1990 г. январь — февраль *т. 45, вып. 1 (271)*
УСПЕХИ МАТЕМАТИЧЕСКИХ НАУК

УДК 517.11

МОЖЕТ ЛИ (ИНДИВИДУАЛЬНАЯ) ПОСЛЕДОВАТЕЛЬНОСТЬ НУЛЕЙ И ЕДИНИЦ БЫТЬ СЛУЧАЙНОЙ?

В. А. Успенский, А. Л. Семенов, А. Х. Шень

СОДЕРЖАНИЕ

Введение

Если кто-либо предъявит нам конечную последовательность

(1) 000000000000

или

(2) 010101010101

и скажет, что она получена в результате случайного бросания совершенно симметричной монеты (при соотнесении, например, 0 с гербом и 1 с решеткой), мы усомнимся в честности эксперимента, однако последовательность

(3) 011001011010

не покажется нам подозрительной.

Таким образом, наша интуиция довольно четко выделяет среди последовательностей случайные и неслучайные. Одна из главных задач математики — подкреплять (как, например, в случае теоремы Жордана о разбиении плоскости гомеоморфом окружности) или опровергать (как в случае математических парадоксов) и тем уточнять нашу интуицию. В состоянии ли математика справиться с подобной задачей в данном случае? О последовательностях типа (1) и (2) говорят обычно, что они не случайны, потому что вероятность появления любой из них очень мала (2^{-12}). Но ведь в точности такова же и вероятность появления последовательности (3)! [Мы остановимся подробнее на этой аргументации в дополнении к статье.]

Проблема разделения конечных последовательностей на случайные и неслучайные (а точнее, классификации по «степени случайности», поскольку вряд ли можно провести какую-либо четкую границу) рассматривалась в работах А. Н. Колмогорова и его учеников ([Колм 65], [Колм 69], [Колм 83], [Колм 83а], [Зво Лев 70], [Лев 74], [Лев 76], [Лев 76а], [Лев 76б], [Лев 77], [Лев 80], [Аса 86], [Аса 86а], [Аса 87], [Аса 87а], [Вовк 85], [Вовк 86], [Вовк 86а], [Вовк 86б], [Вовк 87], [Вовк 87а], [Вовк 88], [Вью 86], [Вью 87], [Шень 83]). Заметим, что эта проблема связана с вопросом о теоретическом обосновании метода Монте-Карло.

В настоящем обзоре мы рассматриваем более простую ситуацию, когда рассматриваемые последовательности бесконечны. Все бесконечные последовательности нулей и единиц, возникающие в результате (на этот раз только мысленного) эксперимента, состоящего в бесконечном числе бросаний монеты, имеют каждая одну и ту же вероятность — нулевую. Однако и в этом случае интуиция выделяет среди них случайные и неслучайные. (Здесь, в отличие от случая конечных последовательностей, можно надеяться провести между случайными и неслучайными последовательностями четкую границу.) Изучение бесконечных последовательностей хотя и дальше от практики, чем изучение конечных, но зато проще, и поэтому первое может трактоваться как введение во второе (а само понятие бесконечной последо-

вательности — как математическая модель для понятия «очень длинная последовательность». Подобный подход, когда бесконечное возникает как «аппроксимация конечного», вообще типичен для математики — сравни переход от рациональных чисел к действительным, от мощностей реально встречающихся множеств к натуральным числам или от молекулярно-кинетической теории к уравнению теплопроводности.

В нашей статье мы ограничиваемся последовательностями, состоящими только из нулей и единиц (переход к любому конечному числу значений непринципиален); для множества всех таких последовательностей — как конечных, так и бесконечных — мы закрепляем обозначение Σ; для множества всех конечных последовательностей (иначе — двоичных слов) — обозначение Ξ, а для множества всех бесконечных последовательностей — обозначение Ω. Таким образом, $\Sigma = \Xi \bigcup \Omega$. Часто — в том числе в названии этой статьи — только элементы Ω называются последовательностями; из контекста, как правило, ясно, в каком смысле — узком (элемент Ω) или широком (элемент Σ) употреблен термин «последовательность».

