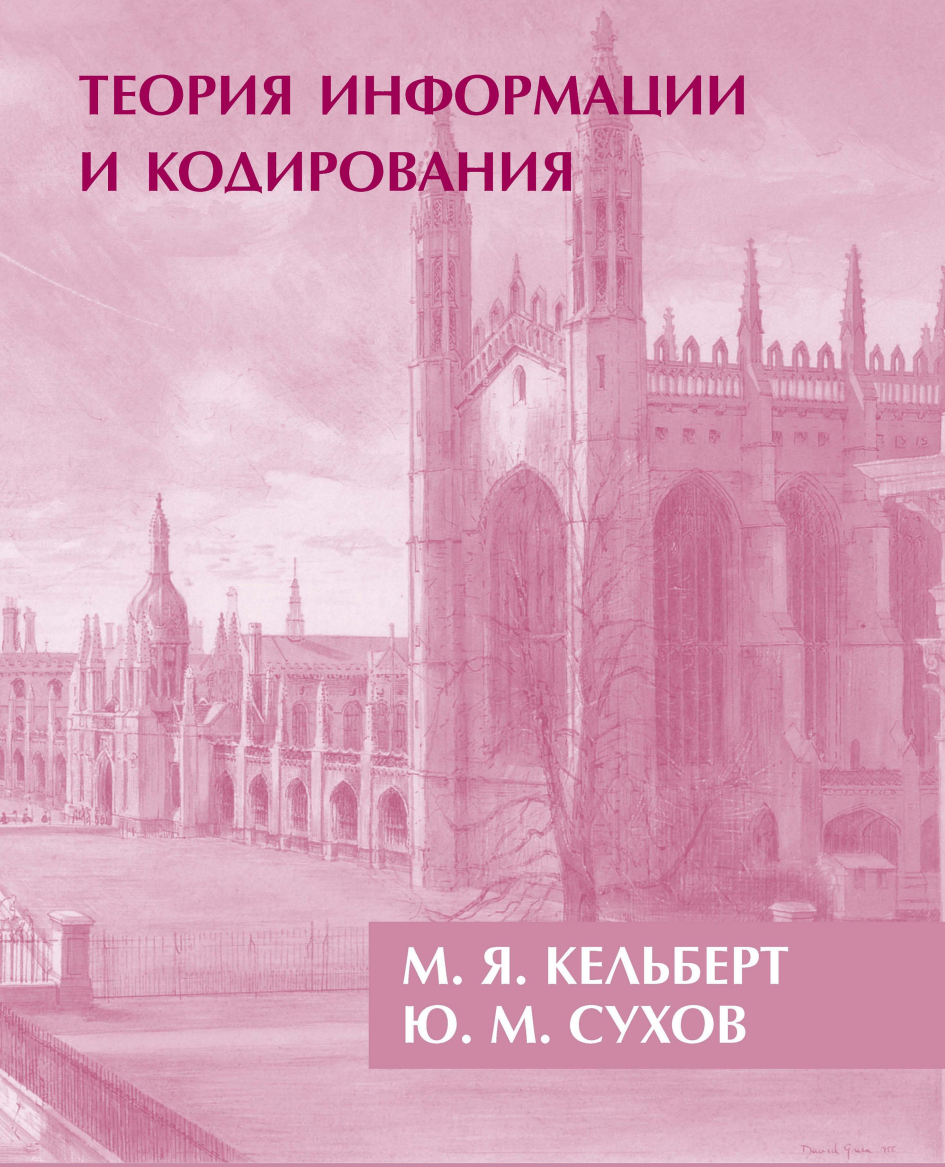
ВЕРОЯТНОСТЬ
и СТАТИСТИКА
в примерах
и задачах

ТЕОРИЯ ИНФОРМАЦИИ И КОДИРОВАНИЯ



М. Я. КЕЛЬБЕРТ
Ю. М. СУХОВ

М. Я. Кельберт, Ю. М. Сухов

Вероятность и статистика в примерах и задачах

Том 3

Теория информации и кодирования

Электронное издание

Москва
Издательство МЦНМО
2016

УДК 519.21
ББК 22.171
К34

Кельберт М. Я., Сухов Ю. М.
Вероятность и статистика в примерах и задачах
Т. 3: Теория информации и кодирования
Электронное издание
М.: МЦНМО, 2016
567 с.
ISBN 978-5-4439-2377-2

Для освоения таких разделов прикладной математики, как теория вероятностей, математическая статистика, теория информации и кодирование, тренировка в решении задач и выработка интуиции важны не меньше, чем изучение доказательств теорем; большое разнообразие задач по этому предмету затрудняет студентам переход от лекций к экзаменационным задачам, а от них — к практике.

Этот том включает стандартный пакет информационно-теоретического материала, обычно читаемого на факультетах информатики и электроники, а также прикладной математики ведущих университетов. При этом излагаются как вероятностные, так и алгебраические аспекты теории информации и кодирования, включая как основы теории, так и некоторые ее современные аспекты. Предмет этой книги критически важен для современных приложений (телекоммуникации, обработка сигналов, информатика, криптография).

Авторы собрали большое количество упражнений, снабженных полными решениями. Эти решения адаптированы к нуждам и умениям учащихся. Необходимые теоретические сведения приводятся по ходу изложения; кроме того, текст снабжен историческими отступлениями.

Подготовлено на основе книги:
Кельберт М. Я., Сухов Ю. М. Вероятность и статистика в примерах и задачах. Т. 3: Теория информации и кодирования. — М.: МЦНМО, 2014. — ISBN 978-5-4439-0154-1.

Издательство Московского центра
непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11,
тел. (499)-241-08-04
http://www.mccme.ru

ISBN 978-5-4439-2377-2

Оглавление

Предисловие

Эта книга частично основывается на нескольких математических курсах Кембриджа: теория информации для 3-го года обучения (который читался, постоянно развиваясь, на протяжении последних четырех десятилетий под разными названиями), кодирование и криптография (более молодой и упрощённый курс, исключающий сложные технические вопросы) и более сложные курсы из части III (представляющей собой кембриджский эквивалент математической магистратуры). Содержание книги построено, по существу, вокруг следующих понятий:

а) энтропия распределения вероятностей как мера «недостоверности» и энтропия на случайный символ как мера «изменчивости» типичных траекторий случайного процесса,

б) кодирование как средство для измерения и использования избыточности информации, генерируемой процессом.

Таким образом, содержание данной книги включает более или менее стандартный пакет информационно-теоретического материала, который можно найти в наше время в учебных курсах по всему миру, в основном читаемых на факультетах информатики и электроники и иногда теории вероятностей и математической статистики. Что отличает эту книгу от остальных, так это, прежде всего, широкий спектр примеров (отличительная черта всей нашей серии учебников «*Вероятность и статистика, в примерах*», опубликованной в издательстве Кембриджского университета). Большинство из этих примеров соответствуют уровню, принятому на экзаменах математических курсов в Кембридже. Таким образом, наши читатели могут сами понять, какого уровня они достигли или собираются достичь.

