от вероятности стирания.

Пример 4.5 (обобщение примеров 4.3 и 4.4). *Канал с искажением и стиранием* и двумя параметрами ε, δ задаётся входным алфавитом $\mathscr{X} = \{0, 1\}$, выходным алфавитом $\mathscr{Y} = \{0, E, 1\}$ и матрицей

$$\pi = \begin{pmatrix} 1-\delta-\varepsilon & \varepsilon & \delta \\ \delta & \varepsilon & 1-\delta-\varepsilon \end{pmatrix}.$$

Параметры канала удовлетворяют неравенствам $\varepsilon \geqslant 0$, $\delta \geqslant 0$, $\delta + \varepsilon \leqslant \leqslant 1$. При передаче символа по данному каналу может происходить правильная передача, искажение (на входе 0, а на выходе 1, или наоборот) или стирание (на выходе символ стирания E). Параметр δ представляет собой вероятность искажения переданного символа, а параметр ε — вероятность стирания. Данный канал представляется следующей диаграммой.

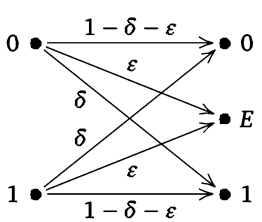

Рис. 4.4. Двоичный канал с искажением и стиранием

Входному распределению $\vec{u} = (\alpha, \beta)$ соответствует выходное распределение

$$\vec{v} = \vec{u} \cdot \pi = ((1-\delta-\varepsilon)\alpha + \delta\beta, \varepsilon, (1-\delta-\varepsilon)\alpha + \delta\beta).$$

Энтропия $H(\vec{v})$ достигает своего максимума при $\alpha = \beta = 1/2$, и этот максимум равен

$$H\left(\frac{1-\varepsilon}{2}, \varepsilon, \frac{1-\varepsilon}{2}\right).$$

Как и в примере 4.4, получим, что энтропия

$$H(Y/X) = H(1-\delta-\varepsilon, \varepsilon, \delta)$$

не зависит от $\vec{p}$. Таким образом,

$$\begin{aligned} C^*(\varepsilon, \delta) &= H\left(\frac{1-\varepsilon}{2}, \varepsilon, \frac{1-\varepsilon}{2}\right) - H(1-\delta-\varepsilon, \varepsilon, \delta) = \\ &= -2 \cdot \frac{1-\varepsilon}{2} \log_2 \frac{1-\varepsilon}{2} + (1-\delta-\varepsilon)\log_2(1-\delta-\varepsilon) + \delta \log_2 \delta = \\ &= (1-\varepsilon)(1-\log_2(1-\varepsilon)) + (1-\delta-\varepsilon)\log_2(1-\delta-\varepsilon) + \delta \log_2 \delta = \\ &= (1-\varepsilon)\left(1 - h\left(\frac{\delta}{1-\varepsilon}\right)\right). \end{aligned}$$

Рассмотрим некоторые частные случаи полученного результата. Если $\varepsilon = 0$, то стирания никогда не происходят и мы приходим к случаю ДСК с параметром δ. Подстановка $\varepsilon = 0$ приводит к выражению

$$C^*(0, \delta) = 1 - h(\delta),$$

что совпадает с результатом в примере 4.3.

Если $\delta = 0$, то искажения никогда не происходят и мы приходим к случаю канала со стиранием с параметром ε. Подстановка $\delta = 0$ приводит к выражению $C^*(\varepsilon, 0) = 1 - \varepsilon$, что совпадает с результатом в примере 4.4. Кроме того, на отрезке $0 \leqslant \varepsilon \leqslant 1$, $\varepsilon + \delta = 1$, получаем $C^* = 1 - \varepsilon$, а на отрезке $0 \leqslant \varepsilon \leqslant 1$, $\varepsilon + 2\delta = 1$, получаем $C^* = 0$.

График зависимости C^* от параметров ε, δ представлен на следующем рисунке.

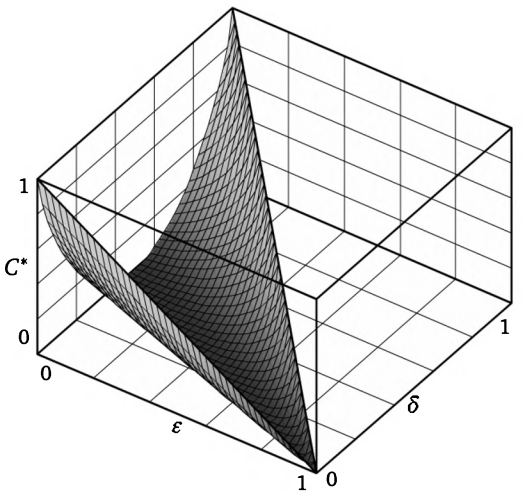

Рис. 4.5. Пропускная способность канала с искажением и стиранием

§ 4.3. Соединение каналов

Если даны два или более каналов связи, то их можно комбинировать различными способами, получая каналы связи с новыми свойствами.

Определение 4.7. Пусть даны два канала связи без памяти $(\mathscr{X}_1, \mathscr{Y}_1, \pi_1)$ и $(\mathscr{X}_2, \mathscr{Y}_2, \pi_2)$, $q_i = |\mathscr{X}_i|$, $s_i = |\mathscr{Y}_i|$, $i = 1, 2$. Тогда канал связи $(\mathscr{X}, \mathscr{Y}, \pi)$ называется

1) *последовательным соединением* данных каналов, если $\mathscr{Y}_1 = \mathscr{X}_2$ и $\pi = \pi_1 \cdot \pi_2$;
2) *параллельным соединением* данных каналов, если $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2$, $\mathscr{Y} = \mathscr{Y}_1 \times \mathscr{Y}_2$ и переходные вероятности имеют вид

$$\pi((y_1, y_2)/(x_1, x_2)) = \pi_1(y_1/x_1) \cdot \pi_2(y_2/x_2);$$

3) *суммой* данных каналов, если $\mathscr{X} = \mathscr{X}_1 \cup \mathscr{X}_2$, $\mathscr{X}_1 \cap \mathscr{X}_2 = \varnothing$, $\mathscr{Y} = \mathscr{Y}_1 \cup \mathscr{Y}_2$, $\mathscr{Y}_1 \cap \mathscr{Y}_2 = \varnothing$ и матрица π имеет блочный вид: $\pi = \begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{pmatrix}$.

Последовательное соединение каналов означает, что выходной символ первого канала подаётся на вход второго канала. При этом равенство $\pi = \pi_1 \cdot \pi_2$ представляет собой просто другую форму записи формулы полной вероятности. Действительно, если X_1 — сл. символ на входе первого канала, $Y_1 = X_2$ — сл. символ на выходе первого (и на входе второго) канала, Y_2 — сл. символ на выходе второго канала, то для $x_i \in \mathscr{X}_1$, $y_j \in \mathscr{Y}_1$ и $z_k \in \mathscr{Y}_2$ совместное распределение X_1, Y_1, Y_2 имеет вид

$$\mathsf{P}(X_1 = x_i, Y_1 = y_j, Y_2 = z_k) = \mathsf{P}(X_1 = x_i)\pi_1(y_j/x_i)\pi_2(z_k/y_j), \qquad (4.5)$$

и при этом

$$\mathsf{P}(Y_2 = z_k/X_1 = x_i) = \sum_{j=1}^{s_1} \mathsf{P}(Y_2 = z_k/X_2 = y_j)\mathsf{P}(Y_1 = y_j/X_1 = x_i) =$$

$$= \sum_{j=1}^{s_1} \pi_1(y_j/x_i)\pi_2(z_k/y_j).$$

Параллельное соединение каналов означает, что на вход канала поступает пара сл. символов (X_1, X_2), на выходе получается пара сл. символов (Y_1, Y_2), и при этом передача $X_1 \to Y_1$ и $X_2 \to Y_2$ происходит независимо, в соответствии с переходными вероятностями соответствующих каналов.

Сумма каналов означает, что на вход нового канала подаётся символ либо из алфавита $\mathscr{X}_1$, либо из алфавита $\mathscr{X}_2$, и затем этот символ передаётся по одному из двух каналов.

В следующих трёх теоремах мы будем предполагать, что канал связи $(\mathscr{X}, \mathscr{Y}, \pi)$ является последовательным или параллельным соединением или суммой двух каналов связи $(\mathscr{X}_1, \mathscr{Y}_1, \pi_1)$ и $(\mathscr{X}_2, \mathscr{Y}_2, \pi_2)$ и что пропускные способности этих каналов равны соответственно C^*, C_1^* и C_2^*.

Теорема 4.3. *Если канал связи $(\mathscr{X}, \mathscr{Y}, \pi)$ является последовательным соединением двух каналов $(\mathscr{X}_1, \mathscr{Y}_1, \pi_1)$ и $(\mathscr{X}_2, \mathscr{Y}_2, \pi_2)$, то справедливо неравенство*

$$C^* \leqslant \min\{C_1^*, C_2^*\}.$$

Доказательство. Дважды применяя утверждение 1.9 (с. 32), можем записать

$$I(X_1; (Y_1, Y_2)) = I(X_1; Y_2) + I(X_1; Y_1/Y_2) = I(X_1; Y_1) + I(X_1; Y_2/Y_1). \qquad (4.6)$$

Согласно (4.5),

$$\mathsf{P}(Y_2 = z_k / X_1 = x_i,\, Y_1 = y_j) = \frac{\mathsf{P}(X_1 = x_i)\pi_1(y_j/x_i)\pi_2(z_k/y_j)}{\mathsf{P}(X_1 = x_i)\pi_1(y_j/x_i)} = \\ = \pi_2(z_k/y_j) = \\ = \mathsf{P}(Y_2 = z_k / Y_1 = y_j),$$

т. е. Y_2 не зависит от X_1 при условии Y_1. Из утверждения 1.11 (с. 33) получаем

$$I(X_1;\, Y_2/Y_1) = 0. \tag{4.7}$$

Далее, в силу утверждения 1.10 (с. 32) имеем

$$I(X_1;\, Y_1/Y_2) \geqslant 0. \tag{4.8}$$

Подставляя (4.7) и (4.8) в (4.6), получим неравенство

$$I(X_1;\, Y_2) \leqslant I(X_1;\, Y_1). \tag{4.9}$$

Поменяв в соотношении (4.6) ролями величины X_1 и Y_2, получим

$$I(Y_2;\, (X_1, Y_1)) = I(X_1;\, Y_2) + I(Y_1;\, Y_2/X_1) = I(Y_1;\, Y_2) + I(X_1;\, Y_2/Y_1).$$

Принимая во внимание (4.7) и неравенство $I(Y_1;\, Y_2/X_1) \geqslant 0$, получим

$$I(X_1;\, Y_2) \leqslant I(Y_1;\, Y_2). \tag{4.10}$$

Из (4.9) и (4.10) следует, что

$$I(X_1;\, Y_2) \leqslant \min\{I(X_1;\, Y_1),\, I(Y_1;\, Y_2)\} \leqslant \min\{C_1^*,\, C_2^*\},$$

что и требовалось доказать. □

Теорема 4.4. *Если канал связи* $(\mathscr{X}, \mathscr{Y}, \pi)$ *является параллельным соединением двух каналов* $(\mathscr{X}_1, \mathscr{Y}_1, \pi_1)$ *и* $(\mathscr{X}_2, \mathscr{Y}_2, \pi_2)$, *то справедливо равенство*

$$C^* = C_1^* + C_2^*.$$

Доказательство этой теоремы по существу не отличается от доказательства теоремы 4.1. Пусть дано некоторое распределение

$$U_{1,2}(x_1, x_2) = \mathsf{P}(X_1 = x_1,\, X_2 = x_2)$$

входной пары сл. символов (X_1, X_2), и тем самым заданы частные распределения

$$U_1(X_1 = x_1) = \sum_{x_2} \mathsf{P}(X_1 = x_1, X_2 = x_2),$$
$$U_2(X_2 = x_2) = \sum_{x_1} \mathsf{P}(X_1 = x_1, X_2 = x_2).$$

Тогда так же, как в доказательстве теоремы 4.1, получим соотношения

$$H(Y_1, Y_2/X_1, X_2) = H(Y_1/X_1) + H(Y_2/X_2),$$
$$H(Y_1, Y_2) \leqslant H(Y_1) + H(Y_2),$$

причём в последнем неравенстве достигается равенство, если сл. символы X_1 и X_2 независимы. Следовательно,

$$I((Y_1, Y_2); (X_1, X_2)) \leqslant I(Y_1; X_1) + I(Y_2; X_2) \leqslant C_1^* + C_2^*,$$

и равенства достигаются, если сл. символы X_1 и X_2 независимы, а их распределения обращают в максимум величины $I(Y_1; X_1)$ и $I(Y_2; X_2)$, соответственно. □

Теорема 4.5. *Если канал связи $(\mathscr{X}, \mathscr{Y}, \pi)$ является суммой двух каналов $(\mathscr{X}_1, \mathscr{Y}_1, \pi_1)$ и $(\mathscr{X}_2, \mathscr{Y}_2, \pi_2)$, то справедливо равенство*

$$2^{C^*} = 2^{C_1^*} + 2^{C_2^*}.$$

Доказательство. Пусть на вход канала связи $(\mathscr{X}, \mathscr{Y}, \pi)$ поступает сл. символ X с распределением $\vec{u} = (u_1, \ldots, u_q)$, а на выходе появляется сл. символ Y с распределением $\vec{v} = (v_1, \ldots, v_s)$.

По определению 4.7

$$q = |\mathscr{X}| = |\mathscr{X}_1| + |\mathscr{X}_2| = q_1 + q_2,$$
$$s = |\mathscr{Y}| = |\mathscr{Y}_1| + |\mathscr{Y}_2| = s_1 + s_2.$$

Представим вектор $\vec{u}$ в виде $\vec{u} = (\vec{u}', \vec{u}'')$, где

$$\vec{u}' = (u_1, \ldots, u_{q_1}), \quad \vec{u}'' = (u_{q_1+1}, \ldots, u_q),$$

и аналогично вектор $\vec{v}$ в виде $\vec{v} = (\vec{v}', \vec{v}'')$, где

$$\vec{v}' = (v_1, \ldots, v_{s_1}), \quad \vec{v}'' = (v_{s_1+1}, \ldots, v_s).$$

Из формулы $\vec{\nu} = \vec{u} \cdot \pi$ и вида матрицы $\pi = \begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{pmatrix}$ следует, что $\vec{\nu}' = \vec{u}' \cdot \pi_1$ и $\vec{\nu}'' = \vec{u}'' \cdot \pi_2$.

Обозначим через α сумму координат вектора $\vec{u}'$. Ясно, что сумма координат вектора $\vec{\nu}'$ также равна α. Введём вероятностные векторы $\vec{a}'$, $\vec{a}''$, $\vec{b}'$, $\vec{b}''$ следующим образом:

$$\vec{a}' = \begin{cases} \frac{1}{\alpha} \cdot \vec{u}', & \alpha > 0, \\ \left(\frac{1}{q_1}, \ldots, \frac{1}{q_1}\right), & \alpha = 0, \end{cases} \qquad \vec{a}'' = \begin{cases} \frac{1}{1-\alpha} \cdot \vec{u}'', & \alpha < 1, \\ \left(\frac{1}{q_2}, \ldots, \frac{1}{q_2}\right), & \alpha = 1, \end{cases}$$

получаем $\vec{b}' = \vec{a}' \cdot \pi_1$, $\vec{b}'' = \vec{a}'' \cdot \pi_2$.

В новых обозначениях

$$\vec{u} = (\alpha\vec{a}', (1-\alpha)\vec{a}''), \qquad \vec{\nu} = (\alpha\vec{b}', (1-\alpha)\vec{b}'').$$

Пусть X_1 и Y_1 — сл. символы на входе и выходе канала $(\mathscr{X}_1, \mathscr{Y}_1, \pi_1)$, X_2 и Y_2 — сл. символы на входе и выходе канала $(\mathscr{X}_2, \mathscr{Y}_2, \pi_2)$, и при этом X_1 и X_2, имеют распределения $\vec{a}'$, $\vec{a}''$ соответственно. Тогда Y_1 и Y_2 имеют распределения $\vec{b}'$, $\vec{b}''$ соответственно.

Вычислим среднюю взаимную информацию

$$I(X;Y) = H(Y) - H(Y/X).$$

Имеем:

$$\begin{aligned} H(Y) &= -\sum_{j=1}^{s} \nu_j \log_2 \nu_j = \\ &= -\sum_{j=1}^{s_1} \nu_j \log_2 \nu_j - \sum_{j=s_1+1}^{s} \nu_j \log_2 \nu_j = \\ &= -\sum_{j=1}^{s_1} \alpha b'_j \log_2(\alpha b'_j) - \sum_{j=1}^{s_2} (1-\alpha) b''_j \log_2\left((1-\alpha) b''_j\right) = \\ &= -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha) - \\ &\quad - \alpha \sum_{j=1}^{s_1} b'_j \log_2 b'_j - (1-\alpha) \sum_{j=1}^{s_2} b''_j \log_2 b''_j = \\ &= h(\alpha) + \alpha H(Y_1) + (1-\alpha) H(Y_2). \end{aligned}$$

Далее,

$$H(Y/X) = \sum_{i=1}^{q} u_i H(\vec{\pi}_i) =$$
$$= \sum_{i=1}^{q_1} u_i H(\vec{\pi}'_i, \underbrace{0, \ldots, 0}_{s_2}) + \sum_{i=1}^{q_2} u_{q_1+i} H(\underbrace{0, \ldots, 0}_{s_1}, \vec{\pi}''_i) =$$
$$= \alpha \sum_{i=1}^{q_1} a'_i H(\vec{\pi}'_i) + (1-\alpha) \sum_{i=1}^{q_2} a''_i H(\vec{\pi}''_i) =$$
$$= \alpha H(Y_1/X_1) + (1-\alpha) H(Y_2/X_2).$$

Таким образом,

$$I(X;Y) = H(Y) - H(Y/X) =$$
$$= h(\alpha) + \alpha H(Y_1) + (1-\alpha) H(Y_2) -$$
$$- \alpha H(Y_1/X_1) - (1-\alpha) H(Y_2/X_2) =$$
$$= h(\alpha) + \alpha I(X_1;Y_1) + (1-\alpha) I(X_2;Y_2) \leqslant$$
$$\leqslant h(\alpha) + \alpha C_1^* + (1-\alpha) C_2^*.$$

В последнем неравенстве достигается равенство, когда при фиксированном α в векторе $\vec{u} = (\alpha\vec{a}', (1-\alpha)\vec{a}'')$ взяты такие $\vec{a}'$ и $\vec{a}''$, для которых достигаются максимальные значения

$$\max I(X_1;Y_1) = C_1^*, \qquad \max I(X_2;Y_2) = C_2^*.$$

Таким образом, вычисление пропускной способности суммы двух каналов связи сводится к нахождению максимума функции

$$f(\alpha) = h(\alpha) + \alpha C_1^* + (1-\alpha) C_2^*$$

в промежутке $0 \leqslant \alpha \leqslant 1$. Из уравнения

$$f'(\alpha) = \log_2 \frac{1-\alpha}{\alpha} + C_1^* - C_2^* = 0$$

находим стационарную точку

$$\alpha_0 = \frac{2^{C_1^*}}{2^{C_1^*} + 2^{C_2^*}}.$$

Нетрудно вычислить значение в этой точке: $f(\alpha_0) = \log_2(2^{C_1^*} + 2^{C_2^*})$. Так как это значение больше, чем значения $f(0) = C_2^*$ и $f(1) = C_1^*$ в крайних точках промежутка, получаем $C^* = f(\alpha_0)$. □

§ 4.4. Геометрическое представление пропускной способности

Если число различных символов на входе и на выходе канала связи невелико, то величине пропускной способности можно придать наглядное геометрическое истолкование.

Рассмотрим вначале геометрическое представление $(m-1)$-мерного симплекса $S_{m-1} = \{(p_1, \ldots, p_m)\colon p_i \geqslant 0, \sum_{i=1}^{m} p_i = 1\}$, на котором задана функция $H(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log_2 p_i$ (см. § 1.4). При небольших m такие симплексы изображены на следующем рисунке.

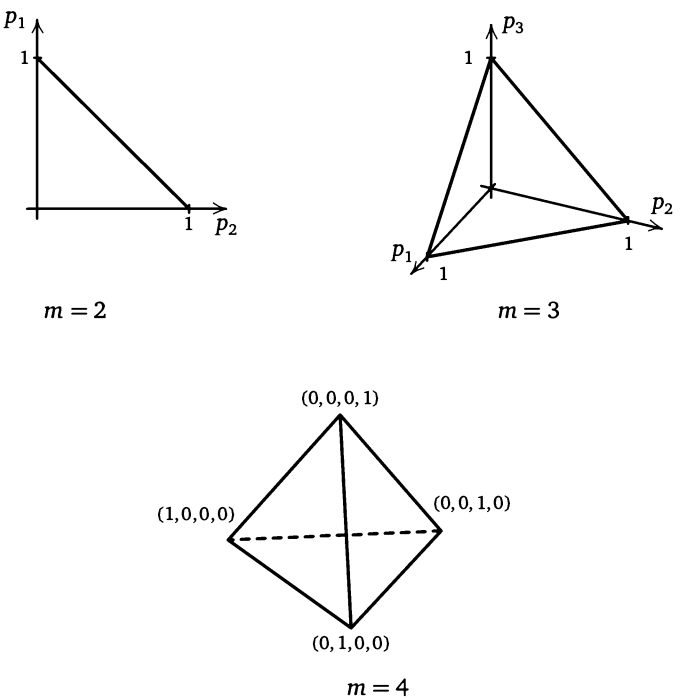

Рис. 4.6. Единичные симплексы размерности 1, 2 и 3

При $m = 2$ получается отрезок (одномерная фигура), соединяющий на плоскости (p_1, p_2) точки (0, 1) и (1, 0); при $m = 3$ получается треугольник (плоская фигура) с вершинами (0, 0, 1), (0, 1, 0) и (1, 0, 0)

в трёхмерном пространстве (p_1, p_2, p_3); при $m = 4$ получается тетраэдр (трёхмерная фигура) в четырёхмерном пространстве, и так далее.

Рассмотрим подробнее случай $m = 3$. Представим, что от каждой точки (p_1, p_2, p_3) симплекса S_2 перпендикулярно плоскости симплекса отложен отрезок длины $H(p_1, p_2, p_3)$. Концы этих отрезков образуют в пространстве поверхность, которую мы будем называть графиком функции $H(p_1, p_2, p_3)$. Будем обозначать эту поверхность $\mathscr{H}_3$. Из свойств функции энтропии (см. главу 1) ясно, что это — выпуклая поверхность, симметричная относительно каждой из трех плоскостей, перпендикулярных плоскости симплекса и проходящих через вершину треугольника и середину противоположной стороны; кроме того, эта поверхность переходит в себя при повороте на 120° вокруг прямой, перпендикулярной плоскости симплекса и проходящей через его центр $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

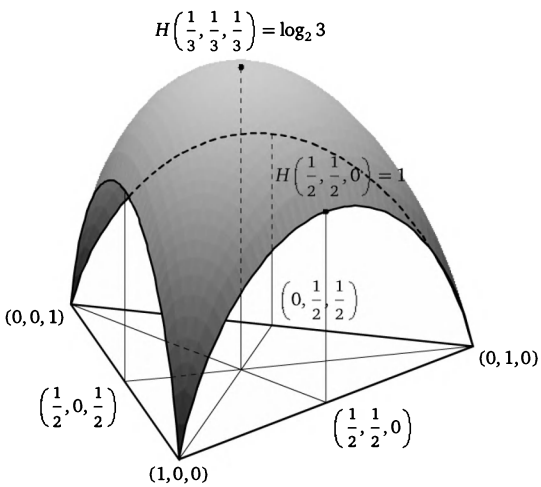

Рис. 4.7. Поверхность $\mathscr{H}_3$

Можно сказать, что эта поверхность представляет собой симметричную «крышу», опирающуюся на три точки в вершинах симплекса, с тремя входами-арками над каждой стороной симплекса.

Теперь мы подготовлены к рассмотрению геометрического смысла

пропускной способности (см. следующий рисунок). Предположим, что на входе и на выходе канала имеется по три различных символа, т. е. $q = s = 3$. Пусть $\pi_{3\times 3}$ — матрица переходных вероятностей канала, а через $\vec{\pi}_i$ обозначена i-я строка матрицы π, $i = 1, 2, 3$. Рассмотрим в симплексе S_2 треугольник $\triangle_1$ с вершинами в точках $\vec{\pi}_1$, $\vec{\pi}_2$ и $\vec{\pi}_3$. Если задано некоторое входное распределение $\vec{u} = (u_1, u_2, u_3)$, то соответствующее выходное распределение $\vec{v} = \vec{u} \cdot \pi = u_1\vec{\pi}_1 + u_2\vec{\pi}_2 + u_3\vec{\pi}_3$ будет представлять некоторую точку, лежащую в треугольнике $\triangle_1$.

Обозначим через A_i точки на поверхности $\mathscr{H}_3$, соответствующие вершинам треугольника $\triangle_1$, т. е. точки, отстоящие от симплекса на расстоянии $H(\vec{\pi}_i)$, и пусть $\triangle_2$ — треугольник с вершинами в точках A_i, $i = 1, 2, 3$.

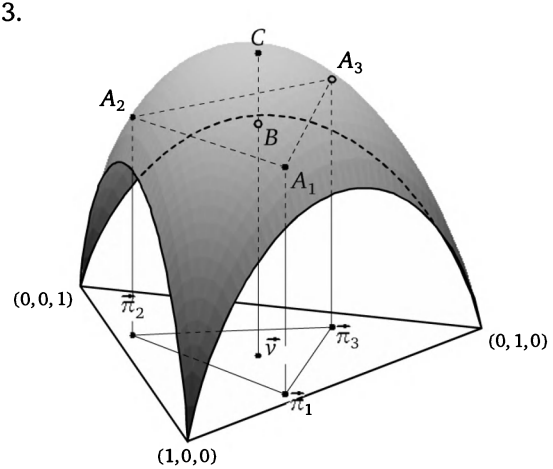

Рис. 4.8. Геометрическое представление пропускной способности

Прямая, перпендикулярная плоскости симплекса S_2 и проходящая через точку $\vec{v}$, пересекает треугольник $\triangle_2$ в некоторой точке B, отстоящей от симплекса на расстоянии $u_1H(\vec{\pi}_1) + u_2H(\vec{\pi}_2) + u_3H(\vec{\pi}_3)$. Далее, эта прямая пересекает поверхность $\mathscr{H}_3$ в некоторой точке C, отстоящей от симплекса на расстоянии $H(\vec{v})$. Расстояние между точками B и C равно

$$|BC| = H(\vec{v}) - \sum_{i=1}^{3} u_i H(\vec{\pi}_i) = I(X; Y),$$

где X и Y — сл. символы на входе и выходе канала соответственно.

Когда входное распределение $\vec{u}$ пробегает множество всех возможных значений, точка $\vec{v}$ пробегает треугольник $\triangle_1$, точка B пробегает треугольник $\triangle_2$, а точка C скользит по поверхности $\mathscr{H}_3$. Теперь ясно, что при некотором входном распределении $\vec{u}$ расстояние $|BC|$ будет максимальным — это максимальное значение и есть величина пропускной способности нашего канала связи.

§ 4.5. Задачи

4.1. Вычислить пропускную способность и оптимальное входное распределение для канала связи, заданного матрицей π переходных вероятностей. Если матрица содержит параметр p, построить график зависимости пропускной способности от p.

а) $\begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}$; б) $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$; в) $\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$;

г) $\begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$; д) $\begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$; е) $\begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$;

ж) $\begin{pmatrix} \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$; з) $\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$; и) $\begin{pmatrix} \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{5}{16} & \frac{3}{8} & \frac{5}{16} \\ \frac{3}{8} & \frac{1}{16} & \frac{9}{16} \end{pmatrix}$;

к) $\begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{pmatrix}$; л) $\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$;

м) $\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$ (единичная);

н) $\begin{pmatrix} \frac{1}{n} & \dots & \frac{1}{n} \\ \dots & \dots & \dots \\ \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$ (равномерная);

о) $\pi = \alpha S + (1-\alpha)T$, где $0 \leqslant \alpha \leqslant 1$, S и T — подстановочные матрицы;

п) $\begin{pmatrix} \frac{1-p}{2} & \frac{1-p}{2} & \frac{p}{2} & \frac{p}{2} \\ \frac{p}{2} & \frac{p}{2} & \frac{1-p}{2} & \frac{1-p}{2} \end{pmatrix}$; р)* $\begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}$;

с) $\begin{pmatrix} 1-p & p & 0 \\ p & 1-p & 0 \\ p & 0 & 1-p \end{pmatrix}$; т) $\begin{pmatrix} 1-p & p & 0 \\ p & 1-p & 0 \\ 0 & 0 & 1 \end{pmatrix}$;

у) $\begin{pmatrix} p & 0 & \dots & 0 & 1-p \\ 0 & p & \dots & 0 & 1-p \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p & 1-p \\ 1-p & 0 & \dots & 0 & p \end{pmatrix}$;

ф) $\begin{pmatrix} p & 1-p & 0 & 0 \\ 1-p & p & 0 & 0 \\ 0 & 0 & p & 1-p \\ 0 & 0 & 1-p & p \end{pmatrix}$.

4.2. Случайный символ X, порождаемый дискретным источником с алфавитом $\mathscr{X} = \{0, 1, 2\}$ и распределением $\vec{p} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$, передаётся одновременно по двум каналам связи с матрицами переходных вероятностей $\pi_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ и $\pi_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$. Пусть Y и Z — сл. символы на выходе первого и второго канала соответственно. Вычислить $H(X)$, $H(Y)$, $H(Z)$, $H(Y, Z)$, $I(X; Y)$, $I(X; Z)$, $I(X; Y/Z)$, $I(X; (Y, Z))$.

4.3. Два ДСК с параметрами $p_1 = \frac{1}{4}$ и $p_2 = \frac{1}{3}$ соединены последовательно. Пусть X и Y — входной и выходной символы первого ДСК, Y и Z — входной и выходной символы второго ДСК и X

* Канал связи с такой матрицей π называют Z-каналом, так как его диаграмма похожа на букву Z.

принимает значения 0 и 1 с вероятностями $\frac{1}{2}$. Вычислить величины $I(X;Y)$, $I(Y;Z)$, $I(X;Z)$.