Наша задача — выделить среди элементов Ω случайные. Традиционная теория вероятностей оказывается бессильной перед подобной задачей. Она вообще не утверждает ничего ни про какую отдельную последовательность, а говорит только про совокупности последовательностей. Если в теории вероятностей и говорят «возьмем случайную последовательность», то это всего лишь вольность речи, «abus de langage» по Бурбаки: когда в дальнейшем про эту «взятую» последовательность нечто утверждается, это означает лишь, что некое свойство выполнено для «подавляющего большинства» последовательностей.

В то же время ясно, что задача дать математическое определение понятию «случайная последовательность» весьма интересна. Впервые эта задача была рассмотрена фон Мизесом в [Миз 19]; о его подходе к определению понятия случайной последовательности (в терминологии Мизеса — «Kollektiv») см. далее, в главе 6. Там же мы обсудим и дальнейшее развитие подхода Мизеса, именуемого также частотным подходом.

Принципиально иной — сложностной — подход к определению понятия случайной последовательности был предложен А. Н. Колмогоровым ([Колм 65], [Колм 69]) и разработан далее Л. А. Левиным ([Лев 73]) и К. П. Шнорром ([Шно 73], [Шно 77]). Этому подходу посвящена глава 3 нашего обзора.

Наконец, учеником А. Н. Колмогорова шведским математиком П. Мартин-Лёфом ([Мар 66], [Мар 66а], [Мар 70]) был разработан количественный, или теоретико-мерный, подход к уточнению понятия случайности. В [Зво Лев 70, § 4, п. 1] после указания некоторых трудностей, связанных с частотным подходом, говорится: «В 1965 г. П. Мартин-Лёфу удалось, основываясь на идеях А. Н. Комогорова, дать свободное от подобных трудностей определение понятия случайной последовательности. Идея А. Н. Колмогорова состояла в том, чтобы «не случайными» считать последовательности, в которых наблюдается достаточно много закономерностей, где под закономерностью подразумевается любое проверяемое свойство последовательности, присущее лишь узкому их классу (достаточно малому по мере)». Количественному (теоретико-мерному) подходу посвящена глава 2.

Замечательно, что каждый из указанных трех подходов к определению того, что такое индивидуальная случайная последовательность, оформляется в виде математического (на уровне сегодняшних представлений о математической строгости) построения на основе теории алгоритмов. Умение дать такое определение (точнее, несколько вариантов таких определений, некоторые из которых следует, быть может, рассматривать как окончательные) авторы склонны считать одним из высших достижений теории алгоритмов. Тем самым теория алгоритмов позволяет дать утвердительный (хотя, быть может, и не окончательный) ответ на вопрос, вынесенный в заголовок статьи.

Как уже отмечалось, каждому из трех подходов к определению случайности — частотному (Р. фон Мизес), сложностному (А. Н. Колмогоров) и количественному (П. Мартин-Лёф) посвящена соответствующая глава статьи. Каждую из этих глав авторы старались сделать по возможности независимой от остальных. Однако сперва все эти подходы кратко излагаются в главе 1, чтобы читатели, ограничившие свое чтение этой главой, могли получить первоначальное представление о сути дела.

Промежуточным между сложностным и теоретико-мерным является подход, связанный с так называемыми вероятностными машинами, составляющий содержание главы 5. Глава 4 содержит сравнение результатов сложностного и количественного подходов.