Второе отличие этой книги от большинства других книг по теории информации или теории кодирования заключается в том, что она охватывает оба возможных направления: вероятностное и алгебраическое. Как правило, эти направления исследований представлены в *разных* монографиях, учебниках и курсах, зачастую написанных людьми, работающими на разных факультетах. Подготовке этой книги способствовало то, что её авторы имели давние связи с Институтом проблем передачи информации Российской академии наук (ИППИ), в котором традиционно изучается широкий спектр проблем. Достаточно упомянуть, среди прочих, такие имена, как Роланд Добрушин, Рафаил Хасьминский, Марк Пинскер, Владимир Блиновский, Вячеслав Прелов, Борис Цыбаков, Камиль Зигангиров (теория вероятностей и математическая статистика), Валентин Афанасьев,

Сергей Гельфанд, Валерий Гоппа, Инна Грушко, Григорий Кабатянский, Григорий Маргулис, Юрий Сагалович, Алексей Скоробогатов, Михаил Цфасман, Леонид Бассалыго, Виктор Зиновьев, Виктор Зяблов (алгебра, комбинаторика, геометрия и теория чисел), которые работали или продолжают работать в ИППИ (было время, когда все они размещались в пяти комнатах в центре Москвы). Традиции преподавания математических тем в теории информации и кодирования в Кембридже восходят первоначально к Питеру Уиттлу (теория вероятностей и оптимизация) и позднее к Чарльзу Голди (теория вероятностей), Ричарду Пинчу (алгебра и геометрия), Тому Кёрнеру и Кейту Карну (анализ) и Тому Фишеру (теория чисел).

Мы также хотели бы добавить, что книга написана авторами, имеющими математическое образование (и остающимися до сих пор математиками), которые, тем не менее, имеют сильную тягу к приложениям, невзирая на все сопутствующие проблемы, возникающие в процессе прикладной работы: неопределённость, неточность, дискуссионность (включая, разумеется, личностный фактор) и последнее, но отнюдь не менее важное — необходимость эффективно применять на практике математические идеи. Авторы твердо считают, что математизация является основным путём к выживанию и совершенствованию в современном конкурентном мире и, следовательно, математику необходимо воспринимать всерьёз и изучать добросовестно.

Обе вышеупомянутые концепции (энтропия и коды), формирующие основу информационно-теоретического подхода к случайным процессам, были введены Шенноном в 1940-х гг., в довольно завершённой форме в публикациях [S, SW]. Конечно, понятие энтропии уже существовало в термодинамике, и его очень хорошо осознавали Больцман и Гиббс на стыке XIX и XX столетий. Коды также эффективно применялись на практике со времен античного мира. Но именно Шеннон полностью оценил роль этих понятий и положил их в основу современного информационно-теоретического подхода к случайным процессам. Не будучи профессиональным математиком, он не всегда давал полные доказательства своих конструкций. (Может быть, он и не задумывался о них.) В соответствующих разделах мы прокомментируем некоторые довольно деликатные моменты в отношениях Шеннона с математическим сообществом. К счастью, похоже, это его сильно не тревожило. (В отличие от Больцмана, который был особенно чувствителен к внешним отзывам и принимал их, пожалуй, слишком близко к сердцу.) Шеннон, несомненно, понимал всю ценность своих открытий, и, по нашему мнению, они ставят его в один ряд с такими выдающимися математиками, как Винер и фон Нейман.

Будет справедливо отметить, что имя Шеннона по-прежнему доминирует как в вероятностном, так и в алгебраическом направлениях в современной теории информации и кодирования. Это довольно необычно, учитывая,

что мы говорим о вкладе человека, который работал в этой области более чем 40 лет назад. (Хотя в отношении нескольких сложных вопросов Шеннон, вероятно, мог бы повторить слова Эйнштейна, переформулировав их так: «С тех пор как математики вторглись в теорию связи, я перестал что-либо понимать в ней».)

За годы, что прошли после работ Шеннона, в математике и электротехнике произошли большие изменения, не говоря уже о компьютерных науках. Кто бы мог предвидеть в 1940–1950-х гг., что соперничество между подходами Шеннона в теории информации и Винера в кибернетике получит такое завершение? Действительно, кибернетика обещала огромные (даже фантастические) выгоды для всего человечества, в то время как теория информации только утверждала, что в определенных пределах можно достичь скромной цели исправления ошибок при передаче.

Книга Винера [W] в 1950–1960-х гг. пленила умы мыслителей практически во всех областях интеллектуальной деятельности. В частности, кибернетика стала серьезной политической проблемой в Советском Союзе и его странах-сателлитах: сначала она была объявлена «буржуазной антинаучной теорией», а затем ей придали неоправданно большое значение. (Цитата из критического обзора кибернетики в ведущем в советской идеологии журнале «*Вопросы философии*» 1953 г. гласит: «Империалистам не удаётся разрешить противоречия умирающего капиталистического общества. Они не могут предотвратить неизбежный экономический кризис. И поэтому они пытаются найти решение не только в бешеной гонке вооружений, но и в идеологической войне. В глубоком отчаянии они прибегают к помощи псевдонауки, которая даёт им некоторые проблески надежды продлить их существование». Советский «*Краткий словарь по философии*» (1954 г.), имевший тираж в сотни тысяч экземпляров, определял кибернетику как «реакционную псевдонауку, которая появилась в США после первой мировой войны и позднее распространилась в других капиталистических странах: вид современного механицизма». Однако под давлением ведущих советских физиков, завоевавших авторитет после успехов советской ядерной программы, тот же самый журнал «*Вопросы философии*» в 1955 г. опубликовал позитивный отзыв о кибернетике. Среди авторов этой статьи были Алексей Ляпунов и Сергей Соболев, выдающиеся советские математики.)

Любопытно, что в недавно опубликованной биографии Винера [CS] указывается, что существуют «тайные правительственные документы (США), которые показывают, как ФБР и ЦРУ следили за Винером в разгар холодной войны, чтобы помешать его социальной активности и растущему влиянию кибернетики в стране и за рубежом». Интересные сравнения можно найти в работе [Hei].

Однако история пошла своим путём. Фримен Дайсон написал в своём обзоре [Dy]: «(Теория Шеннона) была математически элегантной, понятной и легко применимой на практике к проблемам связи. Она была намного более удобной для пользователя, чем кибернетика. Теория стала основой новой дисциплины под названием „теория информации...“ (В настоящее время) в базовый курс подготовки инженеров по электронике входит теория информации, основанная на теории Шеннона, а кибернетика оказалась забытой».

Однако это не совсем так: только на территории бывшего Советского Союза до сих пор работают институты и отделы, в название которых входит слово «кибернетика»: два в Москве, два в Минске, и по одному в Таллине, Тбилиси, Ташкенте и Киеве (последний являлся известнейшим центром компьютерной науки в целом в бывшем СССР). И в Великобритании существуют по крайней мере четыре факультета, в университетах Болтона, Брэдфорда, Халла и Рединга, не считая различных ассоциаций и обществ. Во всём мире общества, связанные с кибернетикой, кажется, процветают, что видно из перечисления названий: от Института метода (Швейцария) или Академии кибернетики (Италия) до аргентинской Ассоциации общей теории систем и кибернетики, Буэнос-Айрес. И мы были рады узнать о существовании Кембриджского кибернетического общества (Бельмонт, Калифорния, США). Напротив, теория информации фигурирует в названиях лишь нескольких организаций. Видимо, давний спор между Шенноном и Винером еще не вполне закончен.