4.4. Вычислить пропускную способность $C_n^*(p)$ для последовательного соединения из n двоичных симметричных каналов с одинаковым параметром p. Найти предел $\lim_{n\to\infty} C_n^*(p)$ как функцию от p.

4.5. Для канала связи с алфавитами $\mathscr{X} = \mathscr{Y} = \{0, 1, 2, 3, 4\}$ переходная вероятность $\pi(y/x)$ равна $\frac{1}{2}$, если $y = x \pm 1 \pmod 5$, и равна 0 в противном случае. Вычислить пропускную способность.

4.6. К случайному входному символу X, принимающему числовые значения $0, 1, \ldots, 10$, в канале связи прибавляется по модулю 11 независимая случайная помеха Z, принимающая значения 1, 2 и 3 с равными вероятностями, так что на выходе получается $Y = X + Z \pmod{11}$. Вычислить пропускную способность и оптимальное входное распределение.

4.7. К случайному входному символу X, принимающему числовые значения 0 и 1, в канале связи прибавляется независимая случайная помеха Z, принимающая значения 0 и a с равными вероятностями, так что на выходе получается $Y = X + Z$. Вычислить пропускную способность как функцию от параметра a.

4.8. К случайному входному символу X, принимающему числовые значения 0, 1, 2, 3, в канале связи прибавляется независимая случайная помеха Z, принимающая значения a и b с равными вероятностями, так что на выходе получается $Y = X + Z$. Вычислить пропускную способность как функцию от параметров a и b. При каких значениях параметров пропускная способность максимальна? минимальна?

4.9. Случайный входной символ X, принимающий числовые значения 0 и 1, в канале связи умножается на независимую случайную помеху Z, принимающую значения 0 и 1 с вероятностями p и $1 - p$ соответственно, так что на выходе получается $Y = X \cdot Z$. Вычислить пропускную способность как функцию от параметра p.

4.10. Пусть дан $(\mathscr{X}, \mathscr{Y}, \pi_1)$ — канал связи с пропускной способностью C^*, и пусть $(\mathscr{Y}, \mathscr{Z}, \pi_2)$ — канал со стиранием, т. е. $\mathscr{Y} =$

$= \{y_1, \ldots, y_s\}$, $\mathscr{Z} = \{y_1, \ldots, y_s, E\}$ и

$$\pi_2 = \begin{pmatrix} 1-p & 0 & \ldots & 0 & p \\ 0 & 1-p & \ldots & 0 & p \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1-p & p \end{pmatrix}.$$

Вычислить пропускную способность последовательного соединения каналов $(\mathscr{X}, \mathscr{Y}, \pi_1)$ и $(\mathscr{Y}, \mathscr{Z}, \pi_2)$.

Глава 5
Теоремы кодирования для дискретных каналов без памяти

§ 5.1. Вероятность ошибочного декодирования

Пусть даны два конечных множества $\mathscr{X}$ и $\mathscr{Y}$ и для $n \geqslant 1$ задан канал связи $(\mathscr{X}^n, \mathscr{Y}^n, \pi^{(n)})$. Как и выше, X^n есть сл. вектор на входе канала связи, а Y^n — соответствующий сл. вектор на выходе канала.

Произвольное подмножество

$$\mathscr{C} = \{x^n(1), \ldots, x^n(M)\} \subseteq \mathscr{X}^n$$

будем называть кодом длины n и объёма M, состоящим из кодовых слов $x^n(i) = (x_1(i), \ldots, x_n(i))$, $1 \leqslant i \leqslant M$.

Предположим, что по каналу связи передаются только кодовые слова кода $\mathscr{C}$. Это означает, что распределение $U_{1,\ldots,n}$ сл. вектора X^n сосредоточено на коде $\mathscr{C}$, т. е. $U_{1,\ldots,n}(x^n) = 0$, если $x^n \in \mathscr{X}^n \setminus \mathscr{C}$.

Определение 5.1. Скоростью передачи информации по каналу связи $(\mathscr{X}^n, \mathscr{Y}^n, \pi^{(n)})$ будем называть величину

$$R_n = \frac{1}{n} H(X^n).$$

Замечание 5.1. Величина R_n имеет смысл среднего количества информации, приходящегося на один символ передаваемого по каналу кодового слова. В частности, если распределение $U_{1,\ldots,n}$ является равномерным на коде $\mathscr{C}$, то

$$R_n = \frac{\log_2 M}{n}. \tag{5.1}$$

Определение 5.2. Пусть $\mathscr{Y}^n = A_1 \sqcup \ldots \sqcup A_M \sqcup A_{M+1}$ — произвольное разбиение множества $\mathscr{Y}^n$ на $M+1$ непересекающихся подмножеств $A_1, \ldots, A_{M+1}$, называемых *решающими областями* декодера. *Множество решений* $W = \mathscr{C} \cup \{\epsilon\}$ состоит из кодовых слов кода $\mathscr{C}$ и дополнительного элемента ϵ, имеющего смысл сообщения об ошибке типа «декодер не может принять решения». *Декодером общего вида* назовём произвольное отображение $\mathscr{D}\colon \mathscr{Y}^n \to W$.

Если сл. вектор Y^n на выходе канала связи принадлежит решающей области A_i, $1 \leqslant i \leqslant M$, то декодер $\mathscr{D}$ принимает решение о том, что на входе канала было кодовое слово $X^n = x^n(i)$; если же $Y^n = y^n \in A_{M+1}$, то декодер объявляет об ошибке ϵ, т. е. декодер не может принять решение в пользу какого-либо кодового слова из кода $\mathscr{C}$.

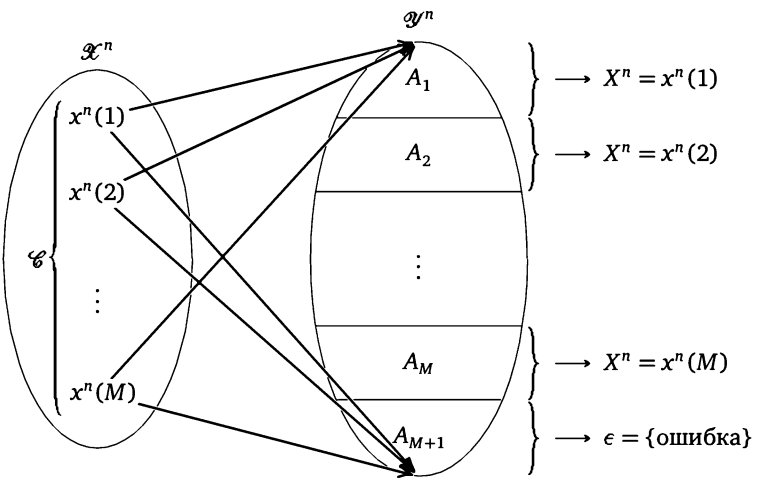

Рис. 5.1. Декодер общего вида

Определение 5.3. Предположим, что $X^n = x^n(i) \in \mathscr{C}$ и $Y^n = y^n \in A_j$, $1 \leqslant i \leqslant M$, $1 \leqslant j \leqslant M+1$. Если при этом $i = j$, то говорим, что принятое сообщение Y^n *правильно декодировано*, а если $i \neq j$, то говорим, что произошла *ошибка декодирования*.

Замечание 5.2. Введём множество

$$\mathscr{E} = \bigcup_{i \neq j} (\{x^n(i)\} \times A_j) \subseteq \mathscr{X}^n \times \mathscr{Y}^n.$$

Тогда, согласно определению 5.3, событие *ошибочное декодирование* означает, что $(X^n, Y^n) \in \mathscr{E}$, или, что то же самое, $\mathscr{D}(Y^n) \neq X^n$.

Пусть заданы канал связи $(\mathscr{X}^n, \mathscr{Y}^n, \pi^{(n)})$, код $\mathscr{C} \subseteq \mathscr{X}^n$ и декодер $\mathscr{D}$. Качество полученной системы передачи информации будем измерять вероятностями принятия ошибочных решений при декодиро-

вании, оставляя пока в стороне вопрос о других критериях верности передачи.

Определение 5.4. *Условной вероятностью ошибочного декодирования* при условии, что было передано кодовое слово $X^n = x^n(i)$, называется величина

$$\lambda_i = \mathsf{P}((X^n, Y^n) \in \mathscr{E} / X^n = x^n(i)).$$

Средней вероятностью ошибочного декодирования называется величина

$$\lambda = \mathsf{P}((X^n, Y^n) \in \mathscr{E}).$$

Замечание 5.3. Далее, когда будет необходимо подчеркнуть зависимость декодера $\mathscr{D}$ или средней вероятности ошибочного декодирования λ от длины n кодового слова, будем добавлять к обозначению индекс n и писать $\mathscr{D}_n$ и $\lambda^{(n)}$.

Утверждение 5.1. *Справедливы формулы*

$$\lambda_i = 1 - \sum_{y^n \in A_i} \pi^{(n)}(y^n / x^n(i)), \tag{5.2}$$

$$\lambda = 1 - \sum_{i=1}^{M} \mathsf{P}_n(x^n(i)) \sum_{y^n \in A_i} \pi^{(n)}(y^n / x^n(i)). \tag{5.3}$$

Доказательство. Непосредственно из определения 5.2 получаем, что условная вероятность принятия решения $\mathscr{D}(Y^n) = x^n(j)$ при условии, что было передано $X^n = x^n(i)$, равна

$$\mathsf{P}(\mathscr{D}(Y^n) = x^n(j) / X^n = x^n(i)) = \sum_{y^n \in A_j} \pi^{(n)}(y^n / x^n(i)).$$

Поэтому

$$\begin{aligned} \lambda_i &= \mathsf{P}((X^n, Y^n) \in \mathscr{E} / X^n = x^n(i)) = \\ &= 1 - \mathsf{P}(\mathscr{D}(Y^n) = x^n(i) / X^n = x^n(i)) = \\ &= 1 - \sum_{y^n \in A_i} \pi^{(n)}(y^n / x^n(i)). \end{aligned}$$

Следовательно, по формуле полной вероятности

$$\lambda = \sum_{i=1}^{M} \lambda_i \mathsf{P}_n(x^n(i)) = 1 - \sum_{i=1}^{M} \mathsf{P}_n(x^n(i)) \sum_{y^n \in A_i} \pi^{(n)}(y^n / x^n(i)),$$

что и требовалось. □

Замечание 5.4. В некоторых случаях важно знать также минимальную и максимальную вероятности ошибочного декодирования

$$\lambda_{\min} = \min_{1 \leqslant i \leqslant M} \lambda_i, \qquad \lambda_{\max} = \max_{1 \leqslant i \leqslant M} \lambda_i.$$

В качестве примеров рассмотрим два важных декодера. Напомним, что если X^n — сл. кодовое слово на входе канала связи, и Y^n — соответствующий сл. вектор на выходе, то для любого кодового слова x^n и любого вектора $y^n \in \mathscr{Y}^n$ определены переходная вероятность

$$\pi^{(n)}(y^n/x^n) = \mathsf{P}(Y^n = y^n/X^n = x^n)$$

и вероятность $\mathsf{P}(Y^n = y^n) = \sum\limits_{x^n \in \mathscr{C}} U_{1,\ldots,n}(x^n)\pi^{(n)}(y^n/x^n)$.

Декодер $\mathscr{D}_{\mathsf{L}}$ — по методу максимального правдоподобия. Алгоритм декодера $\mathscr{D}_{\mathsf{L}}$ состоит в том, что если принят вектор y^n, то вычисляются переходные вероятности $\pi^{(n)}(y^n/x^n)$ для всех кодовых слов $x^n \in \mathscr{C}$ и в качестве результата декодирования выбирается то кодовое слово, для которого эта вероятность максимальна. Если максимальное значение переходной вероятности достигается для нескольких кодовых слов, то выбирается одно из них по некоторому заранее оговорённому правилу, например наименьшее в смысле лексикографического порядка.

Декодер $\mathscr{D}_{\mathsf{AP}}$ — по методу максимальной апостериорной вероятности. Если $\mathsf{P}(Y^n = y^n) > 0$, то определена условная вероятность

$$\mathsf{P}(X^n = x^n/Y^n = y^n) = \frac{U_{1,\ldots,n}(x^n)\pi^{(n)}(y^n/x^n)}{\sum_{z^n \in \mathscr{C}} U_{1,\ldots,n}(z^n)\pi^{(n)}(y^n/z^n)}. \tag{5.4}$$

Эту вероятность обычно называют *апостериорной* вероятностью кодового слова x^n, т. е. вероятностью *после опыта*, после получения вектора y^n из канала связи, в то время как вероятность $\mathsf{P}(X^n = x^n) = U_{1,\ldots,n}(x^n)$ называют *априорной*, т. е. *до опыта*.

Алгоритм декодера $\mathscr{D}_{\mathsf{AP}}$ состоит в том, что если принят вектор y^n, то вычисляются апостериорные вероятности $\mathsf{P}(X^n = x^n/Y^n = y^n)$ для всех кодовых слов $x^n \in \mathscr{C}$ и в качестве результата декодирования выбирается то кодовое слово, для которого эта вероятность максимальна. Если максимальное значение апостериорной вероятности достигается для нескольких кодовых слов, то, как и выше, выбирается одно из них по некоторому заранее оговорённому правилу.

Отметим, что декодер $\mathscr{D}_{\mathsf{L}}$, в отличие от декодера $\mathscr{D}_{\mathsf{AP}}$, по определению не зависит от распределения $U_{1,\dots,n}(x^n)$ на множестве кодовых слов и поэтому является более простым; декодер $\mathscr{D}_{\mathsf{L}}$ зависит только от кода и от канала связи и может применяться для любого распределения $U_{1,\dots,n}(x^n)$. С другой стороны, декодер $\mathscr{D}_{\mathsf{AP}}$, использующий распределение на множестве кодовых слов, является в определённом смысле наилучшим, как показывает следующее

Утверждение 5.2. *Декодер $\mathscr{D}_{\mathsf{AP}}$ обеспечивает минимальную среднюю вероятность λ ошибочного декодирования.*

Доказательство. Пусть заданы канал связи $(\mathscr{X}^n, \mathscr{Y}^n, \pi^{(n)})$, код $\mathscr{C} \subseteq \mathscr{X}^n$ и декодер $\mathscr{D}$ с решающими областями $A_1, \dots, A_{M+1}$. Далее для краткости будем обозначать

$$\mathsf{P}(y^n) = \mathsf{P}(Y^n = y^n), \qquad \mathsf{P}(x^n/y^n) = \mathsf{P}(X^n = x^n/Y^n = y^n).$$

Запишем среднюю вероятность правильного декодирования в виде

$$\begin{aligned} 1 - \lambda &= \mathsf{P}(\mathscr{D}(Y^n) = X^n) = \\ &= \sum_{y^n \in \mathscr{Y}^n} \mathsf{P}(y^n)\mathsf{P}(\mathscr{D}(Y^n) = X^n/Y^n = y^n) = \\ &= \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n)\mathsf{P}(x^n(i)/y^n). \end{aligned}$$

Из такого представления ясно, что величина $1 - \lambda$ максимальна, если каждая решающая область A_i, $i = 1, \dots, M$, декодера $\mathscr{D}$ состоит из таких векторов y^n, для которых $\mathsf{P}(x^n(i)/y^n) \geqslant \mathsf{P}(x^n(j)/y^n)$ при всех $j = 1, \dots, M$. Это условие совпадает с определением декодера $\mathscr{D}_{\mathsf{AP}}$. □

Определение 5.5. Два декодера $\mathscr{D}_1$ и $\mathscr{D}_2$ называем эквивалентными, если для любых $y^n \in \mathscr{Y}^n$ справедливо равенство $\mathscr{D}_1(y^n) = \mathscr{D}_2(y^n)$.

Утверждение 5.3. *Если распределение $U_{1,\dots,n}(x^n)$ на множестве кодовых слов равномерное, то декодеры $\mathscr{D}_{\mathsf{L}}$ и $\mathscr{D}_{\mathsf{AP}}$ эквивалентны.*

Доказательство. Если распределение $U_{1,\dots,n}(x^n)$ равномерное, то формула (5.4) принимает вид

$$\mathsf{P}(X^n = x^n/Y^n = y^n) = \frac{\pi^{(n)}(y^n/x^n)}{\sum\limits_{z^n \in \mathscr{C}} \pi^{(n)}(y^n/z^n)}.$$

Так как знаменатель в правой части не зависит от x^n, апостериорная вероятность $\mathsf{P}(X^n = x^n / Y^n = y^n)$ максимальна по x^n тогда и только тогда, когда максимальна переходная вероятность $\pi^{(n)}(y^n / x^n)$. □

§ 5.2. Обратная теорема кодирования

Наша цель в этой главе — изложение двух принципиальных результатов теории информации: прямой и обратной теорем кодирования для дискретного канала связи без памяти. Вначале мы рассмотрим в данном параграфе обратную теорему кодирования, поскольку её доказательство более короткое и простое. Это доказательство будет опираться на одно важное неравенство, имеющее и самостоятельное значение.

Теорема 5.1 (неравенство Фано). *Если выполнено хотя бы одно из условий*

$$\mathsf{P}(Y^n \in A_{M+1}) = 0 \quad \text{или} \quad \lambda \leqslant 1 - \frac{1}{M},$$

то справедливо неравенство

$$H(X^n / Y^n) \leqslant h(\lambda) + \lambda \log_2 (M - 1), \tag{5.5}$$

где $h(\lambda)$ — функция двоичной энтропии.

Доказательство. Представим среднюю вероятность правильного декодирования, как в доказательстве утверждения 5.2, в виде

$$1 - \lambda = \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n) \mathsf{P}(x^n(i) / y^n),$$

а вероятность ошибочного декодирования — в виде

$$\begin{aligned} \lambda &= \mathsf{P}(\mathscr{D}(Y^n) \neq X^n) = \\ &= \sum_{y^n \in \mathscr{Y}^n} \mathsf{P}(y^n) \mathsf{P}(\mathscr{D}(Y^n) \neq X^n / Y^n = y^n) = \\ &= \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n) \sum_{1 \leqslant j \leqslant M,\, j \neq i} \mathsf{P}(x^n(j) / y^n) + \mathsf{P}(Y^n \in A_{M+1}). \end{aligned}$$

Аналогично выпишем формулу для условной энтропии $H(X^n/Y^n)$,

$$H(X^n/Y^n) = -\sum_{x^n,\, y^n} \mathsf{P}(x^n, y^n) \log_2 \mathsf{P}(x^n/y^n) =$$
$$= -\sum_{i=1}^{M+1} \sum_{y^n \in A_i} \mathsf{P}(y^n) \sum_{j=1}^{M} \mathsf{P}(x^n(j)/y^n) \log_2 \mathsf{P}(x^n(j)/y^n).$$

Используя полученные выражения, запишем разность

$$H(X^n/Y^n) - h(\lambda) - \lambda \log_2(M-1)$$

в виде

$$H(X^n/Y^n) - h(\lambda) - \lambda \log_2(M-1) =$$
$$= H(X^n/Y^n) + (1-\lambda)\log_2(1-\lambda) + \lambda \log_2 \frac{\lambda}{M-1} =$$
$$= -\sum_{i=1}^{M+1} \sum_{y^n \in A_i} \mathsf{P}(y^n) \sum_{j=1}^{M} \mathsf{P}(x^n(j)/y^n) \log_2 \mathsf{P}(x^n(j)/y^n) +$$
$$+ \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n) \mathsf{P}(x^n(i)/y^n) \log_2(1-\lambda) +$$
$$+ \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n) \sum_{1 \leqslant j \leqslant M,\, j \neq i} \mathsf{P}(x^n(j)/y^n) \log_2 \frac{\lambda}{M-1} +$$
$$+ \mathsf{P}(Y^n \in A_{M+1}) \log_2 \frac{\lambda}{M-1} =$$
$$= \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n) \mathsf{P}(x^n(i)/y^n) \log_2 \frac{1-\lambda}{\mathsf{P}(x^n(i)/y^n)} +$$
$$+ \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n) \sum_{j \neq i} \mathsf{P}(x^n(j)/y^n) \log_2 \frac{\lambda}{(M-1)\mathsf{P}(x^n(j)/y^n)} +$$
$$+ \sum_{y^n \in A_{M+1}} \mathsf{P}(y^n) \sum_{j=1}^{M} \mathsf{P}(x^n(j)/y^n) \log_2 \frac{\lambda}{(M-1)\mathsf{P}(x^n(j)/y^n)} =$$
$$= \Sigma_1 + \Sigma_2 + \Sigma_3.$$

Суммы Σ_1, Σ_2 и Σ_3 оценим сверху при помощи неравенства (1.3) (с. 24):

$$\Sigma_1 \leqslant \log_2 e \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n)[1-\lambda-\mathsf{P}(x^n(i)/y^n)],$$

$$\Sigma_2 \leqslant \log_2 e \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n) \sum_{j \neq i} \Big[\frac{\lambda}{M-1} - \mathsf{P}(x^n(j)/y^n)\Big] =$$

$$= \log_2 e \sum_{i=1}^{M} \sum_{y^n \in A_i} \mathsf{P}(y^n)[\lambda - 1 + \mathsf{P}(x^n(i)/y^n)],$$

$$\Sigma_3 \leqslant \log_2 e \sum_{y^n \in A_{M+1}} \mathsf{P}(y^n) \sum_{j=1}^{M} \Big[\frac{\lambda}{M-1} - \mathsf{P}(x^n(j)/y^n)\Big] =$$

$$= \log_2 e \cdot \mathsf{P}(Y^n \in A_{M+1}) \Big(\frac{\lambda M}{M-1} - 1\Big).$$

Так как полученные верхние оценки для сумм Σ_1 и Σ_2 равны по модулю и имеют противоположные знаки, $\Sigma_1 + \Sigma_2 \leqslant 0$, и мы приходим к неравенству

$$H(X^n/Y^n) - h(\lambda) - \lambda \log_2(M-1) \leqslant \log_2 e \cdot \mathsf{P}(Y^n \in A_{M+1}) \Big(\frac{\lambda M}{M-1} - 1\Big) \tag{5.6}$$

При условиях теоремы правая часть неравенства (5.6) не больше нуля, поэтому из (5.6) вытекает искомое неравенство (5.5). □

Замечание 5.5. 1. Неравенство (5.6), которое можно назвать усиленной формой неравенства Фано (5.5), равносильно (5.5), если $\mathsf{P}(Y^n \in A_{M+1}) = 0$ или $\lambda = 1 - \frac{1}{M}$. Условие $\mathsf{P}(Y^n \in A_{M+1}) = 0$ означает, что декодер всегда принимает решение в пользу какого-то кодового слова и никогда не сообщает об ошибке «декодер не может принять решения».

2. Принимая во внимание условие обращения неравенства (1.3) в равенство, получаем, что в неравенстве $\Sigma_1 + \Sigma_2 \leqslant 0$ равенство достигается тогда и только тогда, когда для любых $1 \leqslant i \leqslant M$ и $y^n \in A_i$ выполнены условия

$$\mathsf{P}(x^n(i)/y^n) = 1 - \lambda, \qquad \mathsf{P}(x^n(j)/y^n) = \frac{\lambda}{M-1}, \ j \neq i.$$

Это значит, что для любого принятого вектора $y^n \notin A_{M+1}$ декодер может принять решение в пользу правильного кодового слова с вероятностью $1-\lambda$ или в пользу любого из $M-1$ других кодовых слов

с одинаковыми вероятностями $\lambda/(M-1)$. При $\mathsf{P}(Y^n \in A_{M+1}) = 0$ указанные условия необходимы и достаточны для обращения неравенства Фано (5.5) в равенство.

Если же $\mathsf{P}(Y^n \in A_{M+1}) > 0$, то для обращения усиленного неравенства Фано (5.6) в равенство необходимо, чтобы дополнительно для всех $y^n \in A_{M+1}$ и $1 \leqslant j \leqslant M$ было выполнено условие $\mathsf{P}(x^n(j)/y^n) = \lambda/(M-1)$, из которого следует, что $\lambda = 1 - \frac{1}{M}$.

Замечание 5.6. Рассмотрим функцию

$$g(\lambda) = h(\lambda) + \lambda \log_2(M-1),$$

фигурирующую в правой части неравенства Фано. Очевидно, $g(0) = 0$, $g(1) = \log_2(M-1)$. Из уравнения

$$g'(\lambda) = \log_2 \frac{1-\lambda}{\lambda} + \log_2(M-1) = 0$$

находим единственную стационарную точку $\lambda_0 = 1 - \frac{1}{M}$, в которой функция принимает значение $g(\lambda_0) = \log_2 M$.

В промежутке $0 \leqslant \lambda \leqslant \lambda_0$ функция возрастает от 0 до $\log_2 M$, а в промежутке $\lambda_0 \leqslant \lambda \leqslant 1$ функция убывает от $\log_2 M$ до $\log_2(M-1)$. График функции g изображен на следующем рисунке.

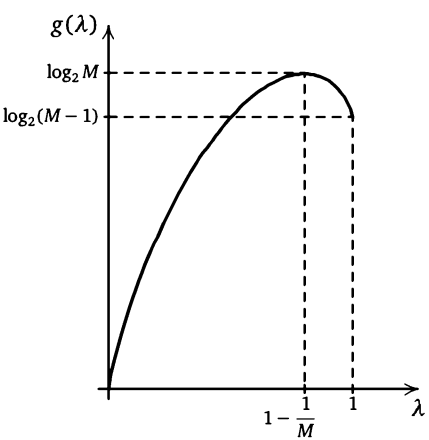

Рис. 5.2. График функции $g(\lambda)$

Рассмотрим подробнее значение $\lambda_0 = 1 - \frac{1}{M}$. Предположим, что на выходе канала связи имеется так называемый случайный декодер, т. е. независимо от принятого из канала вектора $y^n \in \mathcal{Y}^n$ декодер принимает решение, выбирая случайно и равновероятно произвольное кодовое слово из кода $\mathcal{C} = \{x^n(1), \ldots, x^n(M)\}$. При этом, очевидно, правильное декодирование происходит с вероятностью $\frac{1}{M}$, а ошибка декодирования — с вероятностью $\lambda = \lambda_0$. Если же наш декодер не случайный, а устроен некоторым разумным образом, т. е. результат декодирования зависит от принятого вектора и свойств канала связи, естественно предполагать, что $\lambda \leqslant \lambda_0$.

Теорема 5.2 (обратная теорема кодирования). *Пусть $\varepsilon > 0$, и пусть $(\mathcal{X}, \mathcal{Y}, \pi)$ — канал связи с пропускной способностью C^*. Предположим, что для всех $n = 1, 2, \ldots$ заданы такие код $\mathcal{C}^{(n)} \subseteq \mathcal{X}^n$ и распределение $U_{1,\ldots,n}$ на множестве $\mathcal{X}^n$, сосредоточенное на коде $\mathcal{C}^{(n)}$, что скорость передачи $R_n = \frac{1}{n} H(X^n)$ удовлетворяет условию $R_n \geqslant C^* + \varepsilon$.*

Тогда найдётся такое число $\delta > 0$, что для любого $n \geqslant 1$ и любого декодера общего вида $\mathcal{D}_n$ средняя вероятность ошибочного декодирования $\lambda^{(n)}$ удовлетворяет условию $\lambda^{(n)} \geqslant \delta$.

Замечание 5.7. Теорема означает, что если скорость передачи информации больше, чем пропускная способность канала связи, то никаким, даже сколь угодно сложным, способом кодирования и декодирования нельзя добиться, чтобы вероятность ошибочного декодирования стала меньше определённого положительного числа.

Доказательство. теоремы 5.2. Согласно определению 4.5, для любого распределения $U_{1,\ldots,n}$ на множестве $\mathcal{X}^n$ справедливо неравенство

$$I(X^n; Y^n) \leqslant nC^*. \tag{5.7}$$

По условию теоремы,

$$H(X^n) = nR_n \geqslant n(C^* + \varepsilon) > 0.$$

Отсюда, в частности, следует, что число кодовых слов $M_n = |\mathcal{C}^{(n)}|$ удовлетворяет неравенству $M_n \geqslant 2$ и если $\lambda^{(n)} > 1 - \frac{1}{M_n}$, то $\lambda^{(n)} > \frac{1}{2}$.

Если же $\lambda^{(n)} \leqslant 1 - \frac{1}{M_n}$, то по неравенству Фано (теорема 5.1) для

произвольного декодера $\mathscr{D}_n$ имеем

$$H(X^n/Y^n) \leqslant g_n(\lambda^{(n)}) = h(\lambda^{(n)}) + \lambda^{(n)} \log_2(M_n - 1).$$

Следовательно,

$$I(X^n; Y^n) = H(X^n) - H(X^n/Y^n) \geqslant n(C^* + \varepsilon) - g_n(\lambda^{(n)}). \tag{5.8}$$

Из неравенств (5.7) и (5.8) вытекает, что

$$g_n(\lambda^{(n)}) \geqslant n\varepsilon, \tag{5.9}$$

и поэтому $\lambda^{(n)} > 0$ для всех $n = 1, 2, \ldots$ (см. замечание 5.6 о свойствах функции g). С другой стороны, $M_n = |\mathscr{C}^{(n)}| \leqslant |\mathscr{X}^n| \leqslant q^n$, поэтому

$$g_n(\lambda^{(n)}) \leqslant h(\lambda^{(n)}) + \lambda^{(n)} n \log_2 q. \tag{5.10}$$

Из неравенств (5.9) и (5.10) получаем

$$\varepsilon \leqslant \frac{1}{n} h(\lambda^{(n)}) + \lambda^{(n)} \log_2 q. \tag{5.11}$$

Если $\lambda^{(n_k)} \to 0$ для некоторой подпоследовательности n_k, $k = 1, 2, \ldots$, то правая часть (5.11) по этой подпоследовательности n_k также стремится к нулю, и мы приходим к противоречию $\varepsilon \leqslant o(1)$.

Итак, последовательность чисел $\lambda^{(n)} > 0$ такова, что никакая её подпоследовательность не стремится к нулю. Поэтому найдется такое $\delta > 0$, что $\lambda^{(n)} \geqslant \delta$ для любого $n \geqslant 1$, что и требовалось доказать. □

Замечание 5.8. При условиях теоремы 5.2 справедлива так называемая *усиленная обратная теорема кодирования*, согласно которой последовательность $\lambda^{(n)}$ не только отделена от нуля, но и стремится к единице экспоненциально быстро. Точная формулировка и доказательство усиленной обратной теоремы выходят за рамки нашего пособия.