Мы предполагаем, что читатель знаком с началами теории вероятностей и теории меры, позволяя себе, например, иногда употреблять слова «мера на борелевских подмножествах пространства Ω» без разъяснений. К сожалению, другая важная для нас область — теория алгоритмов — до сих пор не входит в общематематическое образование. Авторы считают это анахронизмом (заметим в скобках, что теория алгоритмов и программирование — это не одно и то же, а также что многие курсы, читаемые под названиями «теория алгоритмов», «дискретная математика», «кибернетика» и т. п. в технических вузах, могут лишь дискредитировать любую из перечисленных дисциплин). Тем не менее такова реальность, и мы старались свести к минимуму требования к подготовке читателя в области теории алгоритмов, считая, однако, что слово «привыкнуть» является одним из толкований слова «понять» и что современный читатель (в отличие, скажем, от читателей 50-х годов) достаточно привык к понятию алгоритма и тем самым не нуждается в разъяснении этого понятия. От читателя, в частности, ожидается понимание того, что среди функций бывают *вычислимые* (т. е. обладающие алгоритмами, вычисляющими значение по аргументу). Тем самым и некоторые последовательности из Ω оказываются вычислимыми (т. е. обладающими алгоритмами, вычисляющими n-й член последовательности по заданному n); отметим сразу же, что такие последовательности окажутся неслучайными по любому из приводимых ниже определений случайной последовательности результатов бросания монеты.

Возможности приложения теории алгоритмов к теории вероятностей, о которых рассказывается в настоящей статье, представляются еще одним аргументом в пользу включения основ теории алгоритмов в обязательное университетское образование.

Проблемы обоснования теории вероятностей в течение длительного времени вызывают пристальный интерес математиков и философов, им посвящена обширная литература. Это побудило нас включить в статью дополнение, в котором обсуждаются различные взгляды на основания теории вероятностей с точки зрения излагаемого в статье подхода.

ГЛАВА 1
ОСНОВНЫЕ ПОНЯТИЯ И ФАКТЫ

§ 1.1. Понятие случайности зависит от заданного распределения вероятностей

Прежде чем приступать к какой-либо попытке дать строгое математическое определение понятию «случайная последовательность», необходимо обсудить интуитивные представления об этом понятии — с тем, чтобы полученное определение этим представлениям соответствовало (а то может случиться, что мы сформулируем определение хотя и внешне безупречное, но другого понятия). Вот очевидное, но тем не менее очень важное предварительное замечание.

Пусть в рассматриваемой последовательности из Ω количество нулей примерно вдвое больше количества единиц. Может ли она быть случайной? Нет, если монета симметрична (появления нуля и единицы равновероятны): да, если монета устроена так, что одна ее сторона выпадает с вероятностью 2/3, а другая — с вероятностью 1/3. Мы видим, что само понятие случайности существенным образом зависит от исходного распределения вероятностей. До сих пор мы рассматривали так называемую *бернуллиевскую* ситуацию, или схему испытаний Бернулли. В ней последовательность образуется при независимых испытаниях, в каждом из которых вероятности появления 0 и 1 равны p и q ($p + q = 1$). Усложнением бернуллиевской ситуации является *марковская*: в ней вероятность появления 0 и 1 на n-ом месте зависит от уже появившихся перед этим знаков; ясно, что здесь понятие случайности — свое. Возможны и более сложные ситуации, и в каждой из них возникает свое понятие случайности.

Таким образом, понятие случайности может иметь смысл только по отношению к некоторому исходному распределению вероятностей на пространстве Ω. Среди распределений выделяются *бернуллиевы*, полностью характеризуемые своими p и q, а среди них — *равномерное бернуллиево*, у которого $p = q = 1/2$. В этой первой главе мы, для наглядности, будем заниматься почти исключительно равномерным бернуллиевым распределением, поскольку уже в этом случае достаточно выпукло видна вся возможная теория. Забегая вперед, отметим, что как сложностный, так и количественный подходы (в отличие от частотного) легко обобщаются и на более общий случай произвольного вычислимого распределения на Ω (определение см. в главе 2). Частотный же подход существенно использует то, что распределение вероятностей — бернуллиево (зато не требует вычислимости). Отметим также работу Дэвида [Дэв 85], где предпринимаются попытки дать определение случайности для небернуллиевых распределений в рамках частотного подхода.