В любом случае репутация Винера в области математики остаётся несокрушимой. Достаточно назвать несколько жемчужин его творчества, таких как теорема Пэли—Винера (которая была доказана во время многочисленных посещений Винером Кембриджа), метод Винера—Хопфа и, конечно, особенно близкий нашему сердцу винеровский процесс, чтобы понять его истинную роль в научных исследованиях и приложениях.

Существующие воспоминания об этом гиганте науки изображают Винера сложной и противоречивой личностью. (Название биографии [CS] в этом смысле весьма показательно, хотя такие взгляды оспариваются; см., например, обзор [Mar]. В этой книге мы пытаемся принять более мягкий тон, как, например, в главе о Винере в книге [Ja], с. 386–391.) С другой стороны, имеющиеся документальные записи о жизни Шеннона (так же как и других отцов теории информации и кодирования, в частности Ричарда Хэмминга) дают целостную картину спокойного, умного человека, не лишённого чувства юмора. Мы надеемся, что такое впечатление не будет мешать написанию биографии Шеннона и что в будущем мы увидим столь же много книг о Шенноне, сколько их написано о Винере.

Как было сказано ранее, цель этой книги двояка: обеспечить синтетическое введение в вероятностные и алгебраические аспекты теории, поддерживаемое значительным количеством задач и примеров, и обсудить ряд вопросов, редко представленных в большинстве книг. Главы 1–3 дают введение в основы теории информации и кодирования и обсуждают некоторые современные ответвления этих тем. Мы концентрируемся в этих главах на типичных задачах и примерах (многие из которых возникли в кембриджских курсах) больше, чем на подробном изложении теории, стоящей за ними. Глава 4 даёт краткое введение в более специализированные разделы теории информации. Здесь изложение более сжато и некоторые важные результаты приводятся без доказательств.

В связи с тем, что большая часть текста основана на конспектах лекций и решений различных задач для аудиторных занятий и экзаменов, в книге встречаются неизбежные повторы, многие обозначения и примеры даются на упрощённом языке. Часто мы делали это сознательно, чтобы передать живую атмосферу процесса преподавания и изучения.

Три прекрасные книги [GP], [M] и [CT] оказали особенно сильное влияние на наш учебник. Здесь сыграла свою роль наша долгая дружба с Чарльзом Голди, так же как и знакомство с Томом Ковером. Кроме того, на наш текст оказали влияние такие книги, как [Bl, MWS, R1] и [vL] (мы даже кое-что заимствовали из них). Мы благодарим за гостеприимство Институт Исаака Ньютона Университета Кембриджа (2002–2010 гг.), особенно программу стохастических процессов в коммуникационных науках (январь–июль 2010 г.). Различные части книги обсуждались со многими коллегами из разных учреждений, в первую очередь из Института проблем передачи информации и Института математической геофизики и прогноза землетрясений, Москва. В частности, мы признательны за интерес к книге, проявленный Г. Кабатянским и С. Пироговым, и за замечания, которые возникли в результате последовавших обсуждений. Мы также хотели бы поблагодарить Джеймса Лоуренса (статистическая лаборатория Университета Кембриджа) за его помощь с рисунками.

В ходе заключительного этапа работы над книгой значительную роль сыграла поддержка агентства FAPESP (штат Сан-Пауло, Бразилия), в рамках грантов 2010/17835-0, 2011.20133-0 и 2012.04372-7, а также Ректории Университета Сан-Пауло в рамках гранта 2011.5.764.45.0. Перевод книги на русский язык и техническая сторона её подготовки к печати были выполнены при поддержке РФФИ. Работа переводчика С. Кулешова и редактора О. Широковой заслуживает самой высокой оценки и глубокой благодарности.

Ссылки на том 1 и том 2 относятся к переводам наших книг «Вероятность и статистика в примерах», том 1 и 2 (Probability and Statistics

by Example, Cambridge University Press); страницы даются по русскому изданию. Мы используем стиль этих книг, подавая бóльшую часть материала как «примеры с решениями». Много материала дается в виде задач, взятых из экзаменационных работ (Cambridge Tripos Exam papers), которые сохраняют свой оригинальный стиль.

На протяжении всей книги мы старались развлечь читателя. Когда нашей собственной фантазии не хватало, мы привлекали идеи других авторов, в основном из различных интернет-источников. (К счастью, поток юмора кажется неиссякаемым, и иногда в интернете появляются блестящие высказывания.)

Символ □ указывает на конец отдельной части книги, чтобы отделить её от последующего текста: это относится к примерам, задачам (решениям), определениям, замечаниям и доказательствам.

Глава 1

Основные понятия теории информации

На протяжении всей книги символ P обозначает различные распределения вероятностей. В частности, в гл. 1 этот символ преимущественно обозначает распределение вероятностей последовательности случайных величин (с. в.), характеризующей источники информации. Как правило, это будут последовательности независимых одинаково распределенных случайных величин (н. о. р. с. в.) или цепи Маркова с дискретным временем (ц. м. д. в.); $\mathsf{P}(U_1 = u_1, \ldots, U_n = u_n)$ — *совместная вероятность* события, при котором с. в. $U_1, \ldots, U_n$ принимают значения $u_1, \ldots, u_n$, а $\mathsf{P}(V = v \,|\, U = u, W = w)$ — *условная вероятность*, т. е. вероятность того, что с. в. V принимает значение v, при условии, что с. в. U и W равны u и w соответственно. Символ E закреплён за *математическим ожиданием* с. в. с распределением P.

Символы p и P используются для обозначения различных вероятностей (и связанных с ними объектов, таких как переходная функция ц. м. д. в.). Символ $\#A$ обозначает мощность конечного множества A. Символом $\mathbf{1}$ обозначается *индикаторная* (характеристическая) функция множества. Для логарифмов будем использовать следующие обозначения и правила действия: $\log = \log_2$ и $\forall\, b > 1$: $0 \cdot \log_b 0 = 0 \cdot \log_b \infty = 0$. Далее, при $x > 0$ через $\lfloor x \rfloor$ и $\lceil x \rceil$ мы обозначим максимальное целое число, не превосходящее x, и минимальное целое число, не меньшее x, соответственно. Таким образом, $\lfloor x \rfloor \leqslant x \leqslant \lceil x \rceil$; неравенство превращается в равенство при целых x ($\lfloor x \rfloor$ называется *целой частью* числа x).

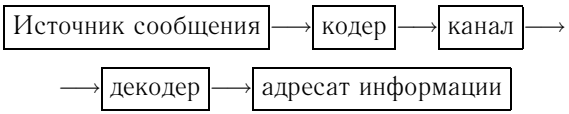
Аббревиатуры л. ч. и п. ч. обозначают соответственно левую и правую части уравнения или неравенства.