§ 5.3. Прямая теорема кодирования

Пусть дана пара сл. векторов X^n, Y^n с совместным распределением

$$p_{X^nY^n}(x^n, y^n) = \mathsf{P}(X^n = x^n, Y^n = y^n), \quad x^n \in \mathscr{X}^n, y^n \in \mathscr{Y}^n,$$

и частными распределениями

$$p_{X^n}(x^n) = \mathsf{P}(X^n = x^n) = \sum_{y^n \in \mathscr{Y}^n} p_{X^n Y^n}(x^n, y^n),$$

$$p_{Y^n}(y^n) = \mathsf{P}(Y^n = y^n) = \sum_{x^n \in \mathscr{X}^n} p_{X^n Y^n}(x^n, y^n).$$

Определение 5.6. Пусть $\varepsilon > 0$. Назовём *ε-типичными*

1) вектор x^n, если $\left|\log_2 \frac{1}{p_{X^n}(x^n)} - H(X^n)\right| < n\varepsilon$;

2) вектор y^n, если $\left|\log_2 \frac{1}{p_{Y^n}(y^n)} - H(Y^n)\right| < n\varepsilon$;

3) пару (x^n, y^n), если ε-типичны x^n и y^n и, кроме того,

$$\left|\log_2 \frac{1}{p_{X^n Y^n}(x^n, y^n)} - H(X^n, Y^n)\right| < n\varepsilon. \tag{5.12}$$

Множество всех ε-типичных пар обозначим $\mathscr{W}_{n,\varepsilon}$.

Для доказательства прямой теоремы кодирования понадобится так называемый *декодер ε-типичных пар*.

Определение 5.7. Пусть по каналу связи $(\mathscr{X}, \mathscr{Y}, \pi)$ передаются кодовые слова кода $\mathscr{C} = \{x^n(1), \ldots, x^n(M)\} \subseteq \mathscr{X}^n$.

Декодером ε-типичных пар назовём такое отображение $\mathscr{D}_\varepsilon : \mathscr{Y}^n \to W$, что если для принятого вектора y^n существует, и притом единственное, кодовое слово $x^n(i) \in \mathscr{C}$, образующее вместе с y^n ε-типичную пару $(x^n(i), y^n)$, то $\mathscr{D}_\varepsilon(y^n) = x^n(i)$; в противном случае $\mathscr{D}_\varepsilon(y^n) = \epsilon$, т. е. декодер объявляет об ошибке.

Рассмотрим один частный случай совместного распределения векторов (X^n, Y^n). Пусть дана пара сл. величин (X, Y) с совместным распределением

$$p_{XY}(x, y) = \mathsf{P}(X = x, Y = y), \quad x \in \mathscr{X}, y \in \mathscr{Y},$$

и частными распределениями

$$p_X(x) = \sum_{y \in \mathscr{Y}} p_{XY}(x, y), \quad p_Y(y) = \sum_{x \in \mathscr{X}} p_{XY}(x, y).$$

Пусть распределение пары (X^n, Y^n) имеет вид

$$p_{X^n Y^n}(x^n, y^n) = \mathsf{P}(X^n = x^n, Y^n = y^n) = \prod_{i=1}^{n} p_{XY}(x_i, y_i),$$

т. е. пары $(X_1, Y_1), \ldots, (X_n, Y_n)$ независимы и распределены так же, как пара (X, Y). Частные распределения векторов X^n и Y^n при этом имеют вид

$$p_{X^n}(x^n) = \mathsf{P}(X^n = x^n) = \sum_{y^n \in \mathscr{Y}^n} p_{X^n Y^n}(x^n, y^n) = \prod_{i=1}^{n} p_X(x_i), \quad x^n \in \mathscr{X}^n,$$

$$p_{Y^n}(y^n) = \mathsf{P}(Y^n = y^n) = \sum_{x^n \in \mathscr{X}^n} p_{X^n Y^n}(x^n, y^n) = \prod_{i=1}^{n} p_Y(y_i), \quad y^n \in \mathscr{Y}^n.$$

Отметим, что

$$H(X^n) = nH(X), \quad H(Y^n) = nH(Y), \quad H(X^n, Y^n) = nH(X, Y). \quad (5.13)$$

При сформулированных предположениях о независимости справедливы следующие три леммы о *совместной асимптотической равнораспределённости*.

Лемма 5.1. *При всех достаточно больших n*

$$\mathsf{P}((X^n, Y^n) \in \mathscr{W}_{n,\varepsilon}) > 1 - \varepsilon.$$

Лемма 5.2. *При всех n*

$$|\mathscr{W}_{n,\varepsilon}| \leqslant 2^{n(H(X,Y)+\varepsilon)}.$$

Лемма 5.3. *Если сл. векторы $\widetilde{X}^n$ и $\widetilde{Y}^n$ независимы и распределены соответственно так же, как векторы X^n и Y^n, то*

$$\mathsf{P}((\widetilde{X}^n, \widetilde{Y}^n) \in \mathscr{W}_{n,\varepsilon}) \leqslant 2^{-n(I(X;Y)-3\varepsilon)}.$$

Кроме того, при всех достаточно больших n

$$\mathsf{P}((\widetilde{X}^n, \widetilde{Y}^n) \in \mathscr{W}_{n,\varepsilon}) \geqslant (1-\varepsilon)2^{-n(I(X;Y)+3\varepsilon)}.$$

Доказательство леммы 5.1. Обозначим через $\mathscr{X}_{n,\varepsilon}$ и $\mathscr{Y}_{n,\varepsilon}$ множества всех ε-типичных векторов x^n и y^n соответственно, а через $\mathscr{Z}_{n,\varepsilon}$ — множество всех пар (x^n, y^n), для которых выполнено свойство (5.12). Тогда множество типичных пар представится в виде

$$\mathscr{W}_{n,\varepsilon} = (\mathscr{X}_{n,\varepsilon} \times \mathscr{Y}_{n,\varepsilon}) \cap \mathscr{Z}_{n,\varepsilon}.$$

К каждому из векторов X^n и Y^n и к паре (X^n, Y^n) применима теорема 2.3 (с. 60) об асимптотической равнораспределённости для

дискретного источника без памяти, поэтому найдутся такие числа n_1, n_2, n_3, что

$$\begin{aligned}
&\text{при } n > n_1 \qquad \mathsf{P}(X^n \notin \mathscr{X}_{n,\varepsilon}) < \varepsilon/3,\\
&\text{при } n > n_2 \qquad \mathsf{P}(Y^n \notin \mathscr{Y}_{n,\varepsilon}) < \varepsilon/3,\\
&\text{при } n > n_1 \qquad \mathsf{P}((X^n, Y^n) \notin \mathscr{Z}_{n,\varepsilon}) < \varepsilon/3.
\end{aligned}$$

Поэтому при $n > \max(n_1, n_2, n_3)$

$$\begin{aligned}
\mathsf{P}((X^n, Y^n) \notin \mathscr{W}_{n,\varepsilon}) &= \mathsf{P}((X^n \notin \mathscr{X}_{n,\varepsilon}) \cup (Y^n \notin \mathscr{Y}_{n,\varepsilon}) \cup ((X^n, Y^n) \notin \mathscr{Z}_{n,\varepsilon})) <\\
&< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

[

Доказательство леммы 5.2. Если (x^n, y^n) — ε-типичная пара, то, согласно (5.12) и (5.13),

$$2^{-n(H(X,Y)+\varepsilon)} < p_{X^nY^n}(x^n, y^n) < 2^{-n(H(X,Y)-\varepsilon)}. \tag{5.14}$$

Применяя левое неравенство (5.14), получим

$$\begin{aligned}
|\mathscr{W}_{n,\varepsilon}| &= \sum_{(x^n, y^n) \in \mathscr{W}_{n,\varepsilon}} 1 <\\
&< \sum_{(x^n, y^n) \in \mathscr{W}_{n,\varepsilon}} p_{X^nY^n}(x^n, y^n) 2^{n(H(X,Y)+\varepsilon)} =\\
&= 2^{n(H(X,Y)+\varepsilon)} \mathsf{P}((X^n, Y^n) \in \mathscr{W}_{n,\varepsilon}) \leqslant\\
&\leqslant 2^{n(H(X,Y)+\varepsilon)}.
\end{aligned}$$

□

Доказательство леммы 5.3. Если (x^n, y^n) — ε-типичная пара, то аналогично (5.14) запишем

$$2^{-n(H(X)+\varepsilon)} < p_{X^n}(x^n) < 2^{-n(H(X)-\varepsilon)}, \tag{5.15}$$

$$2^{-n(H(Y)+\varepsilon)} < p_{Y^n}(y^n) < 2^{-n(H(Y)-\varepsilon)}. \tag{5.16}$$

Применяя правые неравенства (5.15) и (5.16) и лемму 5.2, получим

$$\begin{aligned}
\mathsf{P}((\widetilde{X}^n, \widetilde{Y}^n) \in \mathscr{W}_{n,\varepsilon}) &= \sum_{(x^n, y^n) \in \mathscr{W}_{n,\varepsilon}} p_{X^n}(x^n) p_{Y^n}(y^n) <\\
&< \sum_{(x^n, y^n) \in \mathscr{W}_{n,\varepsilon}} 2^{-n(H(X)-\varepsilon)} 2^{-n(H(Y)-\varepsilon)} =\\
&= 2^{-n(H(X)+H(Y)-2\varepsilon)} |\mathscr{W}_{n,\varepsilon}| \leqslant\\
&\leqslant 2^{-n(H(X)+H(Y)-H(X,Y)-3\varepsilon)} =\\
&= 2^{-n(I(X;Y)-3\varepsilon)},
\end{aligned}$$

что доказывает первую часть леммы. Из левых неравенств (5.15) и (5.16) так же, как и выше, получим

$$\mathsf{P}((\tilde{X}^n, \tilde{Y}^n) \in \mathscr{W}_{n,\varepsilon}) \geqslant 2^{-n(H(X)+H(Y)+2\varepsilon)} |\mathscr{W}_{n,\varepsilon}|. \tag{5.17}$$

Применяя правое неравенство (5.14) и лемму 5.1, при достаточно больших n получим

$$\begin{aligned} |\mathscr{W}_{n,\varepsilon}| &= \sum_{(x^n,y^n)\in\mathscr{W}_{n,\varepsilon}} 1 > \\ &> \sum_{(x^n,y^n)\in\mathscr{W}_{n,\varepsilon}} p_{X^nY^n}(x^n, y^n) 2^{n(H(X,Y)-\varepsilon)} = \\ &= 2^{n(H(X,Y)-\varepsilon)} \mathsf{P}((X^n, Y^n) \in \mathscr{W}_{n,\varepsilon}) \geqslant \\ &\geqslant (1-\varepsilon) 2^{n(H(X,Y)-\varepsilon)}. \end{aligned}$$

Подставляя полученную оценку в (5.17), завершаем доказательство. □

Теорема 5.3 (прямая теорема кодирования). *Пусть $(\mathscr{X}, \mathscr{Y}, \pi)$ — канал связи с пропускной способностью $C^* > 0$. Тогда для любых $\alpha > 0$, $\beta > 0$ при достаточно больших n существуют такой код $\mathscr{C}^{(n)} \subseteq \mathscr{X}^n$ и такой декодер, что*

1) *при равномерном распределении $U_{1,\dots,n}$ на $\mathscr{C}^{(n)}$ скорость передачи $R_n = \frac{1}{n}\log_2 |\mathscr{C}^{(n)}|$ удовлетворяет двойному неравенству $C^* - \alpha < R_n < C^*$;*
2) *максимальная вероятность ошибочного декодирования удовлетворяет неравенству $\lambda_{max} < \beta$.*

Замечание 5.9. Теорема означает, что при достаточно большой длине кода можно добиться, чтобы скорость передачи информации была сколь угодно близка к пропускной способности канала связи, не превосходя её, и при этом вероятность ошибочного декодирования была сколь угодно малой.

Доказательство теоремы 5.3. 1. *Основные параметры.* Выберем произвольно число R так, что

$$C^* - \alpha < R < C^*, \tag{5.18}$$

и положим

$$M = M_n = [2^{nR}] + 1.$$

Зафиксируем некоторое распределение $p(x)$ на множестве $\mathscr{X}$. Пусть X — сл. символ на входе канала с распределением $\mathsf{P}_X(x) = p(x)$, а Y — соответствующий сл. символ на выходе канала, т. е. распределение пары (X, Y) имеет вид

$$p_{XY}(x, y) = p(x)\pi(y/x).$$

Далее предполагаем, что в качестве распределения $p(x), x \in \mathscr{X}$, выбрано оптимальное входное распределение, для которого достигается пропускная способность канала связи, т. е.

$$I(X; Y) = C^*. \tag{5.19}$$

Выберем число ε так, что

$$0 < \varepsilon < \min\left\{\frac{C^* - R}{3}, \frac{\beta}{4}\right\}, \tag{5.20}$$

и зафиксируем множество $\mathscr{W}_{n,\varepsilon}$ ε-типичных пар.

2. *Построение множества кодов и декодера*. Предположим, что выбор сообщения, его кодирование, передача и декодирование происходят следующим образом. Кодом здесь будем называть произвольный набор $\mathscr{U}$ векторов $x^n(i) = (x_1(i), \ldots, x_n(i)) \in \mathscr{X}^n$, $1 \leqslant i \leqslant M$. При этом допускается, что некоторые кодовые слова могут совпадать. Это не влияет по существу на результат, но значительно упрощает доказательство.

Источник сообщений выбирает сл. сообщение w, $1 \leqslant w \leqslant M$, в соответствии с равномерным распределением на множестве $\{1, 2, \ldots, M\}$. Выбранному сообщению w *кодер* ставит в соответствие кодовое слово $X^n = x^n(w)$, которое затем передается по каналу связи $(\mathscr{X}, \mathscr{Y}, \pi)$. К полученному на выходе канала вектору Y^n применяется *декодер типичных пар* $\mathscr{D}_\varepsilon$ (см. определение 5.7), в результате получается кодовое слово $x^n(\hat{w})$ или сообщение ϵ об ошибке.

Если при декодировании получено сообщение ϵ об ошибке или получено кодовое слово $x^n(\hat{w})$, но $\hat{w} \neq w$, то говорим, что произошло ошибочное декодирование. Вероятность ошибочного декодирования равна

$$\lambda(\mathscr{U}) = \frac{1}{M}\sum_{i=1}^{M} \lambda_i(\mathscr{U}),$$

где

$$\lambda_i(\mathscr{U}) = \mathsf{P}(\mathscr{D}_\varepsilon(Y^n) \neq X^n / X^n = x^n(i))$$

есть вероятность ошибочного декодирования при условии, что отправлено сообщение $w = i$.

В соответствии с выбранным распределением $p(x), x \in \mathscr{X}$, кодовому слову $x^n(i) = (x_1(i), \ldots, x_n(i))$ припишем вероятность

$$Q(x^n(i)) = \prod_{j=1}^{n} p(x_j(i)),$$

а коду $\mathscr{U}$ припишем вероятность

$$Q(\mathscr{U}) = \prod_{i=1}^{M} Q(x^n(i)).$$

Будем рассматривать множество $\{\mathscr{U}\}$ всех q^{Mn} возможных кодов $\mathscr{U}$ с заданным на этом множестве распределением $Q(\mathscr{U})$. Другими словами, мы полагаем, что последовательность координат кодовых слов случайного кода $\mathscr{U}$ порождается дискретным источником без памяти с алфавитом $\mathscr{X}$ и заданным на нём распределением $p(x)$.

3. *Оценка средней вероятности ошибочного декодирования.* Средняя по множеству кодов вероятность ошибочного декодирования равна

$$\bar{\lambda} = \sum_{\mathscr{U}} Q(\mathscr{U}) \lambda(\mathscr{U}), \tag{5.21}$$

где сумма распространяется на множество всех q^{Mn} возможных кодов $\mathscr{U}$. По-другому это можно записать в виде

$$\bar{\lambda} = \frac{1}{M} \sum_{i=1}^{M} \bar{\lambda}_i,$$

где

$$\bar{\lambda}_i = \sum_{\mathscr{U}} Q(\mathscr{U}) \lambda_i(\mathscr{U})$$

есть средняя по множеству кодов вероятность ошибочного декодирования при условии, что отправлено сообщение $w = i$.

Рассмотрим случай $w = 1$. В канал связи отправлено кодовое слово $X^n(1)$, и на выходе канала получен сл. вектор Y^n. В соответствии с заданным распределением на множестве кодов, кодовые слова $X^n(1)$, $X^n(2)$, …, $X^n(M)$ рассматриваем как независимые и одинаково распределенные сл. векторы; кроме того, сл. вектор Y^n (выход

канала при входе $X^n(1)$) и кодовые слова $X^n(2), \ldots, X^n(M)$ независимы.

Введём событие

$$\mathscr{E}_i = \{(X^n(i), Y^n) \in \mathscr{W}_{n,\varepsilon}\}, 1 \leqslant i \leqslant M,$$

состоящее в том, что кодовое слово $X^n(i)$ вместе с вектором Y^n образует ε-типичную пару.

Из определения декодера $\mathscr{D}_\varepsilon$ ясно, что при условии $w = 1$ ошибочное декодирование происходит в случае выполнения хотя бы одного из событий: события $\overline{\mathscr{E}}_1$ (отправленное кодовое слово $X^n(1)$ и принятый вектор Y^n не образуют ε-типичную пару) или события $\mathscr{E}_2 \cup \ldots \cup \mathscr{E}_M$ (принятый вектор Y^n образует типичную пару с одним из других кодовых слов). Поэтому

$$\overline{\lambda}_1 = \mathsf{P}(\overline{\mathscr{E}}_1 \cup \mathscr{E}_2 \cup \ldots \cup \mathscr{E}_M) \leqslant \mathsf{P}(\overline{\mathscr{E}}_1) + \sum_{i=2}^{M} \mathsf{P}(\mathscr{E}_i).$$

Поскольку $\mathscr{E}_1 = \{(X^n(1), Y^n) \in \mathscr{W}_{n,\varepsilon}\}$, в силу леммы 5.1 при всех достаточно больших n имеем

$$\mathsf{P}(\overline{\mathscr{E}}_1) \leqslant \varepsilon.$$

Далее, в силу независимости Y^n с каждым из $X^n(2), \ldots, X^n(M)$, по лемме 5.3 для $2 \leqslant i \leqslant M$ имеем

$$\mathsf{P}(\mathscr{E}_i) \leqslant 2^{-n(I(X;Y)-3\varepsilon)}.$$

Следовательно,

$$\begin{aligned}\overline{\lambda}_1 &\leqslant \varepsilon + (M-1)2^{-n(I(X;Y)-3\varepsilon)} \leqslant \\ &\leqslant \varepsilon + 2^{-n(I(X;Y)-R-3\varepsilon)}.\end{aligned}$$

Полученная оценка не зависит от предположения $w = 1$ и может быть выведена аналогично при любом $w = i$, $1 \leqslant i \leqslant M$. Поэтому

$$\overline{\lambda} = \frac{1}{M}\sum_{i=1}^{M} \overline{\lambda}_i \leqslant \varepsilon + 2^{-n(I(X;Y)-R-3\varepsilon)}.$$

В полученной выше оценке множитель при n в показателе степени будет положителен в силу условий (5.20) и (5.19). Действительно,

$$I(X;Y) - R - 3\varepsilon = C^* - R - 3\varepsilon > C^* - R - 3 \cdot \frac{C^* - R}{3} = 0.$$

Таким образом,

$$\overline{\lambda} \leqslant \varepsilon + 2^{-n\delta},$$

где

$$\delta = I(X;Y) - R - 3\varepsilon > 0.$$

Поскольку слагаемое $2^{-n\delta}$ стремится к нулю с ростом n, мы получаем, что при всех достаточно больших n справедливо неравенство

$$\overline{\lambda} \leqslant 2\varepsilon. \tag{5.22}$$

4. *Выбор кода*. Ключевой момент доказательства состоит в возможности выбора кода с нужным условием. Из определения (5.21) и оценки (5.22) следует, что в множестве всех возможных кодов $\mathscr{U}$ найдётся такой код $\mathscr{U}_0$, для которого $\lambda(\mathscr{U}_0) \leqslant 2\varepsilon$. (Это известный *принцип Дирихле*: если среднее арифметическое значение некоторого набора чисел положительно, то в этом наборе существует положительное число.)

Докажем, что код $\mathscr{U}_0$ можно выбрать так, что все его кодовые слова $x^n(i)$ будут различны и все они будут ε-типичными векторами. Для этого вначале докажем, что при достаточно больших n число ε-типичных векторов x^n больше, чем M. Действительно, так как кодовые слова кода $\mathscr{U}_0$ порождаются источником без памяти, выполняется свойство асимптотической равнораспределённости (теорема 2.3, с. 60). Поэтому, согласно утверждению 2.7 (с. 59), при достаточно больших n для мощности множества $\mathscr{X}_{n,\varepsilon}$ ε-типичных векторов x^n справедлива нижняя оценка

$$|\mathscr{X}_{n,\varepsilon}| > (1-\varepsilon)2^{n(H(X)-\varepsilon)}.$$

С другой стороны,

$$M = [2^{nR}] + 1 \leqslant 2^{nR} + 1.$$

Поэтому для доказательства неравенства $|\mathscr{X}_{n,\varepsilon}| > M$ (при всех n, начиная с некоторого) достаточно установить, что

$$H(X) - \varepsilon > R.$$

Это неравенство, в свою очередь, вытекает из неравенства $H(X) \geqslant C^*$ и условий (5.18) и (5.20).

Далее, допустим, что в коде $\mathscr{U}_0$ кодовое слово $x^n(i)$ не является ε-типичным вектором. Тогда, по определению декодера ε-типичных

пар, при посылке в канал кодового слова $x^n(i)$ декодер независимо от принятого вектора y^n всегда будет выдавать сообщение об ошибке, другими словами, $\lambda_i(\mathscr{U}_0) = 1$. Заменим кодовое слово $x^n(i)$ на произвольный ε-типичный вектор из $\mathscr{X}_{n,\varepsilon} \setminus \mathscr{U}_0$ и получим новый код, в котором $\lambda_i(\mathscr{U}_0)$ (а поэтому и $\lambda(\mathscr{U})$) могут только уменьшиться.

Наконец, допустим, что в коде $\mathscr{U}_0$ совпадают два каких-то кодовых слова, $x^n(i) = x^n(j)$, $i \neq j$. В этом случае, также по определению декодера ε-типичных пар, $\lambda_i(\mathscr{U}_0) = \lambda_j(\mathscr{U}_0) = 1$. Так же как и выше, заменим одно из кодовых слов, скажем $x^n(i)$, на произвольный ε-типичный вектор из $\mathscr{X}_{n,\varepsilon} \setminus \mathscr{U}_0$ и получим новый код, в котором $\lambda(\mathscr{U})$ может только уменьшиться.

Итак, зафиксируем код $\mathscr{U}_0$, состоящий из M различных ε-типичных кодовых слов $x^n(i)$, $1 \leqslant i \leqslant M$. Не ограничивая общности, будем считать, что эти кодовые слова упорядочены по невозрастанию условной вероятности ошибки декодирования, т. е. так, что

$$\lambda_1(\mathscr{U}_0) \leqslant \lambda_2(\mathscr{U}_0) \leqslant \ldots \leqslant \lambda_M(\mathscr{U}_0).$$

Рассмотрим новый код $\mathscr{C}^{(n)}$, состоящий из первых $[M/2]$ кодовых слов кода $\mathscr{U}_0$. Покажем (от противного), что для кода $\mathscr{C}^{(n)}$ выполнено неравенство

$$\lambda_{\max}(\mathscr{C}^{(n)}) = \lambda_{[M/2]}(\mathscr{U}_0) \leqslant 4\varepsilon.$$

Действительно, если $\lambda_{[M/2]}(\mathscr{U}_0) > 4\varepsilon$, то

$$\begin{aligned}
\lambda(\mathscr{U}_0) = \frac{1}{M}\sum_{i=1}^{M} \lambda_i(\mathscr{U}_0) &\geqslant \\
&\geqslant \frac{1}{M}\sum_{i=[M/2]+1}^{M} \lambda_i(\mathscr{U}_0) > \\
&> 4\varepsilon(M - [M/2])/M \geqslant \\
&\geqslant 2\varepsilon,
\end{aligned}$$

т. е. приходим к противоречию с условием выбора кода $\mathscr{U}_0$.

Таким образом, для кода $\mathscr{C}^{(n)}$ максимальная вероятность ошибочного декодирования, в силу (5.20), удовлетворяет условию

$$\lambda_{\max}(\mathscr{C}^{(n)}) \leqslant 4\varepsilon < \beta,$$

а скорость передачи

$$R_n = \frac{1}{n}\log_2\left[\frac{M}{2}\right] = \frac{1}{n}\log_2\left[\frac{[2^{nR}]+1}{2}\right]$$

лежит в границах (так как $a-1<[a]\leqslant a$)

$$R - \frac{1}{n} + \frac{1}{n}\log_2(1-2^{1-nR}) < R_n \leqslant R - \frac{1}{n} + \frac{1}{n}\log_2(1+2^{-nR}),$$

и поэтому при достаточно больших n, в силу (5.18), удовлетворяет условию

$$C^* - \alpha < R_n < C^*,$$

что и требовалось доказать. □

Замечание 5.10. Отметим, что в ключевом моменте (выборе кода с нужным условием из множества всех возможных кодов), а потому и в целом, доказательство теоремы неконструктивное — мы доказываем лишь существование нужного кода, но не строим его явно.

Известны различные подходы к доказательству прямой теоремы кодирования, но все опубликованные доказательства являются неконструктивными и так или иначе опираются на одну общую идею, предложенную К. Шенноном и называемую принципом случайного кодирования: если средняя по множеству кодов вероятность ошибочного декодирования мала, то в этом множестве найдется код, для которого вероятность ошибочного декодирования также мала.

Прямая теорема кодирования является одним из важнейших результатов теории информации, поскольку указывает предельные возможности надёжной передачи информации по каналу связи с шумом, но, к сожалению, эта теорема не дает никакого указания на практически приемлемый способ построения кодов и декодеров с нужными свойствами. Эффективное построение кодов и декодеров, свойства которых были бы близки к предельным, описанным в прямой теореме кодирования, остаётся важной открытой проблемой.

§ 5.4. Задачи

5.1. По ДСК(p) передаются равновероятные кодовые слова $x(1) = (000)$ и $x(2) = (111)$. На выходе используется декодер с решающими областями

$A_1 = \{(000), (001), (010), (100)\}$,
$A_2 = \{(110), (101), (011), (111)\}$.

Вычислить вероятности ошибок λ_1, λ_2, λ.

5.2. По ДСК с параметром $p = 0{,}1$ передаются равновероятные кодовые слова $x(1) = (010)$ и $x(2) = (101)$. Вычислить среднюю вероятность ошибочного декодирования λ, если на выходе используется декодер с решающими областями:

1) $A_1 = \{(010), (000), (110), (011)\}$,
   $A_2 = \{(101), (001), (111), (100)\}$;
2) $A_1 = \{(000), (001), (010), (100)\}$,
   $A_2 = \{(110), (101), (011), (111)\}$.

Какой из двух декодеров лучше? Зависит ли выбор лучшего декодера от значения параметра p?

5.3. По ДСК(p) передаются кодовые слова $x(1) = (000)$ и $x(2) = (111)$ с вероятностями α и $1 - \alpha$ соответственно. На выходе используется декодер $\mathscr{D}_{\mathsf{AP}}$ (по методу максимальной апостериорной вероятности). Записать неравенство Фано для этого случая; проверить численно при $p = 0{,}1$, $\alpha = 0{,}5$.

5.4. По ДСК с параметром $p < 0{,}5$ передаются кодовые слова $x^n(1) = (00\ldots0)$ и $x^n(2) = (11\ldots1)$ длины n, с вероятностями α и $1 - \alpha$ соответственно. На выходе используется декодер $\mathscr{D}_{\mathsf{AP}}$. Вычислить среднюю вероятность ошибочного декодирования $\lambda^{(n)}$. Доказать, что $\lambda^{(n)} < p$ при $n > 1$ и $\lim_{n\to\infty} \lambda^{(n)} = 0$. При $p = 0{,}1$ найти наименьшее n, для которого $\lambda^{(n)} \leqslant 10^{-k}$, $k = 2, \ldots, 6$.

5.5. Задать таблично декодер $\mathscr{D}_{\mathsf{L}}$ (по методу максимального правдоподобия), если кодовые слова кода $\mathscr{C} = \{(0000), (0011), (1100), (1111)\}$ передаются:

а) по ДСК с параметром $p < 0{,}5$;
б) по двоичному каналу со стиранием.

5.6. Задать таблично декодер $\mathscr{D}_{\mathsf{AP}}$, если равновероятные кодовые слова кода $\mathscr{C} = \{(00000), (01101), (10111), (11010)\}$ передаются по ДСК с параметром $p < 0{,}5$.

5.7. По ДСК с параметром $p = 0{,}1$ передаются четыре кодовых слова (00000), (01111), (10101), (11010) с вероятностями 0,8, 0,1, 0,05 и 0,05 соответственно. Указать решающие области для декодеров $\mathscr{D}_{\mathsf{L}}$ и $\mathscr{D}_{\mathsf{AP}}$.

5.8. По ДСК(p) передаются восемь равновероятных кодовых слов (000000), (001101), (010110), (011011), (100111), (101010), (110001), (111100). На выходе используется декодер $\mathscr{D}_{\mathsf{L}}$. Вычислить вероятность ошибочного декодирования λ как функцию от параметра p.

5.9. Канал связи задан входным алфавитом $\mathscr{X} = \{x_1, x_2, x_3\}$, выходным алфавитом $\mathscr{Y} = \{y_1, y_2, y_3\}$ и матрицей

$$\pi = \begin{pmatrix} 0{,}5 & 0{,}3 & 0{,}2 \\ 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}1 & 0{,}9 & 0 \end{pmatrix}.$$

Входные символы канала равновероятны. Декодер определяется условием $\mathscr{D}(y_i) = x_i$, $i = 1, 2, 3$. Вычислить вероятности ошибок λ_1, λ_2, λ_3, λ. Проверить выполнение неравенства Фано.