§ 1.2. Три свойства случайности: стохастичность, хаотичность и типичность

Случайные бесконечные последовательности (если, конечно, таковые существуют) обладают рядом свойств. Можно надеяться, что изучение этих свойств поможет нам в поисках формального определения случайной последовательности. Пока такого определения нет, само наличие указанных свойств не может быть доказано, а может лишь быть постулировано как отвечающее нашей интуиции (поддерживаемой эмпирическими наблюдениями и умозрительными спекуляциями).

Простейшее из таких свойств — свойство *устойчивости частот*, а короче — просто *устойчивости*. Для равномерного бернуллиева распределения (а только такие мы изучаем в этой главе) оно заключается в том, что отношение количества единиц в начальном отрезке последовательности к длине этого отрезка стремится к 1/2 при неограниченном возрастании длины начального отрезка. Разумеется, свойством устойчивости частот обладают и многие неслучайные последовательности, например, последовательность

0101010101010....

Однако в этом случае подпоследовательность, составленная из членов с четными номерами, уже не обладает свойством устойчивости. В случайной же последовательности устойчивость выполняется не только для последовательности в целом, но и для многих ее подпоследовательностей (например, для подпоследовательности, образованной членами с четными номерами или членами, идущими вслед за единицами). Очевидным образом невозможно, чтобы устойчивостью обладали все подпоследовательности данной последовательности: ясно, что если мы отберем в точности те члены последовательности, которые равны 0 (или 1), полученная подпоследовательность из одних

нулей (единиц) не будет обладать (применительно к равномерному бернуллиеву распределению) свойством устойчивости частот! Можно пытаться как-то ограничить разрешенные способы выбора подпоследовательности и называть стохастичностью свойство последовательности, состоящее в том, что все ее «законно выбранные» подпоследовательности обладают устойчивостью частот. Интуитивно свойство стохастичности можно описать как отсутствие стратегии, позволяющей систематически выигрывать при игре в орлянку (при этом членам, не включенным в подпоследовательность, соответствуют бросания монеты, при которых не было сделано ставки). Кажется очевидным, что для случайной (по равномерной бернуллиевой мере) последовательности такой стратегии быть не должно, т. е. что всякая случайная последовательность обладает свойством стохастичности.

Частотный подход фон Мизеса как раз и состоит в отождествлении этого свойства со свойством случайности. Понятие «допустимого правила выбора» требует, разумеется, разъяснений, которые могут быть сделаны различными неэквивалентными способами. Обсуждение этих способов и недостатков частотного подхода см. в главе 6.

Другое свойство случайных последовательностей состоит в их хаотичности. *Хаотичность* означает, что случайная последовательность беспорядочна, что в ней не усматривается никакой простой организации (именно поэтому случайную последовательность так трудно указать конкретно), что она сложно устроена, что ее нельзя задать проще, чем выписав все ее члены, и т. п. Сложностной подход Колмогорова состоит в отождествлении хаотичности (= сложноустроенности) со случайностью. Разумеется, понятие «сложно устроенная» нуждается в уточнении.

Третье свойство случайных последовательностей (на нем основан количественный подход Мартин-Лёфа к определению случайности) состоит в их *типичности*: случайная последовательность является «типичным представителем» класса всех последовательностей. Что имеют в виду, говоря, что некто является, к примеру, «типичным представителем мелкопоместного дворянства»? По-видимому, это означает, что данное лицо является мелкопоместным дворянином и не имеет никаких особенностей, выделяющих его среди общей массы мелкопоместных дворян. Другими словами, если мы выделим по какому-то признаку группу мелкопоместных дворян, составляющую малую часть от их общего количества, то данное лицо в нее не попадет. Точно так же свойство типичности бесконечной последовательности ω состоит в том, что если мы выделим какое-то подмножество пространства Ω, составляющее малую часть всего пространства Ω, то последовательность ω не попадает в выбранное подмножество. Это описание легко довести до абсурда: для любой последовательности ω множество $\{\omega\}$, состоящее из единственной последовательности ω, составляет предельно малую часть пространства Ω, и если мы потребуем, чтобы типичная последовательность не входила ни в одно такое множество, то не окажется ни одной типичной последовательности. (Любого мелкопоместного дворянина можно объявить нетипичным, отметив, что лица, имеющие те же имя, фамилию, день и место рождения, составляют ничтожно малую часть всех мелкопоместных дворян.) Чтобы этого не было, нужно рассматривать не все малые подмножества Ω, а только некоторые. Разумный класс таких подмножеств был предложен Мартин-Лёфом; обсуждение его свойств см. в параграфах 3 и 4 этой главы.