§ 1.1. Основные понятия. Неравенство Крафта. Кодирование Хаффмана

Жить эффективно — жить с адекватной информацией.

Норберт Винер (1894–1964),
американский математик

Типичная схема, применяемая при передаче информации, выглядит следующим образом:



Пример 1.1.1. а) Источник сообщения: хор Кембриджского колледжа.

б) Кодер: блок записи Би-Би-Си. Он переводит звук в двоичный массив и записывает его на дорожку компакт-диска. Затем компакт-диск размножают и передают в магазин.

в) Канал: оптовый покупатель компакт-дисков в Англии, переправляющий их в Австралию. На канал воздействует «шум»: возможный ущерб (от механических, электрических, химических и др. воздействий), полученный во время передачи (транспортировки).

г) Декодер: проигрыватель компакт-дисков в Австралии.

д) Адресат информации: аудитории в Австралии.

е) Цель: обеспечение высокого качества звука, несмотря на повреждения.

В самом деле, компакт-диск может повредиться от иглы, если в нём проделать аккуратную дырку, или от крошечной капли кислоты и все еще давать высокое качество звучания при проигрывании (мы категорически не рекомендуем проводить такой эксперимент!). Это наглядно иллюстрирует способности корректирующих кодов. С технической точки зрения типичные цели передачи информации следующие:

1) быстрое кодирование информации,
2) удобная передача закодированных данных,
3) эффективное использование доступного канала (т. е. максимальная передача информации в единицу времени),
4) быстрое декодирование,
5) исправление ошибок (как можно большего количества), внесенных шумом в канале. □

Как правило, эти цели противоречат друг другу, и необходимо найти оптимальное решение. Как раз об этом и говорит данная глава. К со-

жалению, не приходится ожидать идеального решения: теория, которая будет изложена, в основном нацелена на изложение основных принципов. Окончательное решение всегда ложится на плечи ответственного лица (или группы). Большая часть этого параграфа (и вся глава 1) будет посвящена проблемам кодирования. Целями кодирования являются:

i) сжатие данных для уменьшения избыточной информации, содержащейся в сообщении;

ii) защита текста от несанкционированного пользователя;

iii) исправление допущенных ошибок.

Мы начинаем с изучения *источников* и *кодеров*. Источник генерирует последовательность букв (или символов)

$$u_1 u_2 \ldots u_n \ldots, \tag{1.1.1}$$

где $u_j \in I$ $(= I_m)$ — m-элементное множество, часто отождествляемое с $\{1, \ldots, m\}$ (*алфавит источника*). В случае литературного английского языка $m = 26 + 7$, 26 букв плюс 7 символов пунктуации: . , : ; – ( ). (Иногда добавляют знаки ? ! ’ ‘ и ".) В случае английского телеграфа $m = 27$.

Как правило, последовательность (1.1.1) рассматривают как *выборку* из случайного источника, т. е. последовательности с. в.

$$U_1, U_2, \ldots, U_n, \ldots, \tag{1.1.2}$$

и пытаются разработать теорию для разумного класса таких последовательностей.

Пример 1.1.2. а) Простейший пример случайного источника — это последовательность н. о. р. с. в.:

$$\mathsf{P}(U_1 = u_1, U_2 = u_2, \ldots, U_k = u_k) = \prod_{j=1}^{k} p(u_j), \tag{1.1.3а}$$

где $p(u) = \mathsf{P}(U_j = u)$, $u \in I$, — распределение одной с. в. Случайный источник с независимыми одинаково распределёнными символами часто называют *источником Бернулли*.

Частный случай, когда вероятность $p(u)$ не зависит от $u \in I$ (и поэтому равна $1/m$), соответствует *равновероятному источнику Бернулли*.

б) Более общим примером служит *источник Маркова*, где последовательность символов представляет собой дискретную цепь Маркова:

$$\mathsf{P}(U_1 = u_1, U_2 = u_2, \ldots, U_k = u_k) = \lambda(u_1) \prod_{j=1}^{k-1} P(u_j, u_{j+1}), \tag{1.1.3б}$$

где $\lambda(u) = \mathsf{P}(U_1 = u)$, $u \in I$, — начальное распределение и $P(u, u') = \mathsf{P}(U_{j+1} = u' \,|\, U_j = u)$, $u, u' \in I$, — вероятность перехода. Источник Мар-

кова называется *стационарным*, если $\mathsf{P}(U_j = u) = \lambda(u)$, $j \geqslant 1$, т. е. $\lambda = \{\lambda(u),\ u = 1, \ldots, m\}$ — инвариантная вектор-строка матрицы

$$P = \{P(u, v)\} : \sum_{u \in I} \lambda(u) P(u, v) = \lambda(v), \quad v \in I,$$

или, короче, $\lambda P = \lambda$.

«Вырожденным» примером источника Маркова является источник, генерирующий повторяющиеся символы. В этой ситуации

$$\mathsf{P}(U_1 = U_2 = \ldots = U_k = u) = p(u), \ \ u \in I,$$
$$\mathsf{P}(U_k \neq U_{k'}) = 0, \ \ 1 \leqslant k < k', \tag{1.1.3в}$$

где $0 \leqslant p(u) \leqslant 1$ и $\sum\limits_{u \in I} p(u) = 1$. □

Начальный участок последовательности (1.1.1)

$$\mathbf{u}(n) = (u_1, u_2, \ldots, u_n), \quad \text{или} \quad \mathbf{u}(n) = u_1 u_2 \ldots u_n,$$

называется выборочной (из источника) *n-строкой* или *n-словом* (или, короче, строкой или словом), Соответственно рассматриваются случайные n-строки (случайные сообщения):

$$\mathbf{U}^{(n)} = (U_1, U_2, \ldots, U_n), \quad \text{или, иначе,} \quad \mathbf{U}^{(n)} = U_1 U_2 \ldots U_n.$$

Кодер использует *алфавит* $J (= J_q)$, который мы обычно будем записывать как $0, 1, \ldots, q-1$; как правило, число кодирующих символов q меньше m (или даже $q \ll m$); во многих случаях $q = 2$ и $J = \{0, 1\}$ (двоичный кодер). *Код* (также кодировка) — это отображение f, переводящее символ $u \in I$ в конечное слово $f(u) = x_1 \ldots x_s$, знаки которого выбираются из J. Иначе говоря, f отображает I в множество J^* всех возможных строк:

$$f : I \to J^* = \bigcup_{s \geqslant 1} \underbrace{J \times J \times \ldots \times J}_{s \text{ раз}}.$$

Строка $f(u)$, являющаяся образом при отображении f символов $u \in I$, называется *кодовым словом* (в коде f). Говорят, что код имеет (постоянную) длину N, если величина s (длина кодового слова) равна N для всех кодовых слов. Сообщение $\mathbf{u}(n) = u_1 u_2 \ldots u_n$ представляется как сцепление кодовых слов:

$$f(\mathbf{u}^{(n)}) = f(u_1) f(u_2) \ldots f(u_n),$$

т. е. снова как строка из J^*.

Определение 1.1.3. Мы говорим, что данный код является *кодом без потерь*, если из предположения $u \neq u'$ следует, что $f(u) \neq f(u')$ (т. е.