5.10. Канал связи задан входным алфавитом $\mathscr{X} = \{x_1, x_2, x_3\}$, выходным алфавитом $\mathscr{Y} = \{y_1, y_2, y_3\}$ и матрицей

$$\pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{pmatrix}.$$

Входные символы канала имеют вероятности $\frac{1}{2}$, $\frac{1}{4}$ и $\frac{1}{4}$ соответственно. На выходе используется декодер $\mathscr{D}_{\mathsf{AP}}$. Вычислить вероятности ошибок λ_1, λ_2, λ_3, λ. Проверить выполнение неравенства Фано.

Глава 6
Коды, исправляющие ошибки

§ 6.1. Блоковые коды

В этой главе мы рассмотрим методы кодирования и декодирования для надёжной передачи сообщений по каналам связи с шумом. Чтобы понять возникающую здесь задачу, вернёмся к общей схеме передачи информации, описанной в предисловии.

Предположим, что источник сообщений порождает некоторую последовательность $a_{i_1} a_{i_2} \ldots$ символов алфавита $\mathscr{A} = \{a_1, \ldots, a_m\}$. Кодер источника преобразует эту последовательность при помощи алфавитного кодирования φ и выдаёт последовательность $\varphi(a_{i_1})\varphi(a_{i_2}) \ldots = b_{j_1} b_{j_2} \ldots$ символов алфавита $\mathscr{B} = \{b_1, \ldots, b_D\}$. Далее последовательность $b_{j_1} b_{j_2} \ldots$ должна быть передана по дискретному каналу связи с входным алфавитом $\mathscr{X} = \{x_1, \ldots, x_q\}$ и выходным алфавитом $\mathscr{Y} = \{y_1, \ldots, y_s\}$. Для этого последовательность $b_{j_1} b_{j_2} \ldots$ необходимо как-то преобразовать (перекодировать) в последовательность символов алфавита $\mathscr{X}$. Это преобразование (кодирование для канала связи) желательно выбрать так, чтобы выполнить несколько важных требований. Во-первых, мы стремимся достичь возможно более высокой скорости передачи, приближающейся к пропускной способности канала связи. Во-вторых, необходимо на выходе канала связи построить декодер с наименьшей возможной вероятностью ошибочного декодирования. Кроме того, с практической точки зрения важно, чтобы кодирование и декодирование можно было реализовать при помощи алгоритмов, эффективных в смысле трудоёмкости и требуемой памяти.

Известны различные методы кодирования, предназначенные для передачи сообщений по разным видам каналов связи. Метод кодирования выбирают с учётом особенностей помех, воздействующих на передаваемое в канале связи сообщение. Например, в реальных каналах связи оказывается важным соотношение между временем передачи одного символа и средним временем воздействия случайной помехи. Если время воздействия помехи не больше времени передачи одного символа, то с некоторым приближением можно считать, что

последовательная передача символов по каналу происходит независимым образом, и в этом случае нужно такое кодирование, которое позволяло бы обнаруживать и исправлять ошибки приёма отдельных символов. Если время воздействия помехи больше времени передачи одного символа, то помеха воздействует на несколько передаваемых символов, и тогда нужно такое кодирование, которое позволяло бы корректировать так называемые пакеты ошибок. Важно также, как именно воздействует помеха на передаваемый по каналу символ: например, различают ошибки замены, вставки или стирания символов. Методы кодирования различают также по тому, как зависит кодирование очередного символа или блока символов от результатов кодирования предшествующих символов: возможно, что такой зависимости нет, или имеется зависимость от одного, нескольких или всех ранее закодированных символов или блоков символов.

Из всего большого разнообразия методов современной теории кодов, исправляющих ошибки, мы рассмотрим только один класс так называемых блоковых кодов, выделив в нём подкласс линейных блоковых кодов и далее подкласс циклических кодов. Такие коды предназначены для обнаружения и исправления ошибок при передаче символов по дискретным каналам связи без памяти, рассмотренным в главах 4 и 5.

Определение 6.1. Пусть натуральные числа k и n таковы, что $D^k \leqslant q^n$. *Блоковым кодированием* с входным алфавитом $\mathscr{B} = \{b_1, \ldots, b_D\}$, выходным алфавитом $\mathscr{X} = \{x_1, \ldots, x_q\}$, блоком информационных символов длины k и блоком кодовых символов длины n называется произвольное инъективное отображение $f : \mathscr{B}^k \to \mathscr{X}^n$.

Множество $\mathscr{C} = f(\mathscr{B}^k) \subseteq \mathscr{X}^n$ называется блоковым кодом, а его элементы — кодовыми словами.

Замечание 6.1. Предполагается, что введённое выше блоковое кодирование распространяется на множество $\mathscr{B}^*$ всех конечных последовательностей в алфавите $\mathscr{B}$ следующим образом. Если входная последовательность имеет длину, кратную k, то она разбивается на непересекающиеся блоки длины k, и каждый такой блок заменяется на блок длины n в алфавите $\mathscr{X}$ в соответствии с кодированием f независимо от других блоков. Если же длина входной последовательности не кратна k, то последний неполный блок дописывает-

ся до длины k некоторым заранее оговорённым способом, и возвращаемся к предыдущему случаю.

Кодовые слова кода $\mathscr{C} \subseteq \mathscr{X}^n$ предназначены для передачи по дискретному каналу связи без памяти $(\mathscr{X}, \mathscr{Y}, \pi)$ (далее коротко — каналу связи), при этом n символов кодового слова передаются по каналу связи последовательно, один за другим. Если на вход блокового кодера поступает информационный блок $b^k = b_{j_1} b_{j_2} \dots b_{j_k} \in \mathscr{B}^k$, то полученное кодовое слово $x^n = f(b^k) \in \mathscr{C}$ поступает в канал связи, и на выходе канала появляется некоторый вектор $y^n \in \mathscr{Y}^n$.

Далее вектор y^n поступает на вход некоторого декодера общего вида $\mathscr{D}$ (см. определение 5.2, с. 137). Если декодер $\mathscr{D}$ принимает решение, что на входе канала связи было кодовое слово $\tilde{x}^n \in \mathscr{C}$, то к этому кодовому слову применяется обратное преобразование f^{-1}, и в итоге получается информационный блок $\tilde{b}^k = f^{-1}(\tilde{x}^n)$, который может и не совпадать с исходным блоком b^k. Если же декодер $\mathscr{D}$ выдаёт сообщение об ошибке типа «декодер не может принять решение», то делаем вывод о безуспешной передаче; в реальных каналах связи в этом случае может последовать запрос на повторную передачу информационного блока.

Сложная задача состоит в том, чтобы применительно к заданному каналу связи $(\mathscr{X}, \mathscr{Y}, \pi)$ построить блоковое кодирование f и декодер $\mathscr{D}$, позволяющие свести к минимуму вероятность ошибочного декодирования. Кроме того, с точки зрения практического применения важно, чтобы кодирование f и декодер $\mathscr{D}$ допускали простую и надёжную реализацию аппаратными или программными средствами и имели высокое быстродействие.

Ранее в главе 5 были рассмотрены два декодера — декодер по методу максимального правдоподобия и декодер по методу максимальной апостериорной вероятности. Здесь мы введём ещё один важный алгоритм декодирования, основанный на расстоянии Хэмминга.

Определение 6.2. Расстоянием Хэмминга на множестве $\mathscr{X}^n$ называется отображение $\rho : \mathscr{X}^n \times \mathscr{X}^n \to \mathbb{N}_0$, задаваемое формулой

$$\rho(a^n, b^n) = \sum_{i=1}^{n} \chi(a_i \neq b_i),$$

где $a^n = (a_1, a_2, \ldots, a_n) \in \mathscr{X}^n$, $b^n = (b_1, b_2, \ldots, b_n) \in \mathscr{X}^n$ и $\chi(A)$ — индикатор события A. Другими словами, $\rho(a^n, b^n)$ есть число координат, в которых векторы a^n и b^n не совпадают.

Утверждение 6.1. *Для любых $a^n, b^n, c^n \in \mathscr{X}^n$ расстояние Хэмминга обладает следующими свойствами:*

1) (*неотрицательность*) $\rho(a^n, b^n) \geqslant 0$, *причём равенство имеет место тогда и только тогда, когда* $a^n = b^n$;
2) (*симметричность*) $\rho(a^n, b^n) = \rho(b^n, a^n)$;
3) (*неравенство треугольника*) $\rho(a^n, b^n) + \rho(b^n, c^n) \geqslant \rho(a^n, c^n)$.

Доказательство. Свойства 1 и 2 очевидны из определения. Свойство 3 вытекает из неравенства $\chi(a_i \neq b_i) + \chi(b_i \neq c_i) \geqslant \chi(a_i \neq c_i)$, проверяемого путём простого перебора возможных случаев. □

Далее в этой главе, как правило, будем предполагать, что $\mathscr{X} = \mathscr{Y}$, т. е. входной и выходной алфавиты канала связи совпадают, и возможные ошибки в канале сводятся к заменам символов; исключения будут особо оговариваться. Более того, в большинстве случаев будем считать, что $\mathscr{B} = \mathscr{X} = \mathscr{Y}$. Наиболее распространённым случаем является использование двоичного алфавита $\mathscr{B} = \mathscr{X} = \mathscr{Y} = \{0, 1\}$.

В предположении $\mathscr{X} = \mathscr{Y}$ уточним наше понимание терминов *обнаружение* и *исправление* ошибок. Если в канал связи было передано кодовое слово x^n, из канала принят вектор y^n и $\rho(y^n, x^n) = s$, то говорим, что произошло s ошибок, а $n - s$ символов передано правильно. *Обнаружение* ошибок при передаче происходит, если принятый из канала вектор y^n не является кодовым словом кода $\mathscr{C}$. Далее, когда декодер вычисляет кодовое слово $\tilde{x}^n = \mathscr{D}(y^n)$, говорим, что происходит *исправление* ошибок; если при этом оказывается, что результат декодирования $\tilde{x}^n$ совпадает с отправленным в канал кодовым словом x^n, то говорим, что декодер правильно исправил ошибки, в противном случае — декодер ошибся.

Декодер $\mathscr{D}_\rho$ — в ближайшее кодовое слово.

Из утверждения 6.1 ясно, что расстояние Хэмминга действительно является метрикой на пространстве $\mathscr{X}^n$. Это позволяет нам ввести ещё один способ построения декодера — на основе принципа минимального расстояния, или, другими словами, *декодирование*

в ближайшее кодовое слово. Алгоритм декодера в ближайшее кодовое слово состоит в том, что если принят вектор y^n, то вычисляются расстояния $\rho(y^n, x^n)$ для всех кодовых слов $x^n \in \mathscr{C}$, и в качестве результата декодирования выбирается то кодовое слово, для которого это расстояние минимально. Если минимальное расстояние достигается для нескольких кодовых слов, то выбирается одно из них по некоторому заранее оговорённому правилу, например наименьшее в смысле лексикографического порядка. Декодер в ближайшее кодовое слово будем обозначать $\mathscr{D}_\rho$.

Оказывается, в одном важном частном случае декодеры $\mathscr{D}_{\mathsf{L}}$ (по методу максимального правдоподобия) и $\mathscr{D}_\rho$ (в ближайшее кодовое слово) эквивалентны, а именно при использовании двоичного симметричного канала с параметром p (пример 4.1, с. 111), коротко — ДСК(p).

Утверждение 6.2. *Пусть* $\mathscr{X} = \mathscr{Y} = \{0, 1\}$, *канал связи* $(\mathscr{X}, \mathscr{Y}, \pi)$ *представляет собой ДСК*(p), $0 < p \leqslant \frac{1}{2}$. *Тогда декодеры* $\mathscr{D}_{\mathsf{L}}$ *и* $\mathscr{D}_\rho$ *эквивалентны.*

Доказательство. По определению ДСК(p),

$$\pi^{(n)}(y^n/x^n) = \prod_{t=1}^{n} \pi(y_{j_t}/x_{i_t}) = (1-p)^{n-s}p^s, \tag{6.1}$$

где $s = \rho(y^n, x^n)$. Если $\tilde{x}^n$ — другое кодовое слово и $\tilde{s} = \rho(y^n, \tilde{x}^n)$, то из (6.1) получаем

$$\frac{\pi^{(n)}(y^n/\tilde{x}^n)}{\pi^{(n)}(y^n/x^n)} = \frac{(1-p)^{n-\tilde{s}}p^{\tilde{s}}}{(1-p)^{n-s}p^s} = \left(\frac{p}{1-p}\right)^{\tilde{s}-s}.$$

Так как $0 < p/(1-p) \leqslant 1$, неравенство

$$\pi^{(n)}(y^n/\tilde{x}^n) \geqslant \pi^{(n)}(y^n/x^n)$$

равносильно неравенству $\tilde{s} - s \leqslant 0$, или $\tilde{s} \leqslant s$. Это означает, что из двух кодовых слов x^n и $\tilde{x}^n$ декодер $\mathscr{D}_{\mathsf{L}}$ выбирает $\tilde{x}^n$ тогда и только тогда, когда это же верно и для декодера $\mathscr{D}_\rho$. □

Замечание 6.2. Декодеры $\mathscr{D}_{\mathsf{L}}$ и $\mathscr{D}_\rho$, вообще говоря, не эквивалентны. Например, пусть $q = 3$, $\mathscr{X} = \mathscr{Y} = \{0, 1, 2\}$ и канал связи задаётся

матрицей переходных вероятностей

$$\pi = \begin{pmatrix} p & 1-p & 0 \\ 1-p & p & 0 \\ p & 0 & 1-p \end{pmatrix}.$$

Пусть $n = 3$ и код состоит из двух кодовых слов $x^3 = (011)$ и $\tilde{x}^3 = (002)$. Если $y^3 = (000)$, то $\rho(y^3, x^3) = 2$, $\rho(y^3, \tilde{x}^3) = 1$ и отношение

$$\frac{\pi^{(3)}(y^3/\tilde{x}^3)}{\pi^{(3)}(y^3/x^3)} = \frac{p^3}{p(1-p)^2} = \left(\frac{p}{1-p}\right)^2$$

не меньше 1 тогда и только тогда, когда $p \geqslant 1/2$. Отсюда видно, что если $p < 1/2$, то при $y^3 = (000)$ декодеры $\mathscr{D}_{\mathsf{L}}$ и $\mathscr{D}_{\rho}$ дают разные результаты.

Важным параметром, определяющим способность декодера $\mathscr{D}_{\rho}$ обнаруживать и исправлять ошибки, является минимальное кодовое расстояние.

Определение 6.3. Минимальным кодовым расстоянием блокового кода $\mathscr{C} \subseteq \mathscr{X}^n$, $|\mathscr{C}| \geqslant 2$, называется величина

$$d_{\mathscr{C}} = \min_{\substack{a^n, b^n \in \mathscr{C} \\ a^n \neq b^n}} \rho(a^n, b^n).$$

Утверждение 6.3. *Декодер $\mathscr{D}_{\rho}$ обнаруживает любые комбинации из d или меньшего числа ошибок тогда и только тогда, когда $d_{\mathscr{C}} \geqslant d+1$.*

Доказательство. Предположим, что декодер обнаруживает любые комбинации из d или меньшего числа ошибок, и докажем неравенство $d_{\mathscr{C}} \geqslant d+1$ от противного. Если $d_{\mathscr{C}} \leqslant d$, то существуют два различных кодовых слова $x^n, \tilde{x}^n \in \mathscr{C}$, для которых $\rho(x^n, \tilde{x}^n) = s \leqslant d$. Если было передано кодовое слово x^n, а на выходе канала получен вектор $y^n = \tilde{x}^n$, то произошло $s \leqslant d$ ошибок, но декодер не обнаружил их.

Обратно, пусть $d_{\mathscr{C}} \geqslant d+1$. Предположим, что при передаче кодового слова x^n по каналу связи произошла хотя бы одна, но не более d ошибок, т. е. расстояние Хэмминга между переданным кодовым словом x^n и полученным вектором y^n удовлетворяет неравенствам $1 \leqslant \rho(x^n, y^n) \leqslant d$. Отсюда следует, что вектор y^n не является кодовым словом, и поэтому декодер обнаруживает ошибки. □

Утверждение 6.4. *Декодер $\mathscr{D}_\rho$ правильно исправляет любые комбинации из t или меньшего числа ошибок тогда и только тогда, когда $d_{\mathscr{C}} \geqslant 2t+1$.*

Доказательство. Предположим, что $d_{\mathscr{C}} \geqslant 2t+1$, и докажем, что декодер правильно исправляет любые комбинации из t или меньшего числа ошибок. Пусть передано кодовое слово x^n, на выходе канала получен вектор y^n и $\rho(x^n, y^n) \leqslant t$. Для любого кодового слова $\tilde{x}^n \neq x^n$ из определения минимального кодового расстояния и из неравенства треугольника получаем

$$2t+1 \leqslant \rho(x^n, \tilde{x}^n) \leqslant \rho(x^n, y^n) + \rho(y^n, \tilde{x}^n) \leqslant t + \rho(y^n, \tilde{x}^n),$$

откуда следует, что $\rho(y^n, \tilde{x}^n) \geqslant t+1$.

Таким образом, $\rho(y^n, \tilde{x}^n) > \rho(x^n, y^n)$, и поэтому из двух возможных вариантов декодирования x^n и $\tilde{x}^n$ декодер предпочтёт правильный вариант x^n, и тем самым правильно исправит ошибки.

Обратно, пусть декодер правильно исправляет любые комбинации из $t \geqslant 1$ или меньшего числа ошибок. Докажем неравенство $d_{\mathscr{C}} \geqslant 2t+1$ от противного. Пусть $d_{\mathscr{C}} = s \leqslant 2t$. Случай $s = 1$ тривиален. Если $s \geqslant 2$, то рассмотрим два ближайших друг к другу кодовых слова $x^n \neq \tilde{x}^n$ таких, что $\rho(x^n, \tilde{x}^n) = s$. Не ограничивая общности, можно предположить, что кодовые слова x^n и $\tilde{x}^n$ различаются в первых s координатах, т. е.

$$x^n = (a_1, \ldots, a_s, a_{s+1}, \ldots, a_n), \tilde{x}^n = (a'_1, \ldots, a'_s, a_{s+1}, \ldots, a_n),$$

где $a_i \neq a'_i$, $1 \leqslant i \leqslant s$. Обозначим через r наименьшее целое число с условием $s - r \leqslant r$ (при этом $1 \leqslant r \leqslant t$) и рассмотрим вектор

$$y^n = (a'_1, \ldots, a'_r, a_{r+1}, \ldots, a_s, \ldots, a_n).$$

Тогда $\rho(x^n, y^n) = r$, $\rho(y^n, \tilde{x}^n) = s - r$ и y^n не является кодовым словом. Предположим, что передано кодовое слово x^n и получен вектор y^n.

Если s нечётно, $s = 2r - 1$, то $\rho(y^n, \tilde{x}^n) = r - 1 < r = \rho(x^n, y^n)$. Из двух возможных вариантов декодирования x^n и $\tilde{x}^n$ декодер предпочтёт неправильный вариант $\tilde{x}^n$, и тем самым неправильно исправит $r \leqslant t$ ошибок, что противоречит условию.

Если s чётно, $s = 2r$, то $\rho(y^n, \tilde{x}^n) = \rho(x^n, y^n) = r$. Из двух равноправных вариантов декодирования x^n и $\tilde{x}^n$ декодер *может* пред-

почесть неправильный вариант $\tilde{x}^n$, если так предписывает правило разрешения неопределённых ситуаций, и тем самым может неправильно исправить $r \leqslant t$ ошибок, что также противоречит условию. □

Из утверждений 6.3 и 6.4 вытекает

Следствие 6.1. *Если $d_{\mathscr{C}} = d$, то декодер $\mathscr{D}_\rho$ обнаруживает любые комбинации из $d-1$ или меньшего числа ошибок и правильно исправляет любые комбинации из $[(d-1)/2]$ или меньшего числа ошибок.*

В завершение этого параграфа рассмотрим один случай, когда входной алфавит $\mathscr{X}$ и выходной алфавит $\mathscr{Y}$ канала связи не совпадают. Предположим, что алфавит $\mathscr{Y}$ состоит из символов алфавита $\mathscr{X} = \{x_1, \ldots, x_q\}$ и ещё одного дополнительного символа E, называемого символом стирания. Если на входе канала был символ x_i, а на выходе появился символ $x_j \neq x_i$, то говорим, что произошло *искажение* переданного символа; если же на выходе появился символ E, то говорим, что произошло *стирание* переданного символа. Примером такого канала может служить двоичный канал со стиранием (с. 111). На множестве $\mathscr{Y}^n$ введём новую метрику, называемую обобщённым расстоянием Хэмминга.

Определение 6.4. Обобщённым расстоянием Хэмминга на множестве $\mathscr{Y}^n$, $\mathscr{Y} = \{x_1, \ldots, x_q, E\}$, называется отображение $\hat{\rho} : \mathscr{Y}^n \times \mathscr{Y}^n \to \to \mathbb{R}_+$, задаваемое формулой

$$\hat{\rho}(a^n, b^n) = \sum_{i=1}^{n} \hat{\chi}(a_i, b_i),$$

где $a^n = (a_1, a_2, \ldots, a_n) \in \mathscr{Y}^n$, $b^n = (b_1, b_2, \ldots, b_n) \in \mathscr{Y}^n$ и

$$\hat{\chi}(a_i, b_i) = \begin{cases} 0, & \text{если } a_i = b_i, \\ 1, & \text{если } a_i \neq b_i, a_i \neq E, b_i \neq E, \\ \dfrac{1}{2}, & \text{если } a_i \neq b_i \text{ и либо } a_i = E, \text{ либо } b_i = E. \end{cases}$$

Замечание 6.3. Точно так же как в утверждении 6.1, можно доказать, что обобщённое расстояние Хэмминга неотрицательно, симметрично, и для него выполняется неравенство треугольника, т. е.

мы действительно имеем метрику на множестве $\mathscr{Y}^n$. Кроме того, нетрудно видеть, что мы можем обобщить определение декодера $\mathscr{D}_\rho$ на случай $\mathscr{Y} = \mathscr{X} \cup \{E\}$, заменив метрику ρ на $\hat{\rho}$.

Утверждение 6.5. *Если $\mathscr{Y} = \mathscr{X} \cup \{E\}$, $d_{\mathscr{C}} = d$, и $t_0 + 2t_1 \leqslant d - 1$, то обобщённый декодер $\mathscr{D}_{\hat{\rho}}$ правильно исправляет любые комбинации из t_0 стираний и t_1 искажений символов.*

Доказательство. Предположим, что при передаче кодового слова x^n в канале связи произошло t_0 стираний и t_1 искажений символов, так что для принятого вектора y^n имеем $\hat{\rho}(x^n, y^n) = t_0/2 + t_1 \leqslant \leqslant (d-1)/2$.

Докажем, что для любого другого кодового слова $\tilde{x}^n \neq x^n$ справедливо неравенство $\hat{\rho}(\tilde{x}^n, y^n) > (d-1)/2$. Действительно, если $\hat{\rho}(\tilde{x}^n, y^n) \leqslant (d-1)/2$, то по неравенству треугольника

$$\hat{\rho}(\tilde{x}^n, x^n) \leqslant \hat{\rho}(\tilde{x}^n, y^n) + \hat{\rho}(y^n, x^n) \leqslant d - 1,$$

что противоречит условию $d_{\mathscr{C}} = d$.

Таким образом, расстояние $\hat{\rho}$ от принятого вектора y^n до кодового слова x^n меньше, чем до любого другого кодового слова, поэтому декодер примет решение в пользу x^n и тем самым правильно исправит произошедшие в канале искажения и стирания символов. □

§ 6.2. Линейные коды

Линейные блоковые коды являются важным классом кодов, для которых описание, оценка параметров, алгоритмы кодирования и декодирования основываются на хорошо развитом математическом аппарате линейной алгебры.

Здесь и далее полагаем, что мощность q кодового алфавита $\mathscr{X}$ есть степень простого числа. Алфавит $\mathscr{X} = \{x_1, \ldots, x_q\}$ будем отождествлять с конечным полем $GF(q)$, а множество $\mathscr{X}^n$ — с n-мерным линейным пространством $V_n(q)$ векторов-строк над полем $GF(q)$. Число n будет фиксировано, поэтому верхний индекс n, указывающий на длину вектора, будем опускать. Элементы пространства $V_n(q)$ будем обозначать полужирными буквами, например, $\mathbf{c} = (c_1, \ldots, c_n)$; нулевой вектор обозначаем $\mathbf{0} = (0, \ldots, 0)$.

Определение 6.5. Линейным блоковым q-ичным кодом длины n называется произвольное подпространство $\mathscr{C}$ линейного пространства $V_n(q)$. Если размерность кода $\mathscr{C}$ равна $\dim \mathscr{C} = k$, то говорят, что $\mathscr{C}$ есть линейный (n, k)-код; если, кроме того, $d_{\mathscr{C}} = d$, то код $\mathscr{C}$ называют линейным (n, k, d)-кодом.

Поскольку здесь рассматриваются только линейные блоковые коды и не рассматриваются другие классы линейных кодов (например, свёрточные), слово «блоковые» будем для краткости опускать.

Замечание 6.4. Из определения 6.5 следует, что если $\mathscr{C}$ есть линейный (n, k)-код, то $0 \leqslant k \leqslant n$ и мощность кода равна $M = |\mathscr{C}| = q^k$.

Определение 6.6. Весом Хэмминга $w(\mathbf{c})$ вектора $\mathbf{c} \in V_n(q)$ называется расстояние $\rho(\mathbf{c}, \mathbf{0})$.

Другими словами, вес Хэмминга $w(\mathbf{c})$ равен числу ненулевых координат вектора $\mathbf{c}$. Очевидно также, что $\rho(\mathbf{u}, \mathbf{v}) = w(\mathbf{u} - \mathbf{v})$.

Утверждение 6.6. *Если $\mathscr{C}$ есть линейный код, то*

$$d_{\mathscr{C}} = \min_{\mathbf{c} \in \mathscr{C},\, \mathbf{c} \neq \mathbf{0}} w(\mathbf{c}).$$

Доказательство. Для ненулевого $\mathbf{c} \in \mathscr{C}$ имеем $w(\mathbf{c}) = \rho(\mathbf{c}, \mathbf{0}) \geqslant d_{\mathscr{C}}$. С другой стороны, если для некоторых двух различных кодовых слов $\mathbf{u}, \mathbf{v} \in \mathscr{C}$ достигается наименьшее значение $\rho(\mathbf{u}, \mathbf{v}) = d_{\mathscr{C}}$, то в силу линейности кода имеем $\mathbf{u} - \mathbf{v} \in \mathscr{C}$, и $w(\mathbf{u} - \mathbf{v}) = d_{\mathscr{C}}$. □

Доказанное утверждение непосредственно следует из определений 6.3, 6.5 и 6.6 и позволяет для линейного кода упростить, по сравнению с произвольным блоковым кодом, вычисление минимального кодового расстояния $d_{\mathscr{C}}$. Действительно, чтобы для кода $\mathscr{C}$ мощности $M = |\mathscr{C}|$ вычислить $d_{\mathscr{C}}$ по определению 6.3, необходимо найти все попарные расстояния между различными кодовыми словами, т. е. перебрать $\binom{M}{2}$ вариантов, а для линейного кода достаточно найти веса ненулевых кодовых слов, т. е. перебрать $M - 1$ вариантов.

Определение 6.7. Пусть $\mathscr{C}$ есть линейный (n, k)-код, $k \geqslant 1$, и векторы $\mathbf{g}_1, \ldots, \mathbf{g}_k$ образуют базис кода $\mathscr{C}$, $\mathbf{g}_i = (g_{i1}, \ldots, g_{in})$, $1 \leqslant i \leqslant k$. Тогда матрица $G = (g_{ij})$ размера $k \times n$, строками которой являются векторы $\mathbf{g}_1, \ldots, \mathbf{g}_k$, называется порождающей матрицей кода $\mathscr{C}$.

Замечание 6.5. Каждое кодовое слово $\mathbf{c} \in \mathscr{C}$ единственным образом представляется в виде линейной комбинации базисных векторов:

$$\mathbf{c} = u_1\mathbf{g}_1 + \ldots + u_k\mathbf{g}_k,$$

где $u_1, \ldots, u_k \in GF(q)$, или в матричном виде

$$\mathbf{c} = \mathbf{u} \cdot G, \quad \mathbf{u} = (u_1, \ldots, u_k) \in V_k(q). \tag{6.2}$$

Формулу (6.2) будем рассматривать как реализацию блокового кодирования (см. определение 6.1), которое блоку $\mathbf{u} = (u_1, \ldots, u_k)$ информационных символов ставит в соответствие блок $\mathbf{c} = (c_1, \ldots, c_n)$ кодовых символов. Таким образом, для линейного кода операция кодирования представляет собой просто умножение блока информационных символов на порождающую матрицу кода.

Напомним, что скалярным произведением векторов $\mathbf{u} = (u_1, \ldots, u_n)$ и $\mathbf{v} = (v_1, \ldots, v_n)$ из $V_n(q)$ называется элемент поля $GF(q)$, определяемый формулой $(\mathbf{u}, \mathbf{v}) = u_1v_1 + \ldots + u_nv_n$. Если $(\mathbf{u}, \mathbf{v}) = 0$, то говорят, что векторы $\mathbf{u}$ и $\mathbf{v}$ ортогональны, и обозначают этот факт в виде $\mathbf{u} \perp \mathbf{v}$. Множество $\mathscr{C}^\perp = \{\mathbf{u} \in V_n(q) : \mathbf{u} \perp \mathbf{v}$ для любого $\mathbf{v} \in \mathscr{C}\}$ векторов, ортогональных всем векторам k-мерного линейного подпространства $\mathscr{C}$, само является линейным подпространством пространства $V_n(q)$ и имеет размерность $n - k$. В теории кодирования говорят, что $\mathscr{C}^\perp$ есть $(n, n-k)$-код, дуальный (или двойственный) (n, k)-коду $\mathscr{C}$.

Определение 6.8. Пусть $\mathscr{C}$ есть линейный (n, k)-код, $k < n$, и H — порождающая матрица двойственного $(n, n-k)$-кода $\mathscr{C}^\perp$. Тогда матрица H называется проверочной матрицей кода $\mathscr{C}$.