Разумеется, можно указать много других свойств индивидуальных случайных последовательностей (например, «случайная последовательность не может быть вычислима») или класса всех случайных последовательностей (например, «случайная последовательность остается случайной при вычислимой перестановке: если $a_n = b_{f(n)}$, где f — вычислимое взаимно-однозначное отображение множества натуральных чисел $\mathbb{N}$ на себя, а b — случайная последовательность, то последовательность a случайна». Однако важно выделить «фундаментальные», «базовые» свойства случайности, след-

ствиями которых являются все остальные. Пожалуй, к числу таких свойств относятся типичность и хаотичность (которые в некотором смысле эквивалентны, см. формулировку теоремы Левина — Шнорра в § 4 этой главы и главу 4). Другие естественные свойства (в частности, стохастичность) являются их следствиями. Можно сказать, что (по крайней мере на сегодняшнем уровне понимания) сущность случайности состоит в типичности и хаотичности, а все другие свойства (в частности, многочисленные варианты стохастичности) являются внешними проявлениями этой сущности.

§ 1.3. Последовательности типические, хаотические и стохастические: пути уточнений

Последовательности, обладающие свойством типичности — соответственно хаотичности, стохастичности — будем называть *типическими* — соответственно *хаотическими*, *стохастическими*. Наша цель — найти математические определения для пока еще расплывчатых понятий типичности, хаотичности, стохастичности. Когда и если эта цель будет достигнута, соответствующие точные понятия можно будет рассматривать как уточнения для понятия случайности (или некоторых аспектов этого понятия). В данном параграфе мы наметим пути достижения названной цели.

1.3.1. Типичность. Мы уже упоминали о вольности речи, использующей оборот «случайная последовательность». Когда говорят, например, что «согласно закону больших чисел в случайной последовательности нулей и единиц предел частоты единиц равен 1/2», имеют в виду, что множество последовательностей, для которых предел частоты равен 1/2, имеет полную меру (т. е. мера его дополнения — множества тех последовательностей, где этот предел не существует или не равен 1/2, равна 0; говоря о мере, мы имеем в виду равномерное бернуллиево распределение вероятностей на Ω). Естественно стремиться дать определение случайной последовательности так, чтобы этот оборот речи (и подобные ему обороты для других свойств) можно было бы понимать буквально.

Уточним сказанное. Пусть имеется некоторое вероятностное пространство X с распределением μ. Множества, имеющие меру 0, будем называть *нулевыми*, или *пренебрежимыми*. Мы хотели бы — если бы это было возможно — определить понятие случайного элемента пространства X так, чтобы для всякого свойства P, которым могут обладать или не обладать элементы X, следующие утверждения (1) и (2) были бы равносильны:

(1) почти все элементы X обладают свойством P (это значит, что множество элементов $x \in X$, не обладающих свойством P, пренебрежимо);

(2) все случайные (относительно распределения μ) элементы пространства X обладают свойством P.

Если нам удалось бы этого достичь, то описанная выше вольность речи стала оправданной (в приведенном примере $X = \Omega$, μ — равномерное бернуллиево распределение, а P — свойство последовательности иметь предел частот, равный 1/2). Посмотрим, что для этого требуется. Предположим, что нам удалось каким-то образом ввести понятие случайного элемента X, причем (1) и (2) равносильны для любого свойства P. Взяв в качестве P свойство «быть случайным», убеждаемся, что множество неслучайных объектов пренебрежимо. С другой стороны, если U — любое пренебрежимое множество, то свойство «не принадлежать U» выполнено для почти всех элементов X, и потому