отображение $f\colon I \to I^*$ — вложение[1]). Код допускает декодирование (или декодируемый), если любая строка из I^* изображает не более одного сообщения. Строка x служит *префиксом* в другой строке y, если $y = xz$, т. е. строка y может быть представлена как результат сцепления x и z. Код называют *свободным от префиксов*, или *беспрефиксным*, если ни одно из кодовых слов не является префиксом другого (например, код постоянной длины беспрефиксный). □

Беспрефиксный код допускает декодирование, однако обратное утверждение неверно.

Пример 1.1.4. Код с трехсимвольным алфавитом источника $I = \{1, 2, 3\}$ и двоичным алфавитом кодера $J = \{0, 1\}$, заданный соотношениями

$$f(1) = 0, \quad f(2) = 01, \quad f(3) = 011,$$

декодируемый, но не свободен от префиксов. □

Теорема 1.1.5 (неравенство Крафта). *Для данных натуральных чисел $s_1, \ldots, s_m$ декодируемый код $f\colon I \to I^*$ с кодовыми словами длин $s_1, \ldots, s_m$ существует тогда и только тогда, когда*

$$\sum_{i=1}^{m} q^{-s_i} \leqslant 1. \tag{1.1.4}$$

Более того, если выполнено неравенство (1.1.4), *то существует беспрефиксный код с кодовыми словами длин $s_1, \ldots, s_m$.*

Доказательство. 1. Достаточность. Пусть неравенство (1.1.4) выполнено. Нам нужно построить беспрефиксный код с кодовыми словами длин $s_1, \ldots, s_m$. Перепишем неравенство (1.1.4) как

$$\sum_{l=1}^{s} n_l q^{-l} \leqslant 1, \tag{1.1.5}$$

или

$$n_s q^{-s} \leqslant 1 - \sum_{l=1}^{s-1} n_l q^{-l},$$

где n_l — количество слов длины l и $s = \max[s_i]$. Распишем последнее неравенство подробнее:

$$n_s \leqslant q^s - n_1 q^{s-1} - \ldots - n_{s-1} q. \tag{1.1.6.1}$$

Поскольку $n_s \geqslant 0$, получаем

$$n_{s-1} q \leqslant q^s - n_1 q^{s-1} - \ldots - n_{s-2} q^2,$$

[1] В оригинале использован термин «one-to-one map», т. е. «взаимно однозначное отображение», но мы следуем здесь российской математической традиции. — *Прим. перев.*

или

$$n_{s-1} \leqslant q^{s-1} - n_1 q^{s-2} - \ldots - n_{s-2} q. \qquad (1.1.6.2)$$

Продолжая рассуждения, последовательно приходим к неравенствам

$$n_{s-2} \leqslant q^{s-2} - n_1 q^{s-3} - \ldots - n_{s-3} q,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$

$$n_2 \leqslant q^2 - n_1 q, \qquad (1.1.6.s-1)$$

$$n_1 \leqslant q. \qquad (1.1.6.s)$$

Отметим, что на самом деле или $n_{i+1} = 0$, или n_i меньше п. ч. неравенства $\forall\, i = 1, \ldots, s-1$ (по определению $n_s \geqslant 1$, так что для $i = s-1$ имеет место вторая возможность). Проделаем следующее. Сначала выпишем n_1 слов длины 1, используя различные символы из J (это возможно в силу неравенства (1.6.s)). Осталось $q - n_1$ неиспользованных символов; далее, сформируем $(q - n_1)q$ слов длины 2, дописывая к каждому неиспользованному символу ещё по одному. Выберем n_2 таких слов (что можно сделать ввиду неравенства (1.6.s–1). У нас все еще остаётся $q^2 - n_1 q - n_2$ неиспользованных слова длины 2. Строим n_3 слов длины 3 и т. д. По построению ни одно из новых слов не будет содержать предыдущих в качестве префикса. Следовательно, построенный код свободен от префиксов.

2. Необходимость. Предположим, что существует декодируемый код с алфавитом J с кодовыми словами длин $s_1, \ldots, s_m$. Вновь положим $s = \max\{s_i\}$ и заметим, что для любого натурального r выполняется равенство

$$(q^{-s_1} + \ldots + q^{-s_m})^r = \sum_{l=1}^{rs} b_l q^{-l},$$

где b_l — количество способов получения из r слов строки длины l.

Ввиду декодируемости эти строки должны быть различными. Значит, $b_l \leqslant q^l$, поскольку q^l — число всех строк длины l. Следовательно,

$$(q^{-s_1} + \ldots + q^{-s_m})^r \leqslant rs,$$

и

$$q^{-s_1} + \ldots + q^{-s_m} \leqslant r^{1/r} s^{1/r} = \exp\left(\frac{1}{r}(\log r + \log s)\right).$$

Так как это верно при любом r, переходя к пределу при $r \to \infty$, мы увидим, что правая часть неравенства стремится к 1. □

Замечание 1.1.6. Код, подчиняющийся неравенству (1.1.4), не обязательно допускает декодирование. □

Леон Г. Крафт вывел неравенство (1.1.4) в своей диссертации на звание доктора философии в Массачусетском технологическом институте в 1949 г.

Одна из основных целей теории состоит в том, чтобы найти «лучший» (т. е. кратчайший) декодируемый (или беспрефиксный) код. Встанем сейчас на вероятностную точку зрения и предположим, что символ $u \in I$ генерируется источником с вероятностью $p(u)$:

$$\mathsf{P}(U_k = u) = p(u).$$

(На данный момент нет необходимости указывать совместную вероятность более чем одного сгенерированного символа.)

Напомним, что данным кодом $f\colon I \to J^*$ мы кодируем букву $i \in I$, предписывая ей кодовое слово $f(i) = x_1 \dots x_{s(i)}$ длины $s(i)$. Для произвольного символа сгенерированное кодовое слово становится случайной строкой из J^*. Пусть f — код без потерь, тогда вероятность получения данной строки как кодового слова для символа в точности равна $p(i)$, если эта строка совпадает с $f(i)$, и 0, если нет такой буквы $i \in I$, для которой это так.