Замечание 6.6. Очевидно, что проверочная матрица H (n, k)-кода $\mathscr{C}$ имеет размер $(n-k) \times n$, а её $n-k$ строк $\mathbf{h}_1, \ldots, \mathbf{h}_{n-k}$ образуют базис кода $\mathscr{C}^\perp$, $\mathbf{h}_i = (h_{i1}, \ldots, h_{in})$, $1 \leqslant i \leqslant n-k$. Из определения двойственного кода следует, что каждая строка $\mathbf{g}_i$ матрицы G ортогональна всем строкам матрицы H. Поэтому справедливо равенство $G \cdot H^T = O$, где H^T означает транспонирование матрицы H, а O есть нулевая матрица размера $k \times (n-k)$. Если для кода $\mathscr{C}$ матрица G является порождающей, а матрица H — проверочной, то для двойственного кода $\mathscr{C}^\perp$ эти матрицы меняются ролями: G является проверочной матрицей, а H — порождающей.

Термин «проверочная матрица» связан с тем, что при помощи матрицы H легко проверить, является ли заданный вектор элементом кода. Непосредственно из определения 6.8 вытекает следующее

Утверждение 6.7. *Пусть $\mathscr{C}$ есть линейный (n, k)-код, $k < n$, H — проверочная матрица кода $\mathscr{C}$ и $\mathbf{y}$ — произвольный вектор из $V_n(q)$. Условие $\mathbf{y} \in \mathscr{C}$ равносильно равенству $\mathbf{y} \cdot H^T = \mathbf{0}$, где $\mathbf{0}$ — нулевой вектор размера $n - k$.*

В следующих двух примерах рассмотрены двоичные коды, т. е. $q = 2$.

Пример 6.1 (код с повторением). Пусть n нечётно, $n = 2m + 1$, и код $\mathscr{C}$ состоит из двух кодовых слов $\mathbf{0} = (0 \ldots 0)$ и $\mathbf{1} = (1 \ldots 1)$. Очевидно, $k = 1$, $G = (1 \ldots 1)$, а в качестве проверочной можно взять матрицу

$$H = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 & 1 \\ 0 & 1 & \ldots & 0 & 0 & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & 0 & 1 \\ 0 & 0 & \ldots & 0 & 1 & 1 \end{pmatrix},$$

представляющую собой единичную матрицу порядка $n - 1$ с приписанным к ней справа столбцом из единиц. Алгоритм кодирования в данном случае представляет собой просто замену символа 0 на вектор $\mathbf{0} = (0 \ldots 0)$, а символа 1 — на вектор $\mathbf{1} = (1 \ldots 1)$, другими словами, один информационный символ передаётся по каналу связи n раз. Очевидно, что для такого кода $d_{\mathscr{C}} = n = 2m + 1$, и, согласно следствию 6.1, декодер $\mathscr{D}_\rho$ обнаруживает любые комбинации из $n - 1 = 2m$ или меньшего числа ошибок и правильно исправляет любые комбинации из $(n - 1)/2 = m$ или меньшего числа ошибок. Декодер $\mathscr{D}_\rho$ описать очень просто: вычисляем вес Хэмминга $w(\mathbf{u})$ принятого вектора $\mathbf{u} = (u_1, \ldots, u_n)$, и если $w(\mathbf{u}) \leqslant m$, то $\mathscr{D}_\rho(\mathbf{u}) = \mathbf{0}$, в противном случае $\mathscr{D}_\rho(\mathbf{u}) = \mathbf{1}$, т. е. решение принимается в пользу 0, если вес принятого вектора не больше m, и в пользу 1, если этот вес больше m. Такое декодирование называют ещё мажоритарным, или «методом голосования».

Пример 6.2 (код с проверкой на чётность). Пусть код $\mathscr{C}$ состоит из всех векторов длины n, имеющих чётный вес. Нетрудно видеть,

что при сложении двух векторов над $GF(2)$ чётного веса получается вектор, также имеющий чётный вес, так что множество $\mathscr{C}$ действительно является линейным кодом мощности 2^{n-1} и размерности $k = n - 1$. При нечётном n мы получаем код, двойственный коду из предыдущего примера, поэтому в качестве порождающей и проверочной матриц можно взять приведённые выше матрицы H и G соответственно. Поскольку любое кодовое слово $\mathbf{c} = (c_1, \ldots, c_n)$ имеет чётный вес, его последний символ c_n имеет вид $c_n = c_1 \oplus \ldots \oplus c_{n-1}$, где знак $\oplus$ обозначает сложение в поле $GF(2)$. Таким образом, алгоритм кодирования в данном случае сводится к тому, что к блоку $(c_1, \ldots, c_{n-1})$ информационных символов добавляется один символ, равный сумме по модулю 2 всех информационных символов. Поскольку $d_{\mathscr{C}} = 2$, согласно следствию 6.1, декодер $\mathscr{D}_\rho$ обнаруживает любые одиночные ошибки и не может правильно исправлять ошибки; фактически можно утверждать больше: декодер обнаруживает любые комбинации из нечётного числа ошибок.

Оказывается, при помощи проверочной матрицы линейного кода можно находить минимальное кодовое расстояние, а именно, справедлива следующая

Теорема 6.1. *Пусть $\mathscr{C}$ есть линейный код и H — его проверочная матрица. Тогда равенство $d_{\mathscr{C}} = d$ равносильно выполнению следующих двух условий.*

1) *Любые $d - 1$ столбцов матрицы H линейно независимы.*
2) *В матрице H существуют d линейно зависимых столбцов.*

Доказательство. Условие $\mathbf{c} \cdot H^T = \mathbf{0}$, которое, в силу утверждения 6.7, является необходимым и достаточным для принадлежности вектора $\mathbf{c} = (c_1, \ldots, c_n)$ коду $\mathscr{C}$, можно переписать в виде

$$c_1 H_1^{\downarrow} + \ldots + c_n H_n^{\downarrow} = \mathbf{0}^{\downarrow}, \tag{6.3}$$

где $H_i^{\downarrow}$ обозначает i-й столбец матрицы H, а $\mathbf{0}^{\downarrow}$ — нулевой вектор-столбец (все столбцы длины $n - k$). Если $w(\mathbf{c}) = t > 0$, причём ненулевыми в векторе $\mathbf{c}$ являются координаты с номерами $i_1, \ldots, i_t$, то условие (6.3) принимает вид

$$c_{i_1} H_{i_1}^{\downarrow} + \ldots + c_{i_t} H_{i_t}^{\downarrow} = \mathbf{0}^{\downarrow}.$$

Это означает, что t столбцов $H_{i_1}^{\downarrow}, \ldots, H_{i_t}^{\downarrow}$ матрицы H линейно зависимы.

Если $d_{\mathscr{C}} = d$, то найдётся вектор $\mathbf{c} \in \mathscr{C}$ веса $w(\mathbf{c}) = d$, и поэтому некоторые d столбцов матрицы H линейно зависимы. По определению минимального кодового расстояния, никакой ненулевой вектор $\mathbf{c}$, вес которого меньше d, не является кодовым словом, поэтому никакие нетривиальные линейные комбинации из $d-1$ или меньшего числа столбцов матрицы H не равны нулю. Таким образом, условия 1 и 2 теоремы необходимы для выполнения равенства $d_{\mathscr{C}} = d$. Достаточность этих условий доказывается аналогично. □

Определение 6.9. Два линейных кода $\mathscr{C}_1$ и $\mathscr{C}_2$ называются эквивалентными, если они имеют одинаковую размерность и код $\mathscr{C}_2$ можно получить из кода $\mathscr{C}_1$ при помощи некоторой фиксированной перестановки координат во всех кодовых словах.

Нетрудно видеть, что введённое отношение на множестве линейных кодов действительно является отношением эквивалентности, т. е. оно рефлексивно, симметрично и транзитивно. Непосредственно из определения вытекает простой критерий эквивалентности кодов.

Утверждение 6.8. *Пусть $\mathscr{C}_1$ и $\mathscr{C}_2$ — линейные (n, k)-коды. Следующие условия равносильны.*

1) *Коды $\mathscr{C}_1$ и $\mathscr{C}_2$ эквивалентны.*
2) *Если $k \geqslant 1$, то существуют такие порождающие матрицы G_1 и G_2 кодов $\mathscr{C}_1$ и $\mathscr{C}_2$, которые отличаются только перестановкой столбцов.*
3) *Если $k < n$, то существуют такие проверочные матрицы H_1 и H_2 кодов $\mathscr{C}_1$ и $\mathscr{C}_2$, которые отличаются только перестановкой столбцов.*

Замечание 6.7. Очевидно, что при одинаковой перестановке координат в двух векторах расстояние Хэмминга между ними не изменится. Отсюда следует, что структура расстояний между кодовыми словами линейного кода не меняется при переходе к эквивалентному коду; в частности, минимальные кодовые расстояния двух эквивалентных кодов равны. Поэтому при декодировании в ближайшее кодовое слово можно считать, что два эквивалентных кода приводят к одинаковым возможностям по обнаружению и исправлению ошибок. Следовательно, в классе эквивалентных кодов можно выбрать

какой-то один код, удобный с точки зрения простоты кодирования, и не рассматривать другие эквивалентные коды.

Определение 6.10. Линейный (n, k)-код $\mathscr{C}$ называют систематическим, если для него существует порождающая матрица вида $G = (I_k P)$, где I_k — единичная матрица порядка k, а $P = (p_{ij})$ — произвольная матрица размера $k \times (n-k)$.

Утверждение 6.9. *Для любого линейного (n, k)-кода с $k \geqslant 1$ существует эквивалентный ему систематический код.*

Доказательство. Пусть $\mathscr{C}$ — линейный (n, k)-код и G — его порождающая матрица. Применяя алгоритм исключения Гаусса, элементарными преобразованиями строк приведём матрицу G к ступенчатому виду

$$G_1 = \begin{pmatrix} 0 & \dots & 1 & * & \dots & 0 & * & \dots & 0 & * & \dots & * \\ 0 & \dots & 0 & 0 & \dots & 1 & * & \dots & 0 & * & \dots & * \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 1 & * & \dots & * \\ & & i_1 & & & i_2 & & & i_k & & & \end{pmatrix},$$

где строки начинаются с нескольких нулей, за которыми следует единица, причём первая единица в первой строке находится в столбце i_1, первая единица во второй строке находится в столбце i_2, ..., первая единица в k-й строке находится в столбце i_k, $(i_1 < i_2 < \dots < i_k)$, а остальные элементы в столбцах $i_1, \dots, i_k$ равны нулю; символ $*$ обозначает произвольный элемент поля $GF(q)$.

При элементарных преобразованиях строк ранг матрицы сохраняется, а получаемые строки также являются векторами кода $\mathscr{C}$. Поэтому строки матрицы G_1 также образуют базис кода $\mathscr{C}$, иначе говоря, матрица G_1 также будет порождающей для кода $\mathscr{C}$. В матрице G_1 выполним перестановку столбцов: поставим столбец i_1 на первое место, столбец i_2 — на второе место, ..., столбец i_k — на k-е место, а остальные столбцы расположим произвольно на местах $k+1, \dots, n$. В результате получим матрицу $G_2 = (I_k P)$, которая будет порождающей для некоторого систематического (n, k)-кода $\mathscr{C}_1$, эквивалентного коду $\mathscr{C}$. □

Для систематического кода имеется простая связь между порождающей и проверочной матрицами.

Утверждение 6.10. *Если $\mathscr{C}$ — систематический линейный (n, k)-код с порождающей матрицей $G = (I_k P)$, $1 \leqslant k < n$, то в качестве проверочной можно взять матрицу*

$$H = (-P^T I_{n-k}),$$

где $-P^T$ означает матрицу, полученную из P транспонированием и заменой всех элементов на обратные по сложению в поле $GF(q)$, а I_{n-k} есть единичная матрица порядка $n-k$.

Доказательство. Достаточно проверить, что любые две строки $\vec{G}_i$ и $\vec{H}_j$ матриц G и H ортогональны. Действительно,

$$\vec{G}_i = (\ 0, \quad \ldots, \quad 1, \quad \ldots, \quad 0, \quad p_{i1}, \ \ldots, \ p_{ij}, \ \ldots, \ p_{i,\,n-k}),$$
$$\vec{H}_j = (-p_{1j}, \ \ldots, \ -p_{ij}, \ \ldots, \ -p_{kj}, \quad 0, \quad \ldots, \quad 1, \quad \ldots, \quad 0\ ),$$

поэтому $(\vec{G}_i, \vec{H}_j) = -p_{ij} + p_{ij} = 0$. □

Замечание 6.8. 1. Для систематического кода с порождающей матрицей $G = (I_k P)$ формула (6.2) кодирования блока $\mathbf{u} = (u_1, \ldots, u_k)$ информационных символов в блок $\mathbf{c} = (c_1, \ldots, c_n)$ кодовых символов принимает особенно простой вид: $\mathbf{c} = \mathbf{u} \cdot G = (\mathbf{u}, \mathbf{u} \cdot P)$. Таким образом, первые k символов кодового слова $\mathbf{c}$ совпадают с информационными символами, а следующие $n-k$ символов, называемых *проверочными*, представляют собой некоторые линейные комбинации информационных символов, причём коэффициенты линейных комбинаций заданы матрицей P,

$$c_{k+i} = u_1 p_{1i} + u_2 p_{2i} + \ldots + u_k p_{ki}, \quad 1 \leqslant i \leqslant n-k.$$

2. Если при этом $q = 2$, т. е. $\mathscr{C}$ является двоичным систематическим линейным (n, k)-кодом, и на множестве кодовых слов задано равномерное распределение, то выражение (5.1) (с. 137) для скорости передачи принимает вид $R_n = k/n$, т. е. для двоичного систематического кода скорость передачи равна отношению числа информационных символов к общей длине кодового слова.

Перейдём к задаче о построении декодера для линейного кода. Эта задача приводит к необходимости рассмотреть разбиение пространства $V_n(q)$ на смежные классы по коду $\mathscr{C}$.

Напомним, что два вектора $\mathbf{u}$ и $\mathbf{v}$ из $V_n(q)$ называются эквивалентными по коду $\mathscr{C}$, если $\mathbf{u} - \mathbf{v} \in \mathscr{C}$. Это отношение действительно

является отношением эквивалентности, т. е. оно рефлексивно, симметрично и транзитивно. Всё пространство $V_n(q)$ разбивается на классы эквивалентных векторов, и эти классы называются смежными классами по коду $\mathscr{C}$. Если $\mathbf{u}$ — представитель некоторого смежного класса, то этот класс можно записать в виде

$$\mathbf{u} + \mathscr{C} = \{\mathbf{u} + \mathbf{c} \mid \mathbf{c} \in \mathscr{C}\}.$$

Если размерность кода $\mathscr{C}$ равна k, то мощность каждого смежного класса равна q^k, а общее число смежных классов равно q^{n-k}.

Предположим, что $\mathbf{c}$ есть кодовое слово, переданное в канал связи, и $\mathbf{y}$ — принятый вектор. Вектор $\mathbf{e} = \mathbf{y} - \mathbf{c}$ будем называть *вектором ошибок*. Очевидно, что задача декодирования, т. е. восстановления кодового слова $\mathbf{c}$ по принятому вектору $\mathbf{y}$, эквивалентна восстановлению вектора ошибок $\mathbf{e}$. Поскольку ошибки в канале связи могут быть любыми, вектор $\mathbf{e}$ может, вообще говоря, принимать произвольные значения из $V_n(q)$. Однако тот факт, что код $\mathscr{C}$ является линейным подпространством пространства $V_n(q)$, позволяет существенно сократить перебор возможных вариантов значений вектора $\mathbf{e}$. Действительно, условие

$$\mathbf{y} - \mathbf{e} = \mathbf{c} \in \mathscr{C}$$

означает, что подлежащий определению вектор ошибок $\mathbf{e}$ лежит в *том же смежном классе*, что и принятый вектор $\mathbf{y}$, тем самым перебор вариантов при поиске вектора $\mathbf{e}$ сокращается до одного заданного смежного класса $\mathbf{y} + \mathscr{C}$.

Зафиксируем некоторую нумерацию кодовых слов линейного (n, k)-кода $\mathscr{C}$, причём нулевому кодовому слову $\mathbf{0}$ припишем номер 1,

$$\mathscr{C} = \{\mathbf{c}_1 = \mathbf{0}, \mathbf{c}_2, \ldots, \mathbf{c}_{q^k}\}.$$

Определение 6.11. Таблицей стандартного расположения для линейного (n, k)-кода $\mathscr{C}$ называется таблица $S = (\mathbf{y}_{ij})$ размера $q^{n-k} \times q^k$, составленная из векторов пространства $V_n(q)$ по следующему правилу.

1. В первой строке $\vec{S}_1$ таблицы S записаны q^k векторов кода $\mathscr{C}$,

$$\vec{S}_1 = (\mathbf{y}_{11}, \ldots, \mathbf{y}_{1,q^k}) = (\mathbf{c}_1 = \mathbf{0}, \mathbf{c}_2, \ldots, \mathbf{c}_{q^k}).$$

2. Если строки $\vec{S}_1, \ldots, \vec{S}_{i-1}$ уже построены, то строка $\vec{S}_i$ строится так: в качестве вектора $\mathbf{y}_{i1}$ выбираем произвольный вектор из

$V_n(q)$, который ещё не встретился в построенных первых $i-1$ строках, а остальные элементы i-й строки вычисляем по формуле

$$\mathbf{y}_{ij} = \mathbf{y}_{i1} + \mathbf{c}_j, \quad 2 \leqslant j \leqslant q^k.$$

Замечание 6.9. Указанный способ построения таблицы S гарантирует, что в ней по одному разу будут записаны все q^n векторов пространства $V_n(q)$. Покажем (от противного), что все элементы таблицы различны. Действительно, i-ю строку таблицы S можно записать в виде $\mathbf{y}_{i1} + \mathscr{C}$, другими словами, в строке записан смежный класс по коду $\mathscr{C}$, и поэтому все элементы в строке различны. Пусть в строках с номерами $i < s$ есть одинаковые элементы, $\mathbf{y}_{ij} = \mathbf{y}_{st}$, тогда

$$\mathbf{y}_{s1} = \mathbf{y}_{i1} + \mathbf{c}_j - \mathbf{c}_t \in \mathbf{y}_{i1} + \mathscr{C}.$$

Следовательно, вектор $\mathbf{y}_{s1}$ является также элементом строки $\vec{S}_i$, что противоречит правилу построения таблицы S. Итак, строками таблицы стандартного расположения являются все различные смежные классы по коду $\mathscr{C}$. Отметим также, что таблицу стандартного расположения при заданных ограничениях можно построить не единственным образом.

Определение 6.12. Пусть $\mathscr{C}$ — линейный код. Декодером $\mathscr{D}_S$ на основе таблицы стандартного расположения S называется такое отображение $\mathscr{D}_S : V_n(q) \to \mathscr{C}$, что если принятый вектор $\mathbf{y}$ совпадает с элементом $\mathbf{y}_{ij}$ таблицы S, то $\mathscr{D}_S(\mathbf{y}) = \mathbf{c}_j$.

Замечание 6.10. Алгоритмически декодер $\mathscr{D}_S$ описать просто: принятый вектор $\mathbf{y}$ ищем в таблице S, и если находим его на месте (i, j), то в качестве результата декодирования берём вектор, расположенный в таблице S на месте $(1, j)$, т. е. в том же столбце и в первой строке. Поскольку в таблице S по одному разу встречаются все векторы пространства $V_n(q)$, поиск будет всегда успешным, а результат декодирования определяется однозначно. Трудоёмкость поиска в худшем случае пропорциональна q^n.

Существенное уменьшение трудоёмкости декодера $\mathscr{D}_S$ возможно за счёт использования понятия *синдрома* вектора, определяемого при фиксированной проверочной матрице.

Определение 6.13. Пусть $\mathscr{C}$ — линейный код и H — его проверочная матрица. Синдромом $s(\mathbf{y})$ вектора $\mathbf{y} \in V_n(q)$ называется $(n-k)$-мерный вектор $s(\mathbf{y}) = \mathbf{y} \cdot H^T$.

Утверждение 6.11.

1. *Условия* $\mathbf{y} \in \mathscr{C}$ *и* $s(\mathbf{y}) = \mathbf{0}$ *равносильны.*
2. *Векторы* $\mathbf{y}$ *и* $\mathbf{y}'$ *лежат в одном смежном классе по коду* $\mathscr{C}$ *тогда и только тогда, когда* $s(\mathbf{y}) = s(\mathbf{y}')$.

Доказательство. Первая часть утверждения непосредственно следует из определения синдрома и утверждения 6.7. Вторая часть утверждения следует из того, что условие $\mathbf{y} - \mathbf{y}' \in \mathscr{C}$ равносильно равенству $(\mathbf{y} - \mathbf{y}') \cdot H^T = \mathbf{0}$, или, что то же самое, $\mathbf{y} \cdot H^T = \mathbf{y}' \cdot H^T$. □

Замечание 6.11. Имеется всего q^{n-k} различных значений синдрома, по числу смежных классов, и поэтому синдром можно рассматривать просто как q-ичный $(n-k)$-разрядный номер смежного класса. Таблице стандартного расположения S можем сопоставить новую таблицу S' размера $q^{n-k} \times 2$, первый столбец которой совпадает с первым столбцом таблицы S, а во втором столбце записаны синдромы соответствующих смежных классов:

$$S' = \begin{bmatrix} \mathbf{y}_{11} & s(\mathbf{y}_{11}) \\ \vdots & \vdots \\ \mathbf{y}_{i1} & s(\mathbf{y}_{i1}) \\ \vdots & \vdots \\ \mathbf{y}_{q^{n-k},1} & s(\mathbf{y}_{q^{n-k},1}) \end{bmatrix}.$$

При помощи таблицы S' декодер $\mathscr{D}_S$ можно представить в виде эквивалентного, но более простого алгоритма: для принятого вектора $\mathbf{y}$ вычисляем синдром $s(\mathbf{y})$ и ищем этот синдром во втором столбце таблицы S'; если $s(\mathbf{y}) = s(\mathbf{y}_{i1})$, то полагаем $\mathscr{D}_S(\mathbf{y}) = \mathbf{y} - \mathbf{y}_{i1}$.

В таком варианте алгоритма этап поиска принятого вектора $\mathbf{y}$ в таблице S (трудоёмкость в худшем случае пропорциональна q^n) заменяется на поиск синдрома во втором столбце таблицы S' (трудоёмкость в худшем случае пропорциональна q^{n-k}); если же строки таблицы S' предварительно упорядочить, например, в порядке лексикографического возрастания синдромов, то этап поиска при декодировании вообще исключается. Кроме того, для такого вариан-

та алгоритма требуется существенно меньше памяти для хранения таблицы.

Определение 6.14. Вектор $\mathbf{y}$ называют лидером смежного класса $\mathbf{y}+\mathscr{C}$, если $\mathbf{y}$ имеет минимальный вес среди векторов этого смежного класса. Таблицу стандартного расположения S называют таблицей Слепяна, если она удовлетворяет дополнительному условию: в каждой строке $\vec{S}_i$ в качестве первого элемента $\mathbf{y}_{i1}$ выбирается лидер смежного класса, т. е. вектор *минимального веса*, который ещё не встретился в построенных первых $i-1$ строках таблицы.

Пример 6.3. Для двоичного линейного $(3, 1)$-кода $\mathscr{C}=\{(000), (111)\}$ одна из возможных таблиц Слепяна имеет вид

$$S=\begin{bmatrix} 000 & 111 \\ 001 & 110 \\ 010 & 101 \\ 100 & 011 \end{bmatrix}.$$

Утверждение 6.12. 1. *Декодер $\mathscr{D}_S$ правильно декодирует принятый вектор $\mathbf{y}=\mathbf{y}_{ij}$ в кодовое слово $\mathbf{c}_j$ в том и только том случае, когда вектор ошибок $\mathbf{e}$ совпадает с вектором $\mathbf{y}_{i1}$.*

2. *Если S — таблица Слепяна, то декодеры $\mathscr{D}_S$ и $\mathscr{D}_\rho$ эквивалентны.*

Доказательство. Первая часть утверждения вытекает из сопоставления равенств $\mathbf{y}=\mathbf{c}+\mathbf{e}$ и $\mathbf{y}_{ij}=\mathbf{c}_j+\mathbf{y}_{i1}$.

Докажем вторую часть утверждения. В случае линейного кода *любой декодер* ищет вектор ошибок $\mathbf{e}$ в том же смежном классе, в котором лежит принятый вектор $\mathbf{y}$, при этом $\rho(\mathbf{y}, \mathbf{c})=w(\mathbf{e})$. Следовательно, декодер $\mathscr{D}_\rho$ должен выбирать в качестве вектора ошибок $\mathbf{e}$ лидера смежного класса $\mathbf{y}+\mathscr{C}$ и вычислять $\mathbf{c}=\mathbf{y}-\mathbf{e}$. Нетрудно видеть, что это в точности совпадает с декодированием $\mathscr{D}_S$ по таблице Слепяна. □

Итак, если в качестве таблицы стандартного расположения взять таблицу Слепяна, то табличный декодер $\mathscr{D}_S$ сводится к описанному выше декодеру $\mathscr{D}_\rho$ в ближайшее кодовое слово.

Покажем, что в одном важном частном случае таблица Слепяна является наилучшей среди всех таблиц стандартного расположения в том смысле, что она обеспечивает наименьшую вероятность оши-

бочного декодирования. Согласно определениям 5.3 (с. 138) и 5.4 (с. 139), вероятность ошибочного декодирования равна

$$\lambda = \mathsf{P}(\mathscr{D}(Y^n) \neq X^n),$$

где X^n — случайное кодовое слово на входе канала связи, Y^n — случайный вектор на выходе канала связи и $\mathscr{D}$ — применяемый декодер. Поскольку нам нужно сравнивать вероятности ошибочного декодирования для различных декодеров, будем подчёркивать зависимость вероятности λ от применяемого декодера $\mathscr{D}$ обозначением $\lambda(\mathscr{D})$.

Теорема 6.2. *Пусть $\mathscr{C}$ — двоичный линейный (n, k)-код, на коде $\mathscr{C}$ задано равномерное распределение*

$$\mathsf{P}(X^n = \mathbf{c}) = 2^{-k}, \quad \mathbf{c} \in \mathscr{C},$$

а в качестве канала связи используется двоичный симметричный канал с параметром $p \leqslant 1/2$. Если S — произвольная таблица стандартного расположения и $\widetilde{S}$ — таблица Слепяна, то $\lambda(\mathscr{D}_S) \geqslant \lambda(\mathscr{D}_{\widetilde{S}})$.

Доказательство. Рассмотрим два способа доказательства теоремы. Первый способ сводится к применению нескольких ранее доказанных утверждений. Второй способ, связанный с простыми, но несколько громоздкими выкладками, фактически повторяет идеи доказательства упомянутых утверждений и приводится только с учебной целью, как более подробное объяснение.

1. Декодер $\mathscr{D}_{\mathsf{AP}}$ обеспечивает минимальную среднюю вероятность ошибочного декодирования (утверждение 5.2, с. 141). При равномерном распределении на множестве кодовых слов декодер $\mathscr{D}_{\mathsf{AP}}$ эквивалентен декодеру $\mathscr{D}_{\mathsf{L}}$ (утверждение 5.3, с. 141), а для двоичного симметричного канала с $p \leqslant 1/2$ декодер $\mathscr{D}_{\mathsf{L}}$ эквивалентен декодеру $\mathscr{D}_{\rho}$ (утверждение 6.2, с. 165). Кроме того, при использовании таблицы Слепяна $\widetilde{S}$ декодеры $\mathscr{D}_{\rho}$ и $\mathscr{D}_{\widetilde{S}}$ эквивалентны (утверждение 6.12). Следовательно, при условиях теоремы декодер $\mathscr{D}_{\widetilde{S}}$ также обеспечивает минимальную среднюю вероятность ошибочного декодирования.

2. В двоичном случае покоординатное суммирование векторов над полем GF(2) обозначаем символом $\oplus$.

Пусть передано кодовое слово $X^n = \mathbf{c}_j$ и Y^n — принятый вектор. Согласно п. 1 утверждения 6.12, при использовании декодера $\mathscr{D}_S$ бу-

дут правильно декодированы те и только те значения $Y^n = \mathbf{y}$, которые расположены в j-м столбце таблице S, т. е.

$$\mathbf{y} = \mathbf{y}_{ij}, \quad 1 \leqslant i \leqslant 2^{n-k},$$

в частности, $\mathbf{y}_{1j} = \mathbf{c}_j$. Обозначим $w_{ij} = \rho(\mathbf{y}_{1j}, \mathbf{y}_{ij})$. Так как $\mathbf{y}_{ij} = \mathbf{y}_{i1} \oplus \mathbf{y}_{1j}$ и $\mathbf{y}_{11} = \mathbf{c}_1 = \mathbf{0}$, получаем

$$w_{ij} = w(\mathbf{y}_{i1}) = w_{i1}, \quad 1 \leqslant j \leqslant 2^k. \tag{6.4}$$

По определению двоичного симметричного канала

$$\mathsf{P}(Y^n = \mathbf{y}_{ij} / X^n = \mathbf{y}_{1j}) = p^{w_{ij}} (1-p)^{n-w_{ij}},$$

следовательно, условная вероятность правильного декодирования при условии $X^n = \mathbf{y}_{1j}$ равна

$$\mathsf{P}(\mathscr{D}_S(Y^n) = X^n / X^n = \mathbf{y}_{1j}) = \sum_{i=1}^{2^{n-k}} \mathsf{P}(Y^n = \mathbf{y}_{ij} / X^n = \mathbf{y}_{1j}) = \sum_{i=1}^{2^{n-k}} p^{w_{ij}} (1-p)^{n-w_{ij}},$$

а безусловная вероятность правильного декодирования равна

$$1 - \lambda(\mathscr{D}_S) = \mathsf{P}(\mathscr{D}_S(Y^n) = X^n) = \sum_{j=1}^{2^k} \mathsf{P}(X^n = \mathbf{y}_{1j}) \mathsf{P}(\mathscr{D}_S(Y^n) = X^n / X^n = \mathbf{y}_{1j}) = 2^{-k} \sum_{j=1}^{2^k} \sum_{i=1}^{2^{n-k}} p^{w_{ij}} (1-p)^{n-w_{ij}}.$$

Изменяя порядок суммирования и принимая во внимание (6.4), получаем

$$1 - \lambda(\mathscr{D}_S) = 2^{-k} \sum_{i=1}^{2^{n-k}} \sum_{j=1}^{2^k} p^{w_{ij}} (1-p)^{n-w_{ij}} = 2^{-k} \sum_{i=1}^{2^{n-k}} 2^k p^{w_{i1}} (1-p)^{n-w_{i1}} = (1-p)^n \sum_{i=1}^{2^{n-k}} \left(\frac{p}{1-p} \right)^{w_{i1}}.$$

Для таблицы Слепяна $\widetilde{S} = (\widetilde{\mathbf{y}}_{ij})$ точно так же получаем формулу

$$1 - \lambda(\mathscr{D}_{\widetilde{S}}) = (1-p)^n \sum_{i=1}^{2^{n-k}} \left(\frac{p}{1-p}\right)^{\widetilde{w}_{i1}},$$

где $\widetilde{w}_{i1} = w(\widetilde{\mathbf{y}}_{i1})$. По определению таблицы Слепяна, векторы $\widetilde{\mathbf{y}}_{i1}$ являются лидерами своих смежных классов, поэтому при одинаковой нумерации смежных классов в таблицах S и $\widetilde{S}$ справедливы неравенства $\widetilde{w}_{i1} \leqslant w_{i1}$, $1 \leqslant i \leqslant 2^{n-k}$. Следовательно, при $p \leqslant 1/2$

$$\left(\frac{p}{1-p}\right)^{w_{i1}} \leqslant \left(\frac{p}{1-p}\right)^{\widetilde{w}_{i1}},$$

и из полученных выше формул для вероятностей правильного декодирования вытекает нужное неравенство $\lambda(\mathscr{D}_S) \geqslant \lambda(\mathscr{D}_{\widetilde{S}})$. □

Важным вопросом при изучении линейных кодов является выяснение допустимых значений параметров n (длина кодового слова), k (размерность кода) и d (минимальное кодовое расстояние). Если два из трёх параметров заданы, то желательно оптимизировать свойства кода, выбирая значение третьего параметра. При этом выясняется, что возникающие требования к параметрам противоречивы, и на практике при выборе кода приходится принимать компромиссные решения, исходя из особенностей решаемой задачи. Например, если заданы параметры n и k, то желательно сделать параметр d как можно больше, чтобы улучшить возможности по обнаружению и исправлению ошибок; если заданы параметры n и d, то желательно увеличить k, чтобы мощность кода была как можно большей. Из большого числа известных результатов о соотношениях между параметрами n, k и d мы приведём только три — границы Хэмминга, Синглтона и Варшамова—Гилберта.

Утверждение 6.13 (граница Хэмминга). *Для любого линейного q-ичного (n, k, d)-кода $\mathscr{C}$ справедливо неравенство*

$$q^{n-k} \geqslant \sum_{i=0}^{t} \binom{n}{i} (q-1)^i, \tag{6.5}$$

где $t = [(d-1)/2]$.

Доказательство. Рассмотрим шары

$$K_t(\mathbf{c}) = \{\mathbf{y} \in V_n(q) : \rho(\mathbf{y}, \mathbf{c}) \leqslant t\}$$

с центрами в кодовых словах $\mathbf{c} \in \mathscr{C}$. Очевидно, что

$$\bigcup_{\mathbf{c}\in\mathscr{C}} K_t(\mathbf{c}) \subseteq V_n(q).$$

Докажем (от противного), что если кодовые слова $\mathbf{c}$ и $\mathbf{c}'$ различны, то шары $K_t(\mathbf{c})$ и $K_t(\mathbf{c}')$ не пересекаются. Действительно, если два этих шара содержат общий элемент $\mathbf{y}$, то по неравенству треугольника

$$\rho(\mathbf{c}, \mathbf{c}') \leqslant \rho(\mathbf{c}, \mathbf{y}) + \rho(\mathbf{y}, \mathbf{c}') \leqslant 2t < d,$$

что противоречит определению минимального кодового расстояния.

Следовательно,

$$\sum_{\mathbf{c}\in\mathscr{C}} |K_t(\mathbf{c})| \leqslant |V_n(q)| = q^n.$$

Заметим, что все q^k шаров $K_t(\mathbf{c})$, $\mathbf{c} \in \mathscr{C}$, равномощны, и мощность каждого из них равна

$$|K_t(\mathbf{c})| = \sum_{i=0}^{t} \binom{n}{i} (q-1)^i,$$

поэтому $q^k \sum\limits_{i=0}^{t} \binom{n}{i} (q-1)^i \leqslant q^n$, что равносильно (6.5). □

Определение 6.15. *Линейный q-ичный (n, k, d)-код $\mathscr{C}$* называется совершенным, или плотно упакованным, если в неравенстве (6.5) достигается равенство.

Утверждение 6.14 (граница Синглтона). *Для любого линейного q-ичного (n, k, d)-кода $\mathscr{C}$ справедливо неравенство*

$$d \leqslant n - k + 1. \tag{6.6}$$

Доказательство. Если $k = n$, то $\mathscr{C} = V_n(q)$, $d = 1$, и нужное неравенство выполнено. Пусть $k < n$. По теореме 6.1, если $d_{\mathscr{C}} = d$, то любые $d - 1$ столбцов проверочной матрицы H линейно независимы. Поскольку столбцы матрицы H имеют длину $n - k$, максимальное число линейно независимых столбцов не может быть больше, чем $n - k$. Следовательно, $d - 1 \leqslant n - k$. □

Замечание 6.12. Формула (6.6) справедлива не только для линейного (n, k, d)-кода, но и для любого блокового кода $\mathscr{C}$ длины n с минимальным кодовым расстоянием $d_{\mathscr{C}} = d$, при этом в качестве параметра k следует взять $k = \log_q |\mathscr{C}|$ (это число не обязательно целое).

Чтобы доказать такое обобщение, для всех кодовых слов блокового кода $\mathscr{C}$ рассмотрим их подвекторы, образованные первыми $n - d + 1$ координатами. Все эти подвекторы различны, так как в противном случае, если два различных кодовых слова совпадают в первых $n - d + 1$ координатах, то расстояние между этими кодовыми словами не превосходит $d - 1$, что противоречит условию $d_{\mathscr{C}} = d$. Итак, мы имеем $|\mathscr{C}| = q^k$ различных векторов длины $n - d + 1$, и это число не может быть больше q^{n-d+1}. Следовательно, $q^k \leqslant q^{n-d+1}$, поэтому $k \leqslant n - d + 1$.

Отметим также, что поскольку в доказательстве границы Хэмминга по существу не использовалось свойство линейности кода, эта граница остаётся верной и для любого блокового кода $\mathscr{C}$, если, как и выше, положить $k = \log_q |\mathscr{C}|$.

Определение 6.16. Линейный q-ичный (n, k, d)-код $\mathscr{C}$ называется кодом с максимально допустимым расстоянием, или МДР-кодом, если в неравенстве (6.6) достигается равенство.

Пример 6.4. Нетрудно проверить, что код из примера 6.3 является и совершенным, и МДР-кодом.

Утверждение 6.15 (граница Варшамова—Гилберта). *Пусть натуральные числа $d \geqslant 2$, n и r таковы, что выполняется неравенство*

$$\sum_{j=0}^{d-2} \binom{n-1}{j} (q-1)^j < q^r.$$

Тогда существует q-ичный линейный код $\mathscr{C}$ длины n, размерности $k \geqslant n - r$ и с минимальным кодовым расстоянием $d_{\mathscr{C}} \geqslant d$.

Доказательство. Построим над полем $GF(q)$ матрицу H размера $r \times n$, в которой любые $d - 1$ столбцов линейно независимы. Для этого в качестве первого столбца $H_1^{\downarrow}$ возьмём произвольный ненулевой столбец длины r. Предположим, что уже построены i столбцов $H_1^{\downarrow}, \ldots, H_i^{\downarrow}$ так, что любые $d - 1$ из них линейно независимы. Рассмотрим всевозможные линейные комбинации $a_1 H_1^{\downarrow} + \ldots + a_i H_i^{\downarrow}$,

в которых среди коэффициентов $a_1, \ldots, a_i \in GF(q)$ имеется не менее одного и не более $d-2$ ненулевых. Число способов составить такие линейные комбинации равно

$$\sum_{j=1}^{d-2} \binom{i}{j} (q-1)^j,$$

и некоторые из этих линейных комбинаций могут представлять одинаковые векторы длины r. Если выполняется неравенство

$$\sum_{j=1}^{d-2} \binom{i}{j} (q-1)^j < q^r - 1, \tag{6.7}$$

то среди всех $q^r - 1$ ненулевых векторов длины r существует такой вектор $b^{\downarrow}$, который не может быть представлен никакой линейной комбинацией указанного вида. Возьмём этот вектор $b^{\downarrow}$ в качестве очередного столбца $H_{i+1}^{\downarrow}$. Так как столбец $H_{i+1}^{\downarrow}$ не выражается в виде линейной комбинации никаких $d-2$ столбцов из $H_1^{\downarrow}, \ldots, H_i^{\downarrow}$, мы получаем, что любые $d-1$ столбцов матрицы $(H_1^{\downarrow}, \ldots, H_{i+1}^{\downarrow})$ линейно независимы.

Левая часть неравенства (6.7) растёт с ростом i, и при некотором i неравенство впервые нарушается. Предположим, что (6.7) верно для $i = 1, 2, \ldots, n-1$ и неверно при $i = n$. Это означает, что при помощи описанного алгоритма построена искомая матрица $H = (H_1^{\downarrow}, \ldots, H_n^{\downarrow})$ размера $r \times n$.

Определим код $\mathscr{C}$ как множество всех векторов из $V_n(q)$, которые ортогональны всем строкам матрицы H, т. е. $\mathscr{C} = \{\mathbf{c} \in V_n(q) \mid \mathbf{c} \cdot H^T = \mathbf{0}\}$. Из теоремы 6.1 следует, что $d_{\mathscr{C}} \geqslant d$. Матрица H, вообще говоря, не является проверочной матрицей кода $\mathscr{C}$, так как строки этой матрицы могут быть линейно зависимыми. Проверочную матрицу H' можно получить из H, вычёркивая строки, которые линейно выражаются через другие строки. Число строк в H' равно $n-k$, где k — размерность кода $\mathscr{C}$, и не превосходит r — числа строк в H, поэтому $k \geqslant n-r$. □

§ 6.3. Циклические коды

В классе линейных кодов, в свою очередь, выделяется важный подкласс циклических кодов, для которых описание, оценка пара-

метров, алгоритмы кодирования и декодирования основываются на теории конечных полей и многочленов над конечными полями.

В этом параграфе нам будет удобно нумеровать координаты векторов из пространства $V_n(q)$ числами $0, 1, \ldots, n-1$, $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1})$.

Определение 6.17. Линейный q-ичный (n, k)-код $\mathscr{C}$ называется циклическим, если для любого кодового слова $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1})$ его циклический сдвиг $\mathbf{c}' = (c_{n-1}, c_0, c_1, \ldots, c_{n-2})$ также принадлежит коду $\mathscr{C}$.

Замечание 6.13. В определении указан циклический сдвиг на один шаг *вправо*, а не влево, но направление сдвига здесь несущественно, так как циклический сдвиг на один шаг влево равносилен циклическому сдвигу на $n-1$ шагов вправо.

Существует естественное взаимно однозначное соответствие между векторами из пространства $V_n(q)$ и многочленами над $GF(q)$ степени не выше $n-1$:

$$\mathbf{c} = (c_0, c_1, \ldots, c_{n-1}) \leftrightarrow c(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}.$$

Кольцо многочленов над полем $GF(q)$ обозначаем через $GF(q)[x]$. Далее нам потребуется факторкольцо

$$\mathscr{R}_n = GF(q)[x]/(x^n - 1),$$

элементами которого являются классы многочленов, сравнимых по модулю $x^n - 1$.

Напомним, что два многочлена $a(x)$ и $b(x)$ называются сравнимыми по модулю многочлена $m(x)$, если $m(x)$ делит разность $a(x) - b(x)$, или, что то же самое, многочлены $a(x)$ и $b(x)$ дают одинаковые остатки при делении на $m(x)$.

Элементы факторкольца $\mathscr{R}_n$ будем обозначать $[c(x)]_{x^n-1}$, или коротко $[c(x)]$. В каждом классе $[c(x)]$ содержится многочлен степени не выше $n-1$, и любые два различных многочлена степени не выше $n-1$ лежат в разных классах факторкольца $\mathscr{R}_n$. Поэтому в качестве системы представителей классов из $\mathscr{R}_n$ можно взять множество всех различных многочленов степени не выше $n-1$,

$$\mathscr{R}_n = \{[c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}] : c_i \in GF(q)\}.$$

Таким образом, имеется взаимно однозначное соответствие между векторами из пространства $V_n(q)$ и классами многочленов из $\mathscr{R}_n$:

$$V_n(q) \ni \mathbf{c} \longleftrightarrow [c(x)] \in \mathscr{R}_n,$$

следовательно, $|\mathscr{R}_n| = |V_n(q)| = q^n$. Нетрудно также показать, что $\mathscr{R}_n$ с операцией сложения и внешней операцией умножения на элементы поля $GF(q)$ представляет собой линейное пространство, изоморфное пространству $V_n(q)$.

Определение 6.18. Если $\mathscr{C}$ — произвольное подмножество пространства $V_n(q)$, то соответствующее подмножество классов из $\mathscr{R}_n$ будем обозначать через $[\mathscr{C}]$.

Замечание 6.14. Очевидно, что если $\mathscr{C}$ — линейный код, то множество $[\mathscr{C}]$ с операцией сложения и внешней операцией умножения на элементы поля $GF(q)$ представляет собой линейное пространство, изоморфное $\mathscr{C}$.

Следующее простое утверждение показывает, что переход к рассмотрению факторкольца $\mathscr{R}_n$ является естественным шагом при изучении циклических кодов.

Утверждение 6.16. *Если* $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1}) \in V_n(q)$ *и* $\mathbf{c}'$ *— циклический сдвиг вектора* $\mathbf{c}$, *то* $[c'(x)] = [x] \cdot [c(x)]$.

Доказательство. Поскольку $[x^n - 1] = [0]$, $[x^n] = [1]$. Имеем

$$\begin{aligned}[x] \cdot [c(x)] &= [x] \cdot [c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}] = \\ &= [c_0 x + c_1 x^2 + \ldots + c_{n-2} x^{n-1} + c_{n-1}\, x^n] = \\ &= [c_{n-1} + c_0 x + c_1 x^2 + \ldots + c_{n-2} x^{n-1}] = \\ &= [c'(x)].\end{aligned}$$

□

Утверждение 6.17. *Если* $\mathscr{C}$ *— циклический код,* $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1})$ *— кодовое слово кода* $\mathscr{C}$ *и* $f(x)$ *— произвольный многочлен над* $GF(q)$, *то класс* $[f(x)] \cdot [c(x)]$ *принадлежит множеству* $[\mathscr{C}]$.

Доказательство. Если $f(x) = f_0 + f_1 x + \ldots + f_t x^t$, то

$$[f(x)] \cdot [c(x)] = \sum_{i=0}^{t} [f_i] \cdot [x^i] \cdot [c(x)].$$

Из определения циклического кода и утверждения 6.16 по индукции следует, что для любого $i \geqslant 0$ класс $[x^i] \cdot [c(x)]$ принадлежит множеству $[\mathscr{C}]$. В силу изоморфизма линейных пространств $\mathscr{C}$ и $[\mathscr{C}]$ каждое слагаемое $[f_i] \cdot [x^i] \cdot [c(x)]$ лежит в $[\mathscr{C}]$, и поэтому вся сумма лежит в $[\mathscr{C}]$. □

Утверждение 6.17 означает, что если $\mathscr{C}$ — циклический код, то множество $[\mathscr{C}]$ является идеалом в факторкольце $\mathscr{R}_n$. Простым упражнением для читателя будет доказательство обратного утверждения: для любого идеала в факторкольце $\mathscr{R}_n$ соответствующее ему множество векторов из $V_n(q)$ образует циклический код. Таким образом, имеется взаимно однозначное соответствие между циклическими кодами в пространстве $V_n(q)$ и идеалами в факторкольце $\mathscr{R}_n$.

Утверждение 6.18. *Пусть $\mathscr{C}$ — циклический код и $g(x)$ — ненулевой многочлен наименьшей степени, для которого $[g(x)] \in [\mathscr{C}]$. Тогда идеал $[\mathscr{C}]$ порождается классом $[g(x)]$ в $\mathscr{R}_n$ и, кроме того, многочлен $g(x)$ делит $x^n - 1$.*

Доказательство. Ясно, что идеал, порождённый классом $[g(x)]$ в факторкольце $\mathscr{R}_n$, является подмножеством множества $[\mathscr{C}]$. Чтобы доказать обратное включение, рассмотрим произвольный класс $[c(x)] \in [\mathscr{C}]$ и разделим $c(x)$ на $g(x)$ с остатком:

$$c(x) = g(x)\, a(x) + r(x), \quad \deg r(x) < \deg g(x).$$

Тогда $[r(x)] = [c(x) - g(x)a(x)] = [c(x)] - [g(x)] \cdot [a(x)] \in [\mathscr{C}]$, а так как $g(x)$ — ненулевой многочлен наименьшей степени с условием $[g(x)] \in [\mathscr{C}]$, получаем, что $r(x) = 0$, т. е. $g(x)$ делит $c(x)$.

Делимость $x^n - 1$ на $g(x)$ доказывается аналогично — достаточно разделить $x^n - 1$ на $g(x)$ с остатком и воспользоваться определением многочлена $g(x)$. □

Доказанное утверждение означает, что любой идеал в факторкольце $\mathscr{R}_n$ является главным идеалом, т. е. он порождается некоторым классом $[g(x)]$. Очевидно, многочлен $g(x)$ при этом можно выбрать унитарным, т. е. со старшим коэффициентом, равным 1.

Идеал, порождаемый классом $[g(x)]$ в $\mathscr{R}_n$, будем обозначать $\langle [g(x)] \rangle$.

Определение 6.19. Пусть $\mathscr{C}$ — циклический код и $g(x)$ — унитарный многочлен наименьшей степени, для которого $\langle [g(x)] \rangle = [\mathscr{C}]$. Тогда говорят, что $g(x)$ является порождающим многочленом кода $\mathscr{C}$.

Замечание 6.15. Порождающий многочлен циклического кода определяется однозначно. Действительно, если $g'(x)$ и $g''(x)$ — два порождающих многочлена, то эти многочлены взаимно делят друг друга, но так как они оба унитарные, $g'(x) = g''(x)$.

Утверждение 6.19. *Пусть $\mathscr{C}$ — циклический (n, k)-код и $g(x)$ — его порождающий многочлен. Тогда* $\deg g(x) = n - k$.

Доказательство. Обозначим $\deg g(x) = \ell$. Если $[c(x)] = [c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}]$ — произвольный элемент идеала $[\mathscr{C}]$, то $g(x)$ делит $c(x)$, т. е. найдётся такой многочлен $f(x) = f_0 + f_1 x + \ldots + f_{n-\ell-1} x^{n-\ell-1}$, что $c(x) = f(x) g(x)$, или

$$[c(x)] = \sum_{i=0}^{n-\ell-1} [f_i] \cdot [x^i g(x)].$$

Таким образом, любой класс $[c(x)] \in [\mathscr{C}]$ представляется в виде линейной комбинации $n - \ell$ классов $[g(x)], [xg(x)], \ldots, [x^{n-\ell-1} g(x)]$. Эти классы линейно независимы, поскольку никакой многочлен $x^i g(x)$ не может быть линейно выражен через многочлены $g(x)$, $xg(x), \ldots, x^{i-1} g(x)$ меньшей степени. Следовательно, размерность $[\mathscr{C}]$ как линейного пространства над $GF(q)$ равна $n - \ell$. Но линейные пространства $\mathscr{C}$ и $[\mathscr{C}]$ изоморфны, и $\dim \mathscr{C} = k$, поэтому $n - \ell = k$, или $\deg g(x) = \ell = n - k$. □

Следствие 6.2. *Пусть $\mathscr{C}$ — циклический (n, k)-код с $k \geqslant 1$ и порождающим многочленом $g(x) = g_0 + g_1 x + \ldots + g_{n-k} x^{n-k}$. Тогда матрица*

$$G = \begin{pmatrix} g_0 & g_1 & \cdots & \cdots & g_{n-k} & 0 & 0 & \ldots & 0 & 0 \\ 0 & g_0 & \cdots & \cdots & g_{n-k-1} & g_{n-k} & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & g_0 & \ldots & \ldots & \ldots & \ldots & g_{n-k-1} & g_{n-k} \end{pmatrix}$$

размера $k \times n$ является порождающей матрицей кода $\mathscr{C}$.

Доказательство. Из доказательства утверждения 6.19 видно, что множество из k классов $[g(x)], [xg(x)], \ldots, [x^{k-1} g(x)]$ является базисом линейного пространства $[\mathscr{C}]$. Следовательно, соответству-

ющие этим классам векторы $(g_0, g_1, \ldots, g_{n-k}, 0, \ldots, 0)$, $(0, g_0, g_1, \ldots$ $\ldots, g_{n-k}, 0, \ldots, 0)$, $\ldots$, $(0, \ldots, 0, g_0, g_1, \ldots, g_{n-k})$ образуют базис кода $\mathscr{C}$. □

Определение 6.20. Пусть $\mathscr{C}$ — циклический код с порождающим многочленом $g(x)$. Тогда многочлен

$$h(x) = \frac{x^n - 1}{g(x)}$$

называется проверочным многочленом кода $\mathscr{C}$.

Утверждение 6.20. *Пусть $\mathscr{C}$ — циклический код с проверочным многочленом $h(x)$. Класс $[c(x)]$ из факторкольца $\mathscr{R}_n$ принадлежит идеалу $[\mathscr{C}]$ тогда и только тогда, когда $[c(x)h(x)] = [0]$.*

Доказательство. Если $[c(x)] \in [\mathscr{C}]$, то $c(x)$ делится на порождающий многочлен $g(x)$, т. е. $c(x) = f(x)g(x)$. Следовательно,

$$[c(x)h(x)] = \left[f(x)g(x)\frac{x^n - 1}{g(x)}\right] = [f(x)(x^n - 1)] = [0].$$

С другой стороны, если $[c(x)h(x)] = [0]$, то $x^n - 1$ делит $c(x)h(x)$. Но $x^n - 1 = g(x)h(x)$, поэтому $g(x)$ делит $c(x)$, так что $[c(x)] \in [\mathscr{C}]$. □

Замечание 6.16. 1. Из утверждения 6.20 видно, что проверочный многочлен для циклического кода играет такую же роль, какую играет проверочная матрица в более общем случае линейного кода.

2. Из утверждения 6.19 следует, что проверочный многочлен циклического (n, k)-кода имеет степень $\deg h(x) = k$.

3. Младшие коэффициенты g_0 и h_0 многочленов $g(x)$ и $h(x)$ отличны от нуля. Действительно, при подстановке значения $x = 0$ в тождество $x^n - 1 = g(x)h(x)$ получим $g_0 h_0 = -1$.

Утверждение 6.21. *Пусть $\mathscr{C}$ — циклический (n, k)-код с проверочным многочленом $h(x) = h_0 + h_1 x + \ldots + h_k x^k$. Тогда двойственный код $\mathscr{C}^\perp$ является циклическим $(n, n-k)$-кодом с порождающим многочленом*

$$g^\perp(x) = h_0^{-1} x^k h\left(\frac{1}{x}\right) = h_0^{-1}(h_k + h_{k-1}x + \ldots + h_0 x^k).$$

Доказательство. Докажем вначале, что код $\mathscr{C}^\perp$ является циклическим. Рассмотрим произвольные векторы $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1}) \in \mathscr{C}$

и $\mathbf{d} = (d_0, d_1, \ldots, d_{n-1}) \in \mathscr{C}^{\perp}$. По определению двойственного кода, $\mathbf{c} \perp \mathbf{d}$. Так как код $\mathscr{C}$ — циклический, вектор $\mathbf{c}' = (c_{n-1}, c_0, c_1, \ldots, c_{n-2})$ также принадлежит коду $\mathscr{C}$, поэтому $\mathbf{c}' \perp \mathbf{d}$, или $c_{n-1}d_0 + c_0d_1 + \ldots + c_{n-2}d_{n-1} = 0$. Следовательно, $\mathbf{c} \perp \mathbf{d}'$, где $\mathbf{d}' = (d_1, d_2, \ldots, d_{n-1}, d_0)$ есть циклический сдвиг вектора $\mathbf{d}$ на один шаг влево. Мы получили, что если вектор $\mathbf{d}$ принадлежит коду $\mathscr{C}^{\perp}$, то и его циклический сдвиг $\mathbf{d}'$ также принадлежит коду $\mathscr{C}^{\perp}$, что и требовалось.

Теперь докажем, что порождающий многочлен $g^{\perp}(x)$ кода $\mathscr{C}^{\perp}$ имеет вид, указанный выше. Рассмотрим вначале крайние случаи $k = 0$ и $k = n$. Если $k = 0$, то код $\mathscr{C}$ состоит из одного нулевого вектора, а $\mathscr{C}^{\perp} = V_n(q)$, при этом $g(x) = x^n - 1$, $h(x) = 1$, $g^{\perp}(x) = 1$, что и требовалось. Случай $k = n$ рассматривается аналогично.

Пусть $0 < k < n$. Рассмотрим многочлен

$$\tilde{h}(x) = h_k + h_{k-1}x + \ldots + h_0x^k$$

и соответствующий ему n-мерный вектор

$$\tilde{\mathbf{h}} = (h_k, h_{k-1}, \ldots, h_0, 0, \ldots, 0).$$

Скалярные произведения вектора $\tilde{\mathbf{h}}$ на строки порождающей матрицы G кода $\mathscr{C}$ (см. следствие 6.2) равны

$$\begin{aligned}
(\tilde{\mathbf{h}}, \vec{G}_1) &= g_0h_k + g_1h_{k-1} + \ldots &&= \operatorname{coeff}_{x^k}(g(x)h(x)) = 0,\\
(\tilde{\mathbf{h}}, \vec{G}_2) &= g_0h_{k-1} + g_1h_{k-2} + \ldots &&= \operatorname{coeff}_{x^{k-1}}(g(x)h(x)) = 0,\\
&\ldots\\
(\tilde{\mathbf{h}}, \vec{G}_k) &= g_0h_1 + g_1h_0 &&= \operatorname{coeff}_{x}(g(x)h(x)) = 0.
\end{aligned}$$

Так как вектор $\tilde{\mathbf{h}}$ ортогонален всем строкам матрицы G, этот вектор лежит в коде $\mathscr{C}^{\perp}$, или, что то же самое, $[\tilde{h}(x)] \in [\mathscr{C}^{\perp}]$.

С другой стороны, $\dim \mathscr{C}^{\perp} = n - k$, поэтому степень порождающего многочлена $g^{\perp}(x)$ кода $\mathscr{C}^{\perp}$ равна k. Унитарный многочлен $g^{\perp}(x)$ делит $\tilde{h}(x)$, и степени этих многочленов равны, следовательно, $g^{\perp}(x) = h_0^{-1}\tilde{h}(x)$. □

Из доказанного утверждения и следствия 6.2 вытекает

Следствие 6.3. *Пусть $\mathscr{C}$ — циклический (n, k)-код с $k < n$ и прове-*

рочным многочленом $h(x) = h_0 + h_1 x + \ldots + h_k x^k$. *Тогда матрица*

$$H = \begin{pmatrix} h_k & h_{k-1} & \ldots & \ldots & h_0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & h_k & \ldots & \ldots & h_1 & h_0 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & h_k & \ldots & \ldots & \ldots & \ldots & h_1 & h_0 \end{pmatrix}$$

размера $(n-k) \times n$ *является проверочной матрицей кода* $\mathscr{C}$.

Замечание 6.17. Поскольку младший коэффициент g_0 порождающего многочлена $g(x)$ отличен от нуля (см. п. 3 замечания 6.16), порождающая матрица G циклического кода (см. следствие 6.2) элементарными преобразованиями строк может быть приведена к виду, указанному в определении 6.10, другими словами, циклический код является систематическим. Здесь нам будет удобнее полагать, что для систематического циклического (n, k)-кода в кодовом слове $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1})$ информационные символы $(c_{n-k}, \ldots, c_{n-1})$ располагаются *справа*, а проверочные символы $(c_0, c_1, \ldots, c_{n-k-1})$ — слева. Действительно, старший коэффициент g_k порождающего многочлена $g(x)$ равен 1, поэтому порождающую матрицу G эквивалентными преобразованиями строк можно привести к виду $G = (PI_k)$.

Опишем алгоритм кодирования для систематического циклического кода. С этой целью введём *информационный многочлен*

$$a(x) = c_{n-k} + c_{n-k+1} x + \ldots + c_{n-1} x^{k-1}$$

и *корректирующий многочлен*

$$r(x) = c_0 + c_1 x + \ldots + c_{n-k-1} x^{n-k-1}.$$

Тогда *кодовый многочлен* $c(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$ представляется в виде

$$c(x) = r(x) + x^{n-k} a(x).$$

Так как порождающий многочлен $g(x)$ делит $c(x)$, $c(x) = g(x) f(x)$, получаем формулу

$$x^{n-k} a(x) = g(x) f(x) - r(x),$$

из которой вытекает *правило кодирования*: чтобы по заданному информационному многочлену $a(x)$ найти кодовый многочлен $c(x)$, нужно найти корректирующий многочлен $r(x)$, представляющий

собой взятый со знаком минус остаток от деления многочлена $x^{n-k}a(x)$ на порождающий многочлен $g(x)$. Далее остаток от деления произвольного многочлена $t(x)$ на многочлен $g(x)$ будем обозначать в виде $t_g(x)$.

Декодер $\mathscr{D}_\rho$ (в ближайшее кодовое слово) в случае циклического кода также можно описать на языке многочленов. Предположим, что $\mathbf{c}$ — переданное в канал связи кодовое слово циклического кода $\mathscr{C}$, $\mathbf{e}$ — вектор ошибок и $\mathbf{y} = \mathbf{c} + \mathbf{e}$ — принятый вектор. Этим векторам соответствуют классы $[c(x)]$, $[e(x)]$ и $[y(x)]$ из факторкольца $\mathscr{R}_n$. Вначале декодер вычисляет остаток $y_g(x)$ (очевидно, остатки $y_g(x)$ и $e_g(x)$ равны). Если остаток $y_g(x)$ равен нулю, то вектор $\mathbf{y}$ принадлежит коду $\mathscr{C}$, декодер не обнаруживает ошибок и в качестве результата декодирования выдаёт $\mathscr{D}_\rho(\mathbf{y}) = \mathbf{y}$. Если же $y_g(x) \neq 0$, то $\mathbf{y} \notin \mathscr{C}$, декодер обнаруживает наличие ошибок в канале связи и пытается их исправить. Для этого ищется многочлен $e'(x)$ с наименьшим числом ненулевых членов, удовлетворяющий условию $y_g(x) = e'_g(x)$, и результатом декодирования объявляется вектор, соответствующий многочлену $y(x) - e'(x)$.