Если f не вложение, то вероятность строки равна сумме членов $p(i)$, для которых кодовое слово $f(i)$ равно данной строке. Таким образом, длина кодового слова тоже случайная величина S с распределением вероятностей

$$\mathsf{P}(S = s) = \sum_{1 \leqslant i \leqslant m} \mathbf{1}(s(i) = s) p(i).$$

Мы ищем такой декодируемый код, чтобы минимизировать *математическое ожидание длины слова*

$$\mathsf{E}S = \sum_{s \geqslant 1} s\, \mathsf{P}(S = s) = \sum_{i=1}^{m} s(i) p(i).$$

Таким образом, возникает следующая задача:

$$\begin{gathered}\text{минимизировать } g(s(1) \dots s(m)) = \mathsf{E}S \text{ при условии} \\ \sum_i q^{-s(i)} \leqslant 1 \text{ (Крафт), где } s(i) \text{ — натуральные числа.}\end{gathered} \tag{1.1.7}$$

Теорема 1.1.7. *Оптимальное решение задачи* (1.1.7) *удовлетворяет неравенству*

$$\min \mathsf{E}S \geqslant h_q(p(1), \dots, p(m)), \tag{1.1.8}$$

где

$$h_q(p(1), \dots, p(m)) = -\sum_i p(i) \log_q p(i). \tag{1.1.9a}$$

Доказательство. Задача (1.1.7) — это целочисленная задача оптимизации. Если заменить условие $s(1), \dots, s(m) \in \{1, 2, \dots\}$ более слабым: $s(i) > 0$, $1 \leqslant i \leqslant m$, то можно использовать теорему Лагранжа об

условном минимуме. Функция Лагранжа в этом случае выглядит так:

$$\mathcal{L}(s(1), \ldots, s(m), z; \lambda) = \sum_i s(i)p(i) + \lambda\Big(1 - \sum_i q^{-s(i)} - z\Big)$$

(здесь $z \geqslant 0$ — резервная переменная). Минимизируя $\mathcal{L}$ по $s(1), \ldots, s(m)$ и z, получаем

$$\lambda < 0, \;\; z = 0 \;\; \text{и} \;\; \frac{\partial \mathcal{L}}{\partial s(i)} = p(i) + q^{-s(i)}\lambda \ln q = 0,$$

откуда следует, что

$$-\frac{p(i)}{\lambda \ln q} = q^{-s(i)}, \quad \text{т. е.} \;\; s(i) = -\log_q p(i) + \log_q(-\lambda \ln q), \quad 1 \leqslant i \leqslant m.$$

Учитывая ограничение $\sum\limits_i q^{-s(i)} = 1$ (резервная переменная $z = 0$), получаем

$$\sum_i p(i)/(-\lambda \ln q) = 1, \quad \text{т. е.} \quad -\lambda \ln q = 1.$$

Следовательно, набор

$$s(i) = -\log_q p(i), \quad 1 \leqslant i \leqslant m,$$

является (единственным) решением задачи минимизации, определяющим значение h_q из формулы (1.1.9а). Мы нашли решение задачи минимизации на бóльшем множестве, чем требовалось, поэтому минимальное значение не больше, чем п. ч. формулы (1.1.8). □

Замечание 1.1.8. Величина h_q, определенная формулой (1.1.9а), играет центральную роль в теории информации. Она называется *q-ичной энтропией* распределения вероятностей $(p(x), x \in I)$ и будет появляться во многих ситуациях. Отметим, что справедлива формула

$$h_q(p(1), \ldots, p(m)) = \frac{1}{\log q} h_2(p(1), \ldots, p(m)),$$

где h_2 — двоичная энтропия,

$$h_2(p(1), \ldots, p(m)) = -\sum_i p(i) \log p(i). \tag{1.1.9б}$$

□

Пример 1.1.9. а) Приведите пример кода без потерь с алфавитом J_q, не удовлетворяющего неравенству Крафта. Приведите пример кода без потерь, в котором математическое ожидание длины кодового слова строго меньше чем $h_q(X)$.

б) Покажите, что «сумма Крафта» $\sum\limits_i q^{-s(i)}$, выписанная по коду без потерь, может быть сколь угодно большой (при достаточно большом алфавите источника).

Решение. а) Рассмотрим алфавит $I = \{0, 1, 2\}$ и код без потерь f с $f(0) = 0$, $f(1) = 1$, $f(2) = 00$. Тогда длины кодовых слов таковы: $s(0) = s(1) = 1$, $s(2) = 2$. Очевидно, что $\sum\limits_{x \in I} 2^{-s(x)} = 5/4$, что противоречит неравенству Крафта. Для с. в. X с $p(0) = p(1) = p(2) = 1/3$ математическое ожидание длины кодового слова равно $\mathsf{E}S(X) = 4/3 < h(X) = \log 3 \approx 1{,}585$.

б) Предположим, что размер алфавита m равен $\#I = 2(2^L - 1)$ для некоторого натурального числа L. Рассмотрим код без потерь, сопоставляющий буквам $x \in I$ кодовые слова 0, 1, 00, 01, 10, 11, 000, ..., достигающие максимальной длины L. Суммой Крафта будет

$$\sum_{x \in I} 2^{-s(x)} = \sum_{l \leqslant L} \sum_{x:\ s(x) = l} 2^{-s(x)} = \sum_{l \leqslant L} 2^l \times 2^{-l} = L,$$

что может быть сколь угодно большим. □

Теорема 1.1.7 получает дальнейшее развитие в теореме 1.1.10.

Теорема 1.1.10 (теорема Шеннона для канала без шума). *Для случайного источника, генерирующего символы с вероятностью $p(i) > 0$, минимальное математическое ожидание длины кодового слова в декодируемом кодере с алфавитом J_q подчиняется неравенству*

$$h_q \leqslant \min \mathsf{E}S < h_q + 1, \tag{1.1.10}$$

где $h_q = -\sum\limits_i p(i) \log_q p(i)$ — q-ичная энтропия источника (ср. с формулой (1.1.9а)).

Доказательство. Л. ч. неравенства следует из теоремы 1.1.7. Для доказательства п. ч. выберем такое натуральное число $s(i)$, что

$$q^{-s(i)} \leqslant p(i) < q^{-s(i)+1}.$$

Отсюда следует нестрогая оценка $\sum\limits_i q^{-s(i)} \leqslant \sum\limits_i p(i) = 1$, т. е. неравенство Крафта. Значит, найдется декодируемый код с кодовыми словами длин $s(1), \ldots, s(m)$. Из п. ч. неравенства вытекает, что

$$s(i) < -\frac{\log p(i)}{\log q} + 1,$$

и поэтому

$$\mathsf{E}S < -\frac{\sum\limits_i p(i) \log p(i)}{\log q} + \sum_i p(i) = \frac{h}{\log q} + 1. \qquad \square$$

Остерегайтесь найти то, что вы ищете.

Ричард Хэмминг (1915–1998),
американский математик и программист

Пример 1.1.11. Поучительное приложение теоремы Шеннона 1.1.10 заключается в следующем. Пусть размер m алфавита источника равен 2^k, и предположим, что буквы $i = 1, \ldots, m$ генерируются равновероятно: $p(i) = 2^{-k}$. Допустим, алфавит кодера — это $J_2 = \{0, 1\}$ (двоичный кодер). Так как двоичная энтропия равна $h_2 = -\log 2^{-k} \sum\limits_{1 \leqslant i \leqslant 2^k} 2^{-k} = k$, нам потребуется (в среднем) по крайней мере k двоичных знаков для декодируемого кодера. Используя термин «бит» для единицы энтропии, мы говорим, что в среднем кодирование требует по крайней мере k битов. Кроме того, теорема 1.1.10 приводит к процедуре кодирования Шеннона—Фано: мы фиксируем в качестве длин кодовых слов такие натуральные числа $s(1), \ldots, s(m)$, что $q^{-s(i)} \leqslant p(i) < q^{-s(i)+1}$, или, что эквивалентно,

$$-\log p(i) \leqslant s(i) < -\log p(i) + 1, \quad \text{т. е.} \quad s(i) = \lceil -\log p(i) \rceil. \qquad (1.1.11)$$

Затем мы строим беспрефиксный код, начиная со слова наименьшей длины $s(i)$, постепенно увеличивая длину слов и следя за тем, чтобы предыдущие кодовые слова не являлись префиксами. Неравенство Крафта гарантирует, что нам это удастся. Полученный код может не быть оптимальным, но средняя длина его кодового слова удовлетворяет неравенству (1.1.11), как в оптимальном коде. □

Оптимальность достигается при *кодировании Хаффмана* $f_m^{\mathrm{H}}\colon I_m \to J_q^*$. Сначала мы обсудим случай двоичной кодировки, т. е. $q = 2$ и $J = \{0, 1\}$. Алгоритм построения бинарного дерева выглядит следующим образом.