Замечание 6.18. Отметим, что для некоторых циклических кодов известны эффективные алгоритмы нахождения указанного выше многочлена $e'(x)$, не требующие перебора большого числа вариантов. Подробное изложение теории эффективных алгоритмов декодирования для циклических кодов не входит в наше учебное пособие, так как это потребовало бы существенно расширить его; по этому вопросу заинтересованный читатель может обратиться к книгам по алгебраической теории кодирования, приведённым в списке литературы.

Пример 6.5. Пусть $n = 7$ и $q = 2$. Разложение многочлена $x^7 - 1 = x^7 \oplus 1$ на неприводимые над полем $GF(2)$ множители имеет вид:

$$x^7 \oplus 1 = (x \oplus 1)(x^3 \oplus x \oplus 1)(x^3 \oplus x^2 \oplus 1).$$

Так как порождающий многочлен циклического кода должен делить многочлен $x^7 \oplus 1$, имеется всего $2^3 = 8$ двоичных циклических кодов длины 7. Рассмотрим циклический код $\mathscr{C}$ с порождающим многочленом $g(x) = 1 \oplus x^2 \oplus x^3$. Для этого кода размерность k равна 4,

корректирующий многочлен равен

$$h(x) = (1 \oplus x)(1 \oplus x \oplus x^3) = 1 \oplus x^2 \oplus x^3 \oplus x^4,$$

а в качестве порождающей и проверочной матриц можно взять

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

По теореме 6.1 (с. 173) минимальное кодовое расстояние равно $d = 3$. Действительно, все столбцы матрицы H различны, так что любые два столбца линейно независимы, и при этом имеются три линейно зависимых столбца (например, первый, третий и четвёртый). Следовательно, декодер $\mathscr{D}_\rho$ правильно исправляет любые одиночные ошибки.

Выпишем все возможные многочлены $e(x)$, соответствующие одиночным ошибкам, и остатки от деления этих многочленов на порождающий многочлен $g(x)$:

$e(x)$	$e_g(x)$
1	1
x	x
x^2	x^2
x^3	$1 \oplus x^2$
x^4	$1 \oplus x \oplus x^2$
x^5	$1 \oplus x$
x^6	$x \oplus x^2$

Закодируем информационный вектор $\mathbf{a} = (1110)$. Для этого информационный многочлен $a(x) = 1 \oplus x \oplus x^2$ умножаем на $x^{n-k} = x^3$, результат $x^3 \oplus x^4 \oplus x^5$ делим с остатком на $g(x)$ и получаем неполное частное $f(x) = 1 \oplus x^2$ и остаток $r(x) = 1$. Кодовому многочлену $c(x) = x^3 a(x) \oplus r(x) = 1 \oplus x^3 \oplus x^4 \oplus x^5$ отвечает кодовое слово $\mathbf{c} = (1001110)$.

Далее, предположим, что при передаче по каналу связи произошло искажение четвёртого символа кодового слова, другими словами, вектор ошибок равен $\mathbf{e} = (0001000)$, а многочлен ошибок равен $e(x) = x^3$. Принятому вектору $\mathbf{y} = (1000110)$ отвечает многочлен

$y(x) = 1 \oplus x^4 \oplus x^5$, который не делится на $g(x)$, так что декодер обнаруживает наличие ошибок и пытается их исправить. Для этого вычисляется остаток $y_g(x) = 1 \oplus x^2$ и по таблице находится соответствующий этому остатку многочлен ошибок x^3, т. е. происходит правильное декодирование.

Важная особенность циклических кодов состоит в том, что для них имеется ещё один способ задания — при помощи корней порождающего многочлена. Мы рассмотрим этот способ и покажем, что он позволяет строить циклические коды, для которых минимальное кодовое расстояние не меньше заданной границы, что даёт возможность гарантированно исправлять заданное число ошибок.

Одно простое замечание сразу даёт ключ к пониманию нового способа задания циклического кода. Пусть $\mathscr{C}$ — циклический код с порождающим многочленом $g(x)$ и β — корень многочлена $g(x)$ в некотором расширении $GF(q^r)$ поля $GF(q)$. Поскольку порождающий многочлен $g(x)$ делит любой кодовый многочлен $c(x)$, элемент β является также корнем многочлена $c(x)$. Кроме того, поскольку $g(x)$ делит $x^n - 1$, справедливо равенство $\beta^n = 1$, поэтому порядок $\mathrm{ord}(\beta)$ элемента β делит n.

Утверждение 6.22. *Пусть $\beta_1, \ldots, \beta_s$ — элементы поля $GF(q^r)$ и n есть наименьшее общее кратное порядков $\mathrm{ord}(\beta_1), \ldots, \mathrm{ord}(\beta_s)$. Обозначим через $[\mathscr{C}]$ множество всех классов $[c(x)]$ из факторкольца $\mathscr{R}_n$, для которых все элементы $\beta_1, \ldots, \beta_s$ являются корнями многочлена $c(x)$:*

$$[\mathscr{C}] = \{[c(x)] \in \mathscr{R}_n \mid c(\beta_1) = \ldots = c(\beta_s) = 0\}.$$

Тогда множество $\mathscr{C}$ векторов из $V_n(q)$, соответствующих классам из $[\mathscr{C}]$, является циклическим кодом.

Доказательство. Вначале покажем, что множество $[\mathscr{C}]$ определено корректно — если элемент β_i является корнем многочлена $c(x)$, то β_i является корнем и для всех многочленов $f(x)$ из класса $[c(x)]$. Действительно, многочлен $x^n - 1$ делит разность $f(x) - c(x)$, т. е. $f(x) = c(x) + t(x)(x^n - 1)$. Так как $\mathrm{ord}(\beta_i)$ делит n, $\beta_i^n = 1$, поэтому

$$f(\beta_i) = c(\beta_i) + t(\beta_i)(\beta_i^n - 1) = 0.$$

Пусть $m_i(x)$ — многочлен над полем $GF(q)$, являющийся минимальным многочленом элемента $\beta_i \in GF(q^r)$, и $g(x)$ — наименьшее общее кратное многочленов $m_1(x), \ldots, m_s(x)$. Очевидно, что элементы $\beta_1, \ldots, \beta_s$ являются корнями многочлена $g(x)$. Покажем, что множество $[\mathscr{C}]$ совпадает с идеалом $\langle [g(x)] \rangle$.

Если $[c(x)] \in \langle [g(x)] \rangle$, то $g(x)$ делит $c(x)$, поэтому $c(\beta_i) = 0$, так что $\langle [g(x)] \rangle \subseteq [\mathscr{C}]$. С другой стороны, если для некоторого многочлена $c(x)$ над полем $GF(q)$ все элементы $\beta_1, \ldots, \beta_s$ являются корнями, то минимальные многочлены $m_i(x)$ делят $c(x)$, поэтому и их наименьшее общее кратное $g(x)$ также делит $c(x)$. Следовательно, $c(x) \in \langle [g(x)] \rangle$, так что $[\mathscr{C}] \subseteq \langle [g(x)] \rangle$. Поскольку множество $[\mathscr{C}]$ является идеалом в факторкольце $\mathscr{R}_n$, соответствующее множество $\mathscr{C}$ векторов из $V_n(q)$ является циклическим кодом. □

Теорема 6.3. *Пусть $\mathscr{C}$ — q-ичный циклический (n, k)-код с порождающим многочленом $g(x)$ и $\alpha^{i_1}, \ldots, \alpha^{i_{n-k}}$ — все корни многочлена $g(x)$ в некотором расширении $GF(q^r)$ поля $GF(q)$, где α есть примитивный элемент поля $GF(q^r)$ и $i_1 \leqslant i_2 \leqslant \ldots \leqslant i_{n-k}$. Тогда минимальное кодовое расстояние $d_{\mathscr{C}}$ больше, чем длина наибольшего интервала последовательных (с шагом 1) чисел в ряду $i_1, i_2, \ldots, i_{n-k}$.*

Доказательство. Из условия теоремы и утверждения 6.22 вытекает, что вектор $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1})$ из $V_n(q)$ принадлежит коду $\mathscr{C}$ в том и только том случае, когда все элементы $\alpha^{i_1}, \ldots, \alpha^{i_{n-k}}$ являются корнями многочлена $c(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$, т. е.

$$c(\alpha^{i_1}) = \ldots = c(\alpha^{i_{n-k}}) = 0.$$

Это условие можно записать в матричном виде

$$\mathbf{c} \cdot F^T = \mathbf{0},$$

где $\mathbf{0}$ есть нулевая вектор-строка из $V_{n-k}(q)$, а матрица F над полем $GF(q^r)$ имеет вид

$$F = \begin{pmatrix} 1 & \alpha^{i_1} & \alpha^{2i_1} & \ldots & \alpha^{(n-1)i_1} \\ 1 & \alpha^{i_2} & \alpha^{2i_2} & \ldots & \alpha^{(n-1)i_2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & \alpha^{i_{n-k}} & \alpha^{2i_{n-k}} & \ldots & \alpha^{(n-1)i_{n-k}} \end{pmatrix}.$$

Матрицу F можно рассматривать как некий аналог проверочной матрицы для кода $\mathscr{C}$, но не над исходным полем $GF(q)$, а над его

расширением $GF(q^r)$ (однако мы не называем матрицу F проверочной, поскольку её строки могут совпадать, а в проверочной матрице строки должны быть линейно независимыми). Для матрицы F справедливо утверждение, аналогичное теореме 6.1 (с. 173), с заменой H на F: если любые d столбцов матрицы F линейно независимы над полем $GF(q)$, то $d_{\mathscr{C}} > d$.

Предположим, что наибольший интервал последовательных (с шагом 1) чисел в ряду $i_1, i_2, \ldots, i_{n-k}$ имеет вид $m_0, m_0+1, \ldots, m_0 + d_0 - 1$. Нам нужно доказать неравенство $d_{\mathscr{C}} > d_0$. Для этого достаточно установить, что любые d_0 столбцов матрицы F линейно независимы над полем $GF(q)$.

Рассмотрим подматрицу F' матрицы F, расположенную в строках $m_0, m_0+1, \ldots, m_0+d_0-1$:

$$F' = \begin{pmatrix} 1 & \alpha^{m_0} & \alpha^{2m_0} & \ldots & \alpha^{(n-1)m_0} \\ 1 & \alpha^{m_0+1} & \alpha^{2(m_0+1)} & \ldots & \alpha^{(n-1)(m_0+1)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & \alpha^{m_0+d_0-1} & \alpha^{2(m_0+d_0-1)} & \ldots & \alpha^{(n-1)(m_0+d_0-1)} \end{pmatrix}.$$

Если мы докажем, что любые d_0 столбцов матрицы F' линейно независимы над $GF(q)$, то это же будет верно и для матрицы F. Возьмём в матрице F' произвольные d_0 столбцов, скажем, с номерами $j_1 < j_2 < \ldots < j_{d_0}$, и рассмотрим квадратную подматрицу

$$F'' = \begin{pmatrix} \alpha^{j_1 m_0} & \alpha^{j_2 m_0} & \ldots & \alpha^{j_{d_0} m_0} \\ \alpha^{j_1 (m_0+1)} & \alpha^{j_2 (m_0+1)} & \ldots & \alpha^{j_{d_0} (m_0+1)} \\ \ldots & \ldots & \ldots & \ldots \\ \alpha^{j_1 (m_0+d_0-1)} & \alpha^{j_2 (m_0+d_0-1)} & \ldots & \alpha^{j_{d_0} (m_0+d_0-1)} \end{pmatrix}.$$

Покажем, что определитель $\det F''$ отличен от нуля. Вынося из первого столбца элемент $\alpha^{j_1 m_0}$, из второго столбца элемент $\alpha^{j_2 m_0}$ и так далее, получим выражение

$$\det F'' = \alpha^{(j_1+j_2+\ldots+j_{d_0})m_0} \cdot W(\alpha^{j_1}, \ldots, \alpha^{j_{d_0}}),$$

где

$$W(\alpha^{j_1}, \ldots, \alpha^{j_{d_0}}) = \det \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \alpha^{j_1} & \alpha^{j_2} & \ldots & \alpha^{j_{d_0}} \\ \alpha^{2j_1} & \alpha^{2j_2} & \ldots & \alpha^{2j_{d_0}} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \alpha^{j_1(d_0-1)} & \alpha^{j_2(d_0-1)} & \ldots & \alpha^{j_{d_0}(d_0-1)} \end{pmatrix}$$

есть известный определитель Вандермонда. Так как элементы $\alpha^{j_1}, \ldots$ $\ldots, \alpha^{j_{d_0}}$ различны, $W(\alpha^{j_1}, \ldots, \alpha^{j_{d_0}}) \neq 0$, поэтому $\det F'' \neq 0$, что и требовалось. □

Теорема 6.3 даёт возможность явно указать способ построения циклических кодов, для которых минимальное кодовое расстояние будет больше заданной границы. Такой способ был открыт независимо в работах Р. Боуза, Д. Рой—Чоудхури и А. Хоквингема, эти коды называют БЧХ-кодами.

Определение 6.21. Выберем целые $m_0, d_0, r \geqslant 1$. Пусть α — примитивный элемент конечного поля $GF(q^r)$, $g(x)$ — унитарный многочлен над $GF(q)$ наименьшей степени, для которого элементы

$$\alpha^{m_0}, \alpha^{m_0+1}, \ldots, \alpha^{m_0+d_0-1}$$

поля $GF(q^r)$ являются корнями, и n — наименьшее общее кратное порядков указанных элементов, $n = \text{НОК}\{\text{ord}(\alpha^i), m_0 \leqslant i \leqslant m_0 + d_0 - 1\}$. БЧХ-кодом с параметрами (m_0, d_0) называется q-ичный циклический код длины n с порождающим многочленом $g(x)$. Число $d_0 + 1$ называется конструктивным кодовым расстоянием БЧХ-кода.

Замечание 6.19. Из определения БЧХ-кода и теоремы 6.3 ясно, что минимальное кодовое расстояние $d_{\mathscr{C}}$, называемое также истинным кодовым расстоянием, не меньше конструктивного кодового расстояния.

Наиболее важный для практических применений класс БЧХ-кодов получается при $q = 2$, $m_0 = 1$, $d_0 = 2t_0$. Для такого двоичного циклического кода $d_{\mathscr{C}} \geqslant 2t_0 + 1$, поэтому декодер в ближайшее кодовое слово правильно исправляет любые комбинации из t_0 или

меньшего числа ошибок. Свойства двоичных БЧХ-кодов описывает следующая

Теорема 6.4. *Для любых $r, t_0 \geqslant 1$ таких, что $t_0 < n/2$, $n = 2^r - 1$, существует двоичный циклический код длины n с конструктивным кодовым расстоянием $2t_0 + 1$, имеющий не более rt_0 проверочных символов.*

Доказательство. Вначале напомним одно простое свойство: если $f(x)$ — многочлен над $GF(2)$ с корнем $\beta \in GF(2^r)$, то элемент β^2 также является корнем $f(x)$. Действительно, пусть $f(x) = f_0 \oplus f_1 x \oplus \ldots \oplus f_m x^m$. Возводя в квадрат равенство $f(\beta) = 0$, получим

$$0 = (f(\beta))^2 = \sum_{i=0}^{m} f_i^2 \beta^{2i} \oplus 2 \sum_{i<j} f_i f_j \beta^{i+j} = \sum_{i=0}^{m} f_i \beta^{2i} = f(\beta^2).$$

Рассмотрим двоичный БЧХ-код с параметрами $m_0 = 1$, $d_0 = 2t_0$. Порождающий многочлен $g(x)$ этого кода имеет корни $\alpha, \alpha^2, \alpha^3, \ldots, \alpha^{2t_0}$, где α — примитивный элемент поля $GF(2^r)$. Длина кодового слова равна $n = \text{НОК}\,\{\text{ord}(\alpha^i),\ 1 \leqslant i \leqslant 2t_0\} = \text{ord}(\alpha) = 2^r - 1$.

Пусть $m_i(x)$ — многочлен над $GF(2)$, являющийся минимальным многочленом элемента $\alpha_i \in GF(2^r)$; тогда $\deg m_i(x) \leqslant r$. Как видно из доказательства утверждения 6.22, порождающим многочленом БЧХ-кода является

$$g(x) = \text{НОК}\,\{m_i(x),\ \ 1 \leqslant i \leqslant 2t_0\}.$$

Из замечания в начале доказательства следует, что $m_{2i}(x)$ делит $m_i(x)$, поэтому при вычислении наименьшего общего кратного $g(x)$ достаточно взять только многочлены $m_i(x)$ с нечётными номерами, т. е.

$$g(x) = \text{НОК}\,\{m_1(x), m_3(x), \ldots, m_{2t_0-1}(x)\}.$$

Так как $g(x)$ есть наименьшее общее кратное t_0 многочленов, степени которых не превосходят r, справедливо неравенство $\deg g(x) \leqslant rt_0$. С другой стороны, по утверждению 6.19, $\deg g(x) = n - k$. Согласно определению систематического кода, число $n - k$ равно числу проверочных символов в кодовом слове. □

Пример 6.6. Построим двоичный БЧХ-код с конструктивным кодовым расстоянием, равным 5. Для этого рассмотрим конечное поле

$GF(16)$ как расширение поля $GF(2)$ корнем α неприводимого над $GF(2)$ многочлена $m(x) = 1 \oplus x \oplus x^4$, при этом поле $GF(16)$ представляем как факторкольцо $GF(2)[x]/m(x)$, и $\alpha = [x]_{m(x)}$. Представления для элементов поля в виде степеней элемента α, элементов факторкольца $GF(2)[x]/m(x)$ и 4-мерных двоичных векторов приведены в следующей таблице.

0	0	0	0000
1	α	x	0100
2	α^2	x^2	0010
3	α^3	x^3	0001
4	α^4	$1 \oplus x$	1100
5	α^5	$x \oplus x^2$	0110
6	α^6	$x^2 \oplus x^3$	0011
7	α^7	$1 \oplus x \oplus x^3$	1101
8	α^8	$1 \oplus x^2$	1010
9	α^9	$x \oplus x^3$	0101
10	α^{10}	$1 \oplus x \oplus x^2$	1110
11	α^{11}	$x \oplus x^2 \oplus x^3$	0111
12	α^{12}	$1 \oplus x \oplus x^2 \oplus x^3$	1111
13	α^{13}	$1 \oplus x^2 \oplus x^3$	1011
14	α^{14}	$1 \oplus x^3$	1001
15	$\alpha^{15} = 1$	1	1000

Рассмотрим двоичный БЧХ-код с параметрами $m_0 = 1$, $d_0 = 4$ как циклический код длины $2^4 - 1 = 15$ с корнями α, α^2, α^3, α^4. Многочлен $m(x)$ имеет корни α, α^2, α^4 и α^8, поэтому порождающий многочлен $g(x)$ равен наименьшему общему кратному многочленов $m(x)$ и $m_3(x)$, где $m_3(x)$ — минимальный многочлен элемента α^3. Многочлен $m_3(x)$ имеет также корни α^6, α^{12}, $\alpha^{24} = \alpha^9$, поэтому

$$m_3(x) = (x - \alpha^3)(x - \alpha^6)(x - \alpha^9)(x - \alpha^{12}) = 1 \oplus x \oplus x^2 \oplus x^3 \oplus x^4.$$

Следовательно, порождающий многочлен кода равен

$$g(x) = \text{НОК}\{m(x), m_3(x)\} = m(x)m_3(x) = 1 \oplus x^4 \oplus x^6 \oplus x^7 \oplus x^8.$$

Порождающему многочлену $g(x)$ соответствует кодовое слово

$$\mathbf{g} = (1000101110000000)$$

веса 5, так что для рассматриваемого кода истинное кодовое расстояние и конструктивное кодовое расстояние совпадают, $d_{\mathscr{C}} = d_0 + 1 = 5$. Следовательно, декодер в ближайшее кодовое слово правильно исправляет любые одиночные и двойные ошибки.

§ 6.4. Коды Хэмминга

Двоичный код Хэмминга является исторически одной из первых общих конструкций кодов для обнаружения и исправления ошибок. В нашем пособии этот код будет хорошим примером применения изложенных выше общих понятий и результатов. Всюду далее $q = 2$.

Для произвольного натурального r положим $n = 2^r - 1$ и рассмотрим матрицу H_r размера $r \times n$ над полем $GF(2)$, составленную из всех n различных ненулевых векторов-столбцов размера r, выписанных в произвольном порядке. Например,

$$H_1 = (1), \quad H_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad H_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Определение 6.22. Кодом Хэмминга $\mathscr{H}_r$ называется двоичный линейный код длины n с проверочной матрицей H_r, т. е.

$$\mathscr{H}_r = \{\mathbf{c} \in V_n(2) \mid \mathbf{c} \cdot H_r^T = \mathbf{0}\}.$$

Например, код $\mathscr{H}_2$ состоит из всех двоичных векторов длины $n = 3$, ортогональных обеим строкам матрицы H_2. Если $\mathbf{c} = (c_0, c_1, c_2)$, то условие $\mathbf{c} \cdot H_r^T = \mathbf{0}$ принимает вид $c_1 \oplus c_2 = 0$, $c_0 \oplus c_2 = 0$, откуда $c_0 = c_1 = c_2$, так что $\mathscr{H}_2 = \{(000), (111)\}$ (см. пример 6.3, с. 180). Отметим, что циклический код в примере 6.5 на с. 194 также является кодом Хэмминга с $r = 3$.

Замечание 6.20. Поскольку порядок следования столбцов в матрице H_r не фиксирован, мы фактически определили не один код, а класс эквивалентных кодов, отличающихся друг от друга только перестановками координат в кодовых словах. Некоторые перестановки координат в кодовых словах кода Хэмминга оставляют код в целом неизменным, например, при $r = 2$ все шесть перестановок координат не меняют код $\{000, 111\}$. Очевидно, что все эквивалентные коды Хэмминга имеют одно и то же минимальное кодовое

расстояние и поэтому одинаковую способность обнаруживать и исправлять ошибки. Однако, с точки зрения практической реализации кодера и декодера, некоторые порядки следования столбцов в проверочной матрице H_r более предпочтительны — такие специальные упорядочения столбцов будут рассмотрены ниже.

Утверждение 6.23. *Параметры кода Хэмминга $\mathscr{H}_r$ равны* $(n, k, d) = (2^r - 1, 2^r - r - 1, 3)$.

Доказательство. Равенство $n = 2^r - 1$ верно по определению. Строки матрицы H_r линейно независимы, поэтому число строк этой матрицы равно $n - k$, где $k = \dim \mathscr{H}_r$. Из равенства $n - k = r$ получаем $k = n - r = 2^r - r - 1$. Минимальное кодовое расстояние равно $d = 3$ в силу теоремы 6.1 (с. 173). Действительно, так как все столбцы матрицы H_r различны, любые два столбца линейно независимы, и при этом имеются три линейно зависимых столбца (например, столбцы, в которых первые две координаты равны 10, 01, 11, а остальные координаты совпадают). Фактически это рассуждение уже проведено при рассмотрении примера 6.5. □

Замечание 6.21. Так как $d = 3$, по следствию 6.1 (с. 168) декодер $\mathscr{D}_\rho$ обнаруживает любые двойные ошибки и правильно исправляет любые одиночные ошибки.

Покажем, что при определённом порядке следования столбцов в проверочной матрице H_r код Хэмминга является циклическим, что позволяет использовать изложенные выше понятия и результаты теории циклических кодов.

Утверждение 6.24. *Семейство эквивалентных кодов Хэмминга $\mathscr{H}_r$ содержит циклический код с порождающим многочленом $g(x)$ над полем $GF(2)$, который является минимальным многочленом примитивного элемента α поля $GF(2^r)$.*

Доказательство. Пусть α — примитивный элемент поля $GF(2^r)$ и $g(x) \in GF(2)[x]$ — минимальный многочлен элемента α, $\deg g(x) = r$. Поле $GF(2^r)$ рассмотрим как линейное пространство над $GF(2)$, а в качестве базиса этого линейного пространства — множество элементов $1, \alpha, \alpha^2, \ldots, \alpha^{r-1}$. Для произвольного элемента $\beta \in GF(2^r)$ обозначим через $\varphi(\beta) = \beta^{\downarrow}$ r-мерный вектор-столбец, составленный

из коэффициентов разложения элемента β по указанному базису. Отображение $\varphi: GF(2^r) \to V_n(2)$ является изоморфизмом линейных пространств над полем $GF(2)$.

Обозначим $n = 2^r - 1$ и рассмотрим матрицу

$$H = ((1)^{\downarrow}, (\alpha)^{\downarrow}, (\alpha^2)^{\downarrow}, \ldots (\alpha^{n-1})^{\downarrow})$$

размера $r \times n$. Поскольку $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$ — все ненулевые элементы поля $GF(2^r)$, в матрице H по одному разу записаны (в некотором порядке) все ненулевые r-мерные вектор-столбцы над $GF(2)$; следовательно, матрица H является проверочной матрицей кода Хэмминга $\mathscr{H}_r$.

Вектор $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1}) \in V_n(2)$ принадлежит коду $\mathscr{H}_r$ тогда и только тогда, когда $\mathbf{c} \cdot H^T = \mathbf{0}$, или

$$c_0(1)^{\downarrow} \oplus c_1(\alpha)^{\downarrow} \oplus c_2(\alpha^2)^{\downarrow} \oplus \ldots \oplus c_{n-1}(\alpha^{n-1})^{\downarrow} = 0^{\downarrow}.$$

Применяя к этому равенству изоморфизм φ^{-1}, получим в поле $GF(2^r)$ равенство

$$c_0 + c_1\alpha + \ldots + c_{n-1}\alpha^{n-1} = 0,$$

которое означает, что элемент α является корнем многочлена $c(x) = c_0 \oplus c_1 x \oplus \ldots \oplus c_{n-1}x^{n-1}$ над полем $GF(2)$. Поскольку $g(x)$ — минимальный многочлен элемента α, $c(x)$ делится на $g(x)$.

Итак, вектор $\mathbf{c}$ принадлежит коду $\mathscr{H}_r$ тогда и только тогда, когда многочлен $c(x)$ делится на $g(x)$. Это означает, что $\mathscr{H}_r$ является циклическим кодом с порождающим многочленом $g(x)$. □

Например, при $r = 3$ можно получить код Хэмминга в циклической форме, если взять $g(x) = 1 \oplus x^2 \oplus x^3$ и, соответственно,

$$H_3 = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Кодирование для кода Хэмминга $\mathscr{H}_r$ удобно описать, представляя этот код как систематический. Действительно, мы можем выбирать произвольный порядок следования столбцов в матрице H_r, в частности, можно сделать так, чтобы последние r столбцов в H_r образовали единичную матрицу, и затем получить нужный вид порождающей матрицы G_r в силу утверждения 6.10 (с. 176). Например, при

$r = 3$ для (7, 4, 3)-кода Хэмминга $\mathscr{H}_3$

$$H_3 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}, \quad G_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Если в кодовом слове $\mathbf{c} = (c_0, c_1, c_2, c_3, c_4, c_5, c_6)$ первые четыре символа c_0, c_1, c_2, c_3 являются информационными, то три проверочных символа c_4, c_5, c_6 вычисляются по формулам

$$c_4 = c_1 \oplus c_2 \oplus c_3, \quad c_5 = c_0 \oplus c_2 \oplus c_3, \quad c_6 = c_0 \oplus c_1 \oplus c_3.$$

Предположим, что кодовые слова кода Хэмминга $\mathscr{H}_r$ передаются по двоичному (не обязательно симметричному) каналу связи, т. е. входной алфавит $\mathscr{X}$ и выходной алфавит $\mathscr{Y}$ совпадают с $GF(2)$. Опишем подробнее алгоритм декодирования $\mathscr{D}_\rho$ в ближайшее кодовое слово. Пусть $\mathbf{c}$ — переданное кодовое слово, $\mathbf{e}$ — вектор ошибок в канале связи и $\mathbf{y} = \mathbf{c} \oplus \mathbf{e}$ — принятый вектор. Вычисляя синдром принятого вектора (см. определение 6.13 на с. 179), получим

$$s(\mathbf{y}) = \mathbf{y} \cdot H_r^T = (\mathbf{c} \oplus \mathbf{e}) \cdot H_r^T = \mathbf{e} \cdot H_r^T.$$

Если произошло искажение только одного символа в кодовом слове, скажем i-го, то в векторе ошибок i-я координата равна 1, а все остальные координаты равны нулю. В этом случае синдром $s(\mathbf{y}) = \mathbf{e} \cdot H_r^T = (H_r)_i^{\downarrow}$ представляет собой i-й столбец проверочной матрицы. Поскольку в проверочной матрице все столбцы различны, значение синдрома однозначно указывает на номер искажённого символа и тем самым позволяет исправить этот символ.

Наличие такого простого алгоритма декодирования для исправления одной ошибки является причиной достаточно широкого практического применения кода Хэмминга. Отметим также, что с точки зрения практической реализации описанного выше декодера удобным является такой порядок следования столбцов в проверочной матрице H_r, при котором каждый i-й столбец является r-разрядным двоичным представлением числа i. При таком порядке столбцов значение синдрома $s(\mathbf{y}) = (H_r)_i^{\downarrow}$ является двоичным представлением номера

искажённой координаты кодового слова. Например, если $r = 3$,

$$H_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix},$$

и при передаче кодового слова $\mathbf{c} = (1111111)$ (проверьте, что это действительно кодовое слово!) произошло искажение третьего символа, то для принятого вектора $\mathbf{y} = (1101111)$ синдром $s(\mathbf{y}) = (011)^T$ есть двоичное представление числа 3.