1. Во-первых, упорядочим буквы $i \in I$ так, что

$$p(1) \geqslant p(2) \geqslant \ldots \geqslant p(m).$$

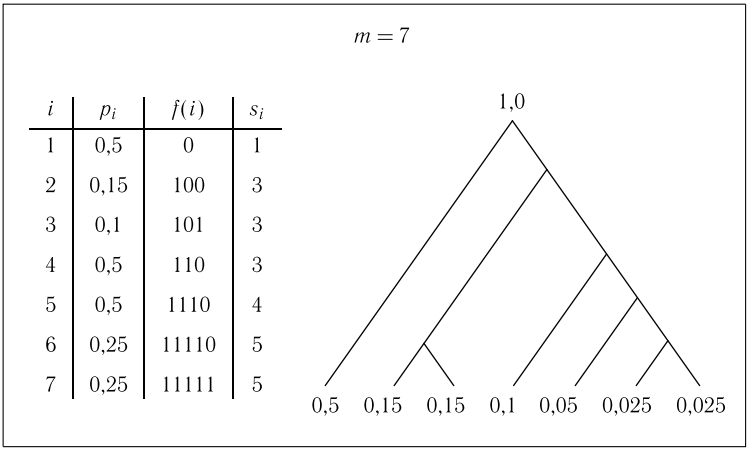
2. Припишем символ 0 букве $m - 1$ и символ 1 букве m.
3. Построим редуцированный алфавит

$$I_{m-1} = \{1, \ldots, m-2, (m-1, m)\}$$

с вероятностями $p(1), \ldots, p(m-2), p(m-1) + p(m)$.

Повторим шаги 1 и 2 с редуцированным алфавитом, и т. д. Мы получаем двоичное дерево, листья которого соответствуют буквам алфавита источника, а вершины — группам объединенных символов. Пример кодирования Хаффмана для $m = 7$ приведён на рис. 1.1.

Количество ветвей, которое мы должны пройти, от листа i до корня, равно $s(i)$. Структура дерева, листья которого отождествляются с буквами алфавита, гарантирует, что код будет свободен от префиксов. Оптимальность двоичного кодирования Хаффмана вытекает из следующих двух простых лемм.



i	p_i	$f(i)$	s_i
1	0,5	0	1
2	0,15	100	3
3	0,1	101	3
4	0,5	110	3
5	0,5	1110	4
6	0,25	11110	5
7	0,25	11111	5

Рис. 1.1

Лемма 1.1.12. *У любого оптимального беспрефиксного двоичного кода длины кодовых слов упорядочены противоположно вероятностям*:

$$p(i) \geqslant p(i') \quad \Rightarrow \quad s(i) \leqslant s(i'). \tag{1.1.12}$$

Доказательство. Если это не так, то можно модифицировать код, поменяв местами кодовые слова для i и i'. Это уменьшит математическое ожидание длины кодового слова, сохранив свободу от префиксов. □

Лемма 1.1.13. *В любом оптимальном свободном от префиксов двоичном коде среди кодовых слов максимальной длины существуют ровно два совпадающих всюду, кроме последнего знака.*

Доказательство. Если это не так, то либо 1) существует единственное слово максимальной длины, либо 2) найдутся два или больше слов максимальной длины, но отличаться друг от друга они будут ранее последнего знака. В обоих случаях последний знак из некоторого слова максимальной длины можно выбросить, не влияя на свободу от префиксов. □

Теорема 1.1.14. *Код Хаффмана — оптимальный среди всех беспрефиксных двоичных кодов.*

Доказательство. Доказательство проведём индукцией по m. При $m = 2$ код Хаффмана f_2^{H} определяется так: $f_2^{\mathrm{H}}(1) = 0$, $f_2^{\mathrm{H}}(2) = 1$ или наоборот, и он является оптимальным. Предположим, что код Хаффмана f_{m-1}^{H} оптимален для I_{m-1} при любом распределении вероятностей. Допустим, далее, что код Хаффмана f_m^{H} не оптимален для I_m при некотором распреде-

лении вероятностей. Иначе говоря, существует другой беспрефиксный код f^*_m, для I_m с меньшим математическим ожиданием длины кодового слова:

$$\mathsf{E}S^*_m < \mathsf{E}S^{\mathrm{H}}_m. \tag{1.1.13}$$

Без ограничения общности можно считать, что распределение вероятностей удовлетворяет неравенствам

$$p(1) \geqslant \ldots \geqslant p(m).$$

По леммам 1.1.12 и 1.1.13 можно перенумеровать кодовые слова в обоих кодах так, что слова, соответствующие $m-1$ и m, будут иметь максимальную длину и отличаться друг от друга лишь в последнем знаке. Это позволяет свести оба кода к I_{m-1}. Действительно, в коде Хаффмана f^{H}_m мы удалим последний знак из $f^{\mathrm{H}}_m(m)$ и $f^{\mathrm{H}}_m(m-1)$, «склеив» эти слова. Эта процедура приведёт нас к коду Хаффмана f^{H}_{m-1}. С кодом f^*_m поступим аналогично, и получим новый беспрефиксный код f^*_{m-1}.

Заметим, что в коде Хаффмана f^{H}_m вклад в $\mathsf{E}S^{\mathrm{H}}_m$ из $f^{\mathrm{H}}_m(m-1)$ и $f^{\mathrm{H}}_m(m)$ равен $s^{\mathrm{H}}(m)(p(m-1)+p(m))$, и после редуцирования он становится равен $(s^{\mathrm{H}}(m)-1)(p(m-1)+p(m))$, т. е. $\mathsf{E}S$ уменьшается на $p(m-1)+p(m)$. В коде f^*_m аналогичный вклад уменьшается с $s^*(m)(p(m-1)+p(m))$ до $(s^*(m)-1)(p(m-1)+p(m))$. Значит, и здесь разница составит $p(m-1)+$ $+\,p(m)$. Все остальные вклады в $\mathsf{E}S^{\mathrm{H}}_{m-1}$ и $\mathsf{E}S^*_{m-1}$ совпадают с соответствующими вкладами в $\mathsf{E}S^{\mathrm{H}}_m$ и $\mathsf{E}S^*_m$. Следовательно, код f^*_{m-1} предпочтительнее, чем f^{H}_{m-1}: $\mathsf{E}S^*_{m-1} < \mathsf{E}S^{\mathrm{H}}_{m-1}$, что противоречит предположению индукции. □

С учетом теоремы 1.1.14 мы получаем следующий результат.