Утверждение 6.25. *Код Хэмминга $\mathscr{H}_r$ является совершенным.*

Доказательство. Необходимо убедиться и том, что неравенство границы Хэмминга (см. утверждение 6.5 на с. 183) обращается в равенство. Для этого достаточно подставить $q = 2$, $n = 2^r - 1$, $k = 2^r - r - 1$ и $t = 1$. □

Замечание 6.22. Доказанное утверждение означает, что шары радиуса 1 с центрами в кодовых словах кода Хэмминга не пересекаются и в совокупности покрывают всё пространство $V_n(2)$. Можно также сказать, что указанные шары являются решающими областями (см. определение 5.2 на с. 137) декодера $\mathscr{D}_\rho$. Отсюда следует, что декодер $\mathscr{D}_\rho$ правильно исправляет любые одиночные искажения, но при двойных искажениях принимает заведомо неправильное решение. Действительно, если при передаче кодового слова $\mathbf{c}$ были искажены два символа, то принятый вектор $\mathbf{y}$ не принадлежит шару радиуса 1 с центром в $\mathbf{c}$, и поэтому $\mathbf{y}$ попадает в шар радиуса 1 с центром в *другом* кодовом слове $\mathbf{c}'$. Следовательно, декодер принимает решение в пользу $\mathbf{c}'$ и ошибается.

Некоторое улучшение декодера при двойных искажениях символов достигается при переходе к так называемому *расширенному* коду Хэмминга $\mathscr{H}_r'$. Для этого кода проверочная матрица H_r' получается из матрицы H_r добавлением одного нулевого столбца (этот столбец становится первым) и затем одной строки из всех единиц (эта строка

также становится первой). Например,

$$H_3' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Кодовое слово расширенного кода $\mathscr{H}_r'$ получается из кодового слова кода $\mathscr{H}_r$ добавлением одного проверочного символа, равного сумме по модулю 2 всех остальных символов (говорят также, что добавляется одна общая проверка на чётность). Для расширенного кода нетрудно доказать (так же, как в утверждении 6.23), что минимальное кодовое расстояние равно $d = 4$, поэтому декодер $\mathscr{D}_\rho$ обнаруживает любые комбинации из трёх или меньшего числа ошибок и правильно исправляет любые одиночные ошибки. Если же происходит искажение двух (или любого и чётного числа) символов, то этот факт выявляется декодером, так как первый символ синдрома $S(\mathbf{y})$ равен нулю, и затем декодер может вычислить кодовые слова, находящиеся на равных расстояниях от принятого вектора.

Утверждение 6.26. *Предположим, что кодовые слова кода $\mathscr{H}_r$ передаются по двоичному симметричному каналу с параметром $p \leqslant 1/2$, а в качестве декодера используется декодер $\mathscr{D}_\rho$. Тогда средняя вероятность ошибочного декодирования не зависит от распределения на коде $\mathscr{H}_r$ и равна*

$$\lambda = 1 - (1-p)^n - np(1-p)^{n-1}.$$

Доказательство. Как показано выше, для кода Хэмминга и декодера $\mathscr{D}_\rho$ решающими областями являются шары радиуса 1 с центрами в кодовых словах. Поэтому для любого кодового слова $\mathbf{c}_i$ с номером i, $1 \leqslant i \leqslant 2^k$, по определению двоичного симметричного канала и согласно утверждению 5.1 (с. 139), условная вероятность ошибочного декодирования равна

$$\begin{aligned} \lambda_i &= 1 - \mathsf{P}(\rho(X^n, Y^n) \leqslant 1 / X^n = \mathbf{c}_i) = \\ &= 1 - (1-p)^n - np(1-p)^{n-1}. \end{aligned}$$

Поскольку λ_i не зависит от i, при любом распределении на множестве кодовых слов получим $\lambda = 1 - (1-p)^n - np(1-p)^{n-1}$. □

§ 6.5. Задачи

6.1. Исследовать зависимость вероятности

$$\lambda = 1 - (1-p)^n - np(1-p)^{n-1}$$

от параметров n и p. При фиксированном n найти асимптотическую формулу для λ при $p \to 0$.

6.2. По двоичному симметричному каналу передаются кодовые слова двоичного линейного кода с проверкой на чётность. Вычислить вероятность того, что декодер $\mathscr{D}_\rho$ не обнаружит ошибки.

6.3. Найти минимальное кодовое расстояние для следующих кодов:

а) {(11000), (10101), (01110)};
б) {(111100), (110011), (001111)};
в) {(00001), (11111), (10100), (01010)};
г) {(101010), (010110), (000001)}.

6.4. Построить двоичный блоковый код из восьми кодовых слов длины 7 с минимальным кодовым расстоянием, не меньшим 4.

6.5. Построить двоичный блоковый код из M кодовых слов длины n с минимальным кодовым расстоянием d или доказать, что такого кода не существует, если параметры (n, M, d) равны: (3, 4, 2); (6, 4, 4); (12, 4, 8); (4, 7, 2); (8, 7, 4); (8, 8, 4); (8, 32, 3); (20, 1000, 7).

6.6. Пусть $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{x}$ — двоичные векторы длины n, любые два из которых находятся на расстоянии d, при этом d чётно. Доказать, что имеется точно один вектор, который находится на расстоянии $\frac{d}{2}$ от каждого из векторов $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$. Доказать, что имеется не более одного вектора, который находится на расстоянии $\frac{d}{2}$ от каждого из векторов $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{x}$.

6.7. Доказать, что при нечётном n двоичный $(n, 1)$-код с повторением является совершенным.

6.8. Выяснить, существует ли совершенный двоичный код длины $n = 147$ с минимальным расстоянием $d = 3$.

6.9. Доказать, что двоичный совершенный код длины n с минимальным расстоянием $d = 7$ может существовать только при $n = 7$ и $n = 23$.

6.10. Доказать, что линейный (5, 3)-код над полем $GF(4)$ с порождающей матрицей $G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 3 \end{pmatrix}$ является совершенным.

6.11. Описать двоичный (n, k)-код с минимальным расстоянием $d = 2$ и наибольшим возможным k.

6.12. Доказать, что в двоичном линейном коде либо все кодовые слова имеют чётный вес, либо половина кодовых слов чётного веса и половина — нечётного.

6.13. По заданной порождающей матрице G двоичного линейного кода найти поверочную матрицу H и параметры n, k и d:

а) $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$; б) $\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$; в) $\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}$;

г) $\begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$; д) $\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$;

е) $\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}$;

ж) $\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}$;

з) $\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$;

и) то же для $q = 3$ и $G = \begin{pmatrix} 0 & 0 & 1 & 1 & 2 & 2 \\ 1 & 1 & 1 & 2 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 & 1 \end{pmatrix}$.

6.14. Линейный код $\mathscr{C}$ называют самодвойственным, если $\mathscr{C} = \mathscr{C}^{\perp}$. Пусть $G = (I_k P)$ — порождающая матрица (n, k)-кода $\mathscr{C}$. Доказать, что код $\mathscr{C}$ самодвойственный тогда и только тогда, когда $P \cdot P^T = P^T \cdot P = I_k$.

6.15. Доказать, что самодвойственный код имеет чётную длину n и размерность $k = \frac{n}{2}$. Построить самодвойственный код длины 8.

6.16. *Прямоугольный двоичный код с числом информационных символов* $k = (a-1) \cdot (b-1)$. Строка из k информационных символов записывается в матрицу размера $(a-1) \times (b-1)$. К этой матрице добавляется справа один столбец так, чтобы число единиц в каждой строке стало чётным. Затем к полученной матрице размера $(a-1) \times b$ добавляется снизу одна строка так, чтобы число единиц в каждом столбце стало чётным. Элементы полученной матрицы размера $a \times b$ выписываются в строку в произвольном порядке, образуя кодовое слово длины $n = a \cdot b$. Найти минимальное кодовое расстояние d описанного прямоугольного кода. Перечислить все прямоугольные коды с $k = 24$, найти для них параметры n и d, выписать порождающие и проверочные матрицы.

6.17. Для двоичного линейного кода с проверочной матрицей

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

найти лидеров смежных классов, содержащих векторы (111101000); (110101011); (100010001); (010010010).

6.18. Для двоичного линейного кода с порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

найти проверочную матрицу, кодовые слова и лидеров смежных классов.

6.19. Для двоичного линейного кода с порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

найти проверочную матрицу, а также синдромы для векторов (01000); (00101); (10010); (11111).

6.20. Для двоичного линейного кода с порождающей матрицей

$$G=\begin{pmatrix}1&0&1&0&1&1\\0&1&1&1&1&0\\0&0&0&1&1&1\end{pmatrix}$$

выписать проверочную матрицу и лидеров смежных классов. Найти число a_i кодовых слов веса $i=0,\dots,6$. Декодировать в ближайшее кодовое слово векторы (1 1 1 0 1 0); (0 0 0 0 1 1); (1 0 1 0 1 0).

6.21. Для двоичного линейного кода с проверочной матрицей

$$H=\begin{pmatrix}0&1&1&1&0&0\\1&0&1&0&1&0\\1&1&0&0&0&1\end{pmatrix}$$

построить таблицу лидеров смежных классов и синдромов. В одном из смежных классов три лидера — перечислить их. Декодировать в ближайшее кодовое слово векторы (110110), (110111), (110001). Для одного из этих векторов декодирование неоднозначно — перечислить все варианты.

6.22. Для троичного линейного кода с порождающей матрицей

$$G=\begin{pmatrix}1&1&1&0\\2&0&1&1\end{pmatrix}$$

декодировать в ближайшее кодовое слово следующие векторы: (2121); (1201); (2222).

6.23. Построить таблицу Слепяна для двоичного линейного кода с проверочной матрицей $H=\begin{pmatrix}0&1&1&1&0&0\\1&0&1&0&1&0\\1&1&0&0&0&1\end{pmatrix}$.

6.24. При помощи кода Хэмминга $\mathcal{H}_3$ с проверочной матрицей

$$H=\begin{pmatrix}1&0&0&1&1&0&1\\0&1&0&1&0&1&1\\0&0&1&0&1&1&1\end{pmatrix}$$

закодировать собщения (0110) и (1010). Кроме того, предполагая, что случилось не более одной ошибки, декодировать

в ближайшее кодовое слово векторы (1100000), (1001010), (1101011), (0110110), (0100111).

6.25. Предположим, что для кода Хэмминга $\mathscr{H}_4$ столбцы проверочной матрицы заданы как 4-разрядные двоичные представления чисел 1, 2, ..., 15 (в таком порядке). Декодировать в ближайшее кодовое слово векторы (001000001100100), (101001110101100), (000100100011000).

6.26. Пусть (11010011) и (11001111) — искажённые слова расширенного кода Хэмминга длины 8. Какое из этих слов содержит одиночную, а какое — двойную ошибку?

6.27. Для (7, 4, 3)-кода Хэмминга обобщённый декодер $\mathscr{D}_{\hat{\rho}}$ правильно исправляет любые комбинации из t_0 стираний и t_1 искажений символов, если $t_0 + 2t_1 \leqslant 2$ (утверждение 6.5, с. 169). Декодировать, если возможно, векторы (11100E0), (0E11101), (01E10E1).

6.28. Для расширенного кода Хэмминга с проверочной матрицей H_3' (с. 207) исправить стирание в векторе (101E0111).

6.29. Пусть кодовые слова (7, 4, 3)-кода Хэмминга передаются по двоичному симметричному каналу с параметром $p = 0,01$ и на выходе канала используется декодер $\mathscr{D}_{\rho}$. Вычислить вероятность необнаружения ошибки и вероятность ошибочного декодирования.

6.30. Доказать, что (4, 2)-код над полем $GF(3)$ с порождающей матрицей $G = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 2 & 2 \end{pmatrix}$ является циклическим. Найти порождающий многочлен $g(x)$ и записать все кодовые слова в виде многочленов, кратных $g(x)$.

6.31. Выяснить, является ли циклическим (5, 3)-код над полем $GF(3)$ с порождающей матрицей $G = \begin{pmatrix} 2 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 2 & 1 & 1 \end{pmatrix}$.

6.32. Сколько существует различных двоичных линейных кодов длины 4? Сколько среди них циклических?

6.33. Сколько существует различных двоичных блоковых кодов длины 7, состоящих из 16 кодовых слов? Сколько среди них линейных, в том числе с минимальным кодовым расстоянием 3?

Сколько среди них циклических, в том числе с минимальным кодовым расстоянием 3?

6.34. Построить все двоичные циклические коды длины 7.

6.35. Проверить, что над полем $GF(3)$ справедливо равенство

$$x^8 - 1 = (x+1)(x+2)(x^2+1)(x^2+x+2)(x^2+2x+1).$$

Используя этот факт, найти число число троичных циклических кодов длины 8 и размерности $k = 0, 1, \ldots, 8$.

6.36. Выписать все кодовые слова двоичного циклического кода размерности $k = 4$ с порождающим многочленом $g(x) = 1 + x + x^3$.

6.37. Для двоичного циклического кода длины $n = 15$ с порождающим многочленом $g(x) = 1 + x + x^4$ найти проверочный многочлен, порождающую и проверочную матрицы и минимальное кодовое расстояние. Те же вопросы для $g(x) = 1 + x^3 + x^4 + x^5 + x^6$; $g(x) = 1 + x^4 + x^6 + x^7 + x^8$.

6.38. Для циклического кода над полем $GF(3)$ с $n = 8$, $k = 4$ и порождающим многочленом $g(x) = (x^2+1)(x^2+x+1)$ найти проверочный многочлен, порождающую и проверочную матрицы. Те же вопросы, если $g(x) = (x^2-1)(x^2+x+1)$.

6.39. Для двоичного циклического кода длины n с порождающим многочленом $g(x)$ доказать утверждения:

— если $g(x)$ делится на $x+1$, то все кодовые слова имеют чётный вес;

— если n нечётно и $g(x)$ не делится на $x+1$, то вектор $(1\ldots1)$ принадлежит коду.

6.40. Блоковый код называют реверсивным, если для любого кодового слова $(c_0, c_1, \ldots, c_{n-1})$ вектор $(c_{n-1}, \ldots, c_1, c_0)$, выписанный в обратном порядке, также является кодовым словом. Доказать, что циклический код с порождающим многочленом $g(x)$ является реверсивным тогда и только тогда, когда для любого корня α многочлена $g(x)$ элемент α^{-1} также является корнем этого многочлена.

6.41. Для циклического кода доказать, что если вектор ошибок e обнаруживается, то и любой вектор ошибок, являющийся циклическим сдвигом вектора e, также обнаруживается.

6.42. Проверить, что многочлен $x^4 + x^3 + x^2 + x + 1$ неприводим над полем $GF(2)$. Используя этот факт, построить проверочную

матрицу для БЧХ-кода, исправляющего две ошибки, длины $n = 15$ и размерности $k = 7$.

Послесловие

Завершая наше краткое введение в теорию информации и кодирования, полезно хотя бы в общих чертах представить её современное состояние и отметить связи с различными разделами математики и той широкой прикладной областью, которую называют информационными технологиями.

Начиная с 1938 года в работах К. Шеннона были созданы основы новых областей исследований в общей теории связи, теории автоматов, электротехники, теории информации и лингвистики и в значительной степени предопределены направления будущего развития кибернетики и математической теории связи. К настоящему времени теория информации и кодирования превратилась в широкую и активно развивающуюся область науки. Значительный вклад в эту теорию внесли такие выдающиеся отечественные учёные, как А. Н. Колмогоров, В. А. Котельников, А. А. Харкевич, А. Я. Хинчин и другие.

Сегодня в этой области работают крупные научные коллективы, издаётся большое количество книг и журналов. Даже неполный список основных монографий и учебников по теории информации и кодирования на русском и английском языках насчитывает несколько сотен книг. Специалистам хорошо известны такие научные журналы, как «Дискретная математика», «Проблемы передачи информации», «Advances in mathematics of communications», «Communications in information and systems», «Designs, codes, and cryptography», «Entropy», «Foundations and trends in communications and information theory», «IEEE Transactions on information theory», «Information and computation», «International journal of information and coding theory».

К настоящему времени в теории информации и кодирования выделились три крупных относительно самостоятельных направления: во-первых, это собственно *теория информации*, во-вторых, *теория кодов, исправляющих ошибки*, и, в-третьих, *теория сжатия данных*.

Собственно *теория информации* связана с количественным измерением информации, порождаемой различными источниками и пе-

редаваемой по каналам связи. Теория информации использует в основном математический аппарат теории вероятностей.

По сравнению с относительно простыми моделями источников сообщений, рассмотренными в учебном пособии, в теории информации изучаются существенно более сложные модели дискретных и непрерывных случайных процессов, информационные характеристики таких процессов и условия их информационной устойчивости.

Относительно моделей каналов связи также можно отметить, что от простых моделей стационарных дискретных каналов без памяти теория переходит к сложным более общим дискретным и непрерывным моделям (каналы связи с обратной связью, с памятью, нестационарные, гауссовские и т. д.). Для таких каналов связи вычисление пропускной способности часто представляет сложную задачу, и для её решения разрабатываются специальные методы и численные итерационные алгоритмы. В теоремах кодирования для различных моделей каналов связи устанавливаются фундаментальные пределы скорости передачи с произвольно малой вероятностью ошибки.

Современная *теория кодов, исправляющих ошибки*, является областью, в которой применяются в основном методы алгебры, алгебраической геометрии, комбинаторики и теории графов. Конечные поля и кольца, матрицы и многочлены, комбинаторные схемы, конечные геометрии являются тем языком, на котором описываются коды, предназначенные для обнаружения и исправления ошибок в различных каналах связи, и соответствующие алгоритмы декодирования.

Такие конструкции, как коды БЧХ, Рида—Соломона, Рида—Маллера, свёрточные коды, турбокоды, коды с малой плотностью проверок на чётность представляют яркие примеры успешного практического применения сложной математической теории и позволяют в некоторых случаях осуществлять передачу информации со скоростью, близкой к пропускной способности канала связи.

В повседневную жизнь вошли такие применения теории кодирования, как CRC-алгоритмы контроля ошибок в компьютерных файлах, помехоустойчивое кодирование при записи звука и изображений на CD и DVD и многие другие. Выдающиеся достижения в космических исследованиях были бы невозможны без использования кодов, исправляющих ошибки, при передаче слабых сигналов на огромные

расстояния между космическими аппаратами и Землёй.

Теория сжатия данных связана в основном с разработкой алгоритмов кодирования информации, порождаемой различными источниками. Характеристики таких алгоритмов сопоставляются с предельными значениями, которые устанавливаются методами вероятностной теории информации. От простых моделей, рассмотренных в нашем пособии (алфавитное кодирование для дискретных источников без памяти и марковских источников), современная теория переходит к таким сложным задачам, как кодирование нескольких зависимых источников, кодирование непрерывных источников при заданном уровне искажений, вычисление предела скорости передачи как функции от уровня искажений непрерывного источника, кодирование источников с неизвестной вероятностной структурой (универсальное кодирование).

Теория сжатия данных имеет очень широкий круг практических приложений, от факсимильных сообщений до телевидения высокой чёткости. Так, известный справочник Д. Саломона [47], в котором приводятся краткие описания основных алгоритмов сжатия данных и их применений, насчитывает более тысячи страниц.

В теории сжатия данных разрабатываются два класса алгоритмов сжатия — без потери информации и с потерей.

Сжатие без потери информации означает, что закодированное (сжатое) сообщение может быть в точности восстановлено при помощи некоторого алгоритма. Сжатие без потери информации применяется для таких источников, где по существу необходимо точное восстановление сообщения после сжатия — текстовые сообщения, базы данных, программное обеспечение и т. п. Здесь применяются, например, адаптивное кодирование по Хаффману (с перестройкой кодового дерева в процессе кодирования в зависимости от кодируемого текста), словарные методы (семейство алгоритмов Лемпела—Зива), арифметическое кодирование, сортировка блоков, кодирование длин серий и другие методы.

Сжатие с потерей информации означает, что закодированное (сжатое) сообщение может быть восстановлено лишь с некоторой точностью. Сжатие с потерей информации применяется для таких источников, сообщения которых воспринимаются органами чувств человека — звуковые сообщения, неподвижные изображения, ви-

део. Поскольку органы слуха и зрения имеют определённые пороги чувствительности и не могут воспринимать, например, звуки вне определённого диапазона частот, свет вне определённого спектра, слишком быстро движущиеся объекты и т. п., при кодировании таких непрерывных сигналов их можно огрублять так, что затем, после декодирования, человек не сможет заметить внесённых при кодировании искажений. За счёт такого огрубления (потери информации) достигают очень высокой степени сжатия данных по сравнению с алгоритмами сжатия без потери информации.

В алгоритмах сжатия с потерей информации используются специфические для этой области математические методы, например, дискретное двумерное косинус-преобразование, вейвлет-преобразования, фрактальные методы. Для достижения высокой степени сжатия в современных алгоритмах используются также довольно тонкие особенности физиологии органов слуха и зрения человека, например, невосприимчивость слуха к определённым частотам после громкого звука на этих частотах, инерционность восприятия движущихся изображений.

Примерами форматов данных и стандартов для неподвижных изображений и видео являются PNG, GIF, BMP, JPEG, JPEG2000, H.264, семейство алгоритмов MPEG. Особые форматы данных и стандарты используются для сжатия звуковой информации.

Кроме названных трёх основных направлений, в теории информации и кодирования развиваются и другие области исследований. Например, на стыке теории информации и теории сложности вычислений развивается алгоритмическая теория информации. Понятие сложности объекта (текста) по А.Н. Колмогорову определяется как минимальный размер программы, порождающей (вычисляющей) данный объект. Колмогоровская сложность тесно связана с энтропией и является, в определённом смысле, более фундаментальным понятием.

Методы и результаты теории информации и кодирования находят широкое применение в различных разделах математики и других областях науки.

Прежде всего отметим приложения теории информации в теории вероятностей и математической статистике. Информационные характеристики (энтропия, взаимная информация) являются функциона-

лами от вероятностных распределений. Эти характеристики, в свою очередь, описывают поведение длинных последовательностей случайных величин и позволяют, например, оценивать вероятности редких событий. Теоретико-информационные методы применяются при изучении вероятностей больших уклонений, в асимптотических оценках вероятностей ошибок статистических критериев, в спектральной теории случайных процессов, предсказании дискретных временных рядов, распознавании образов.

В лингвистике одной из важных прикладных задач является статистическое оценивание энтропии естественных языков.

В физике теория информации находит применение, например, в термодинамике, статистической механике, квантовой теории, теории хаоса. Следует подчеркнуть, что понятие энтропии впервые появилось именно в термодинамике как мера неравномерности распределения частиц по состояниям; фундаментальный второй закон термодинамики устанавливает неубывание энтропии в замкнутой системе. В последние годы активно разрабатывается так называемая квантовая теория информации, изучаются возможности квантовых вычислений и перспективы создания квантовых компьютеров.

В биологии областями приложения теории информации и кодирования являются, например, молекулярная биология и генетика. На стыке возникло такое направление, как «вычислительная биология».

В психологии и физиологии установлены глубокие связи между абстрактными понятиями теории информации и кодирования и сложными конкретными процессами высшей нервной деятельности, особенно процессами функционирования человеческого мозга.

В экономической науке на теоретико-информационном языке описываются методы оптимального управления инвестициями и взаимосвязи между скоростью роста капитала и энтропией рынка.

В криптографии методы теории информации и кодирования используются для определения меры стойкости шифров, при описании схем разделения секрета и в других задачах.

Теория информации и кодирования, возникнув на стыке между чистой математикой и прикладными аспектами информационных технологий, испытывает сильное стимулирующее влияние с обеих сторон и, в свою очередь, оказывает на них значительное воздействие.

Вероятно, перспективы развития теории информации и кодирования связаны и с получением новых глубоких результатов в математике, и с новыми достижениями в информационных технологиях, качественно изменяющими жизнь человеческого общества.

Автор надеется, что предлагаемое учебное пособие будет полезно студентам-математикам, специализирующимся в области информационной безопасности.

Список литературы

1. *Акритас А.* Основы компьютерной алгебры с приложениями. – М.: Мир, 1994. – 544 с.
2. *Аршинов М. Н., Садовский Л. Е.* Коды и математика (рассказы о кодировании). – М.: Наука, 1983. – 144 с.
3. *Берлекэмп Э. Р.* Алгебраическая теория кодирования. – М.: Мир, 1971. – 478 с.
4. *Блейхут Р. Э.* Теория и практика кодов, контролирующих ошибки. – М.: Мир, 1986. – 576 с.
5. *Боровков А. А.* Теория вероятностей. Учебное пособие. – 5-е изд. Сущ. перераб. и доп. – М.: Книжный дом «Либроком», 2009. – 656 с.
6. *Вернер М.* Основы кодирования. – М.: Техносфера, 2004. – 288 с.
7. *Вольфовиц Дж.* Теоремы кодирования теории информации. – М.: Мир, 1967. – 248 с.
8. *Гаврилов Г. П., Сапоженко А. А.* Задачи и упражнения по дискретной математике. Учебное пособие. – 3-е изд. – М.: ФИЗМАТЛИТ, 2004. – 416 с.
9. *Галлагер Р.* Теория информации и надежная связь. – М.: Сов. Радио, 1974. – 719 с.
10. *Глухов М. М.* Алгебра и аналитическая геометрия. – М.: Гелиос АРВ, 2005. – 392 с.
11. *Глухов М. М., Елизаров В. П., Нечаев А. А.* Алгебра: Учебник. – М.: Гелиос АРВ, 2003. – В 2-х томах. – том 1, 336 с.; том 2, 416 с.
12. *Дмитриев В. И.* Прикладная теория информации. – М.: Высшая школа, 1989. – 320 с.
13. *Духин А. А.* Теория информации. – М.: Гелиос АРВ, 2007. – 248 с.
14. *Земор Ж.* Курс криптографии. – М.-Ижевск: НИЦ «Регулярная и хаотическая динамика», Институт компьютерных исследований, 2006. – 256 с.
15. *Зубов А. Ю.* Задачник-практикум по алгебре и аналитической геометрии. – М.: 1988. – 288 с.
16. *Колесник В. Д., Полтырев Г. Ш.* Курс теории информации. – М.: Наука, 1982. – 416 с.

17. *Котоусов А. С.* Теория информации. Учебное пособие для вузов. – М.: Радио и связь, 2003. – 80 с.
18. *Кричевский Р. Е.* Сжатие и поиск информации. – М.: Радио и связь, 1989. – 168 с.
19. *Лидл Р., Пильц Г.* Прикладная абстрактная алгебра. – Екатеринбург: изд-во Урал. ун-та, 1996. – 744 с.
20. *Лидовский В. В.* Теория информации. Учебное пособие. – М.: Компания Спутник+, 2004. – 111 с.
21. *Мак-Вильямс Ф. Дж., Слоэн Н. Дж. А.* Теория кодов, исправляющих ошибки. – М.: Связь, 1989. – 744 с.
22. *Марков А. А.* Введение в теорию кодирования. – М.: Наука, 1982. – 192 с.
23. *Орлов В. А., Филиппов Л. И.* Теория информации в упражнениях и задачах. – М.: Высшая школа, 1976. – 136 с.
24. *Питерсон У., Уэлдон Э.* Коды, исправляющие ошибки. – М.: Мир, 1976. – 594 с.
25. *Свешников А. А.* (ред.) Сборник задач по теории вероятностей, математической статистике и теории случайных функций. – М.: Наука, 1970. – 656 с.
26. *Реньи А.* Трилогия о математике. – М.: Мир, 1980. – 376 с.
27. *Сачков В. Н.* Введение в комбинаторные методы дискретной математики. – М.: МЦНМО, 2004. – 424 с.
28. *Соловьева Ф. И.* Введение в теорию кодирования. Учебное пособие. – Новосибирск: Новосиб. гос. ун-т, 2006. – 127 с.
29. *Файнстейн А.* Основы теории информации. – М.: ИЛ, 1960. – 140 с.
30. *Фано Р.* Передача информации. Статистическая теория связи. – М.: Мир, 1965. – 438 с.
31. *Харин Ю. С., Берник В. И., Матвеев Г. В., Агиевич С. В.* Математические и компьютерные основы криптологии. – Минск: Новое знание, 2003. – 382 с.
32. *Хинчин А. Я.* Избранные труды по теории вероятностей. – М.: ТВП, 1995. – 552 с.
33. *Хэмминг Р. В.* Теория кодирования и теория информации. – М.: Радио и связь, 1983. – 176 с.
34. *Чисар И., Кёрнер Я.* Теория информации. Теоремы кодирования для дискретных систем без памяти. – М.: Мир, 1985. – 400 с.

35. *Шеннон К.* Работы по теории информации и кибернетике. – М.: ИЛ, 1963. – 829 с.
36. *Ширяев А. Н.* Вероятность. – 4-е изд. – М.: МЦНМО, 2007. – В 2-х кн. – к. 1, 552 с.; к. 2, 416 с.
37. *Яблонский С. В.* Введение в дискретную математику. – М.: Наука, 1979. – 272 с.
38. *Яглом А. М., Яглом И. М.* Вероятность и информация. – М.: Наука, 1973. – 512 с.
39. *Ash R. B.* Information theory. – 2nd ed. – Dover Publ., 1990. – 352 pp.
40. *Cover T. M., Thomas J. A.* Elements of information theory. – 2nd ed. – Wiley Interscience, 2006. – 748 pp.
41. *Hankerson D., Harris G. A., Johnson P. D.* Introduction to information theory and data compression. – 2nd ed. – Chapman & Hall / CRC, 2003. – 384 pp.
42. *Huffman W. C., Pless V.* Fundamentals of error-correcting codes. – Cambridge U.P., 2003. – 646 pp.
43. *Lin S., Costello D. J.* Error control coding: fundamentals and applications. – 2nd ed. – Prentice Hall, 2004. – 1272 pp.
44. *Van der Lubbe J. C. A.* Information theory. – Cambridge U.P., 2002. – 350 pp.
45. *MacCay D. J. C.* Information theory, inference, and learning algorithms. – Cambridge U. P., 2003. – 628 pp.
46. *McEliece R. J.* The theory of information and coding: a mathematical framework for communication. – 2nd ed. – Cambridge U.P., 2002. – 397 pp.
47. *Salomon D.* Data compression. The complete reference. – 4th ed. – Springer, 2007. – 1093 pp.
48. *Yeung R. W.* A first course in information theory. – Kluwer Academic / Plenum Publishers, 2002. – 412 pp.

Учебное пособие

Чечёта Сергей Иванович

Введение в дискретную теорию информации и кодирования

Подписано в печать 20.11.2010 г. Формат $60 \times 90\ 1/16$. Бумага офсетная.
Печать офсетная. Печ. л. 17,5. Тираж 1000 экз. Заказ № .

Издательство Московского центра непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11, тел. (499) 241–74–83.

Отпечатано с готовых диапозитивов в ППП «Типография „Наука“».
121099, Москва, Шубинский пер., 6.