Следствие 1.1.15. *Кодирование Хаффмана является оптимальным среди всех декодируемых двоичных кодов.* □

Обобщение конструкции Хаффмана на q-ичные коды (с кодовым алфавитом $J_q = \{0, 1, \ldots, q-1\}$) проводится следующим образом: вместо объединения двух символов $m-1$, $m \in I_m$, имеющих наименьшую вероятность, мы объединяем q из них (тоже с наименьшей вероятностью), повторяя предыдущие рассуждения. Фактически оригинальная работа Хаффмана (1952 г.) была посвящена общему алфавиту кодера. Существуют многочисленные модификации кода Хаффмана, учитывающие разную стоимость кодирования (где некоторые символы $j \in J_q$ более дорогостоящие, чем другие) и другие факторы. Мы не будем обсуждать их в этой книге.

Дэвид Хаффман умер 7 октября 1999 г. в городе Санта-Круз, Калифорния, в возрасте 74 лет. Он был не только блестящим ученым, но и яркой личностью. Легенда говорит, что он изобрёл свой метод кодирования в 1951 г. при написании курсовой работы (часть экзамена в Массачусетском технологическом институте), поставленной перед ним его руководителем, профессором Робертом Фано (который в то время был ближайшим

соратником Шеннона). Фано (род. в 1917 г.) тоже оставил заметный след в теории информации. Его имя не однажды появится на страницах нашей книги (неравенство Фано и обобщённое неравенство Фано). Он родился в известной математической семье: его отец Джино Фано был выдающимся членом итальянской школы алгебраической геометрии (ведущая группа в мире в этой области в первой половине XX века), а его старший брат Уго Фано внёс основополагающий вклад в теоретическую физику.

Пример 1.1.16. Недостаток кодера Хаффмана состоит в том, что длина кодового слова является довольно сложной функцией вероятностей символов $p(1), \ldots, p(m)$. Однако нетрудно получить некоторые оценки. Предположим, что $p(1) \geqslant p(2) \geqslant \ldots \geqslant p(m)$. Докажите, что в любом двоичном коде Хаффмана

1) если $p(1) < 1/3$, то при $m > 2$ буква 1 должна быть закодирована кодовым словом длины не меньше 2,
2) если $p(1) > 2/5$, то длина кодового слова, соответствующего 1, должна равняться 1.

Решение. 1. Возможны два случая: буква 1 а) объединилась с другими буквами раньше последнего шага при построении кода Хаффмана и б) не объединялась. В случае а) имеем $s(1) \geqslant 2$. В случае б) у нас на третьем с конца шаге алгоритма есть символы 1, b и b' и верхняя часть дерева Хаффмана устроена, как на рис. 1.2 а, причём $0 \leqslant p(b), p(b') \leqslant 1 - p(1)$ и $p(b) + p(b') = 1 - p(1)$. Но тогда $\max[p(b), p(b')] > 1/3$, и поэтому $p(1)$ должна быть объединена с $\min[p(b), p(b')]$. Значит, рис. 1.2 а невозможен, и длина кодового слова буквы 1 не меньше 2.

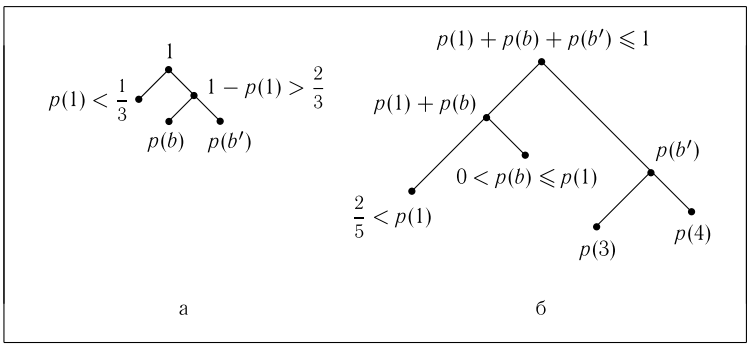


Рис. 1.2

Эта граница является точной, поскольку оба кода

$$\{0,\ 01,\ 110,\ 111\} \quad \text{и} \quad \{00,\ 01,\ 10,\ 11\}$$

являются двоичными кодами Хаффмана для распределения вероятностей 1/3, 1/3, 1/4, 1/12.

2. Пусть теперь $p(1) > 2/5$, а длина кодового слова в коде Хаффмана, соответствующего 1, не меньше 2. Тогда при построении двоичного дерева буква 1 объединялась на каком-то шаге, кроме последнего, с другими буквами. Значит, на каком-то подходящем шаге у нас были символы 1, b и b' с возможным распределением вероятностей

$$\text{А) } p(b') \geqslant p(1) > 2/5, \quad \text{Б) } p(b') \geqslant p(b),$$
$$\text{В) } p(1) + p(b) + p(b') \leqslant 1 \quad \text{и} \quad \text{Г) } p(1),\ p(b) \geqslant \frac{1}{2}p(b').$$

Действительно, если, скажем, $p(b) < p(b')/2$, то b должно быть отобрано для объединения вместо $p(3)$ или $p(4)$, объединение которых давало $p(b')$. Ввиду неравенства Г имеем $p(b) > 1/5$, что делает случай одновременного выполнения неравенств А, Б и В невозможным.

Тогда часть дерева Хаффмана над $p(1)$ устроена, как на рис. 1.2 б, причём $p(3) + p(4) = p(b')$ и $p(1) + p(b') + p(b) \leqslant 1$. Запишем

$$p(1) = 2/5 + \varepsilon, \quad p(b') = 2/5 + \varepsilon + \delta, \quad p(b) = 2/5 + \varepsilon + \delta - \eta,$$

где $\varepsilon > 0$, δ, $\eta \geqslant 0$. Итак,

$$p(1) + p(b') + p(b) = 6/5 + 3\varepsilon + 2\delta - \eta \leqslant 1, \quad \text{и} \quad \eta \geqslant 1/5 + 3\varepsilon + 2\delta.$$

Отсюда получаем

$$p(b) \leqslant 1/5 - 2\varepsilon - \delta < 1/5.$$

Однако так как

$$\max\{p(3),\ p(4)\} \geqslant p(b')/2 \geqslant p(1)/2 > 1/5,$$

вероятность $p(b)$ должна быть объединена с $\min[p(3),\ p(4)]$, т. е. диаграмма, изображённая на рис. 1.2 б, невозможна. Следовательно, длина кодового слова, соответствующего 1, $s(1) = 1$. □

Пример 1.1.17. Предположим, что буквы $i_1, \ldots, i_5$ генерируются с вероятностями 0,45, 0,25, 0,2, 0,05, 0,05. Вычислите математическое ожидание длины кодового слова для кодов Шеннона—Фано и Хаффмана. Проиллюстрируйте оба метода выписыванием декодируемых двоичных кодировок в каждом случае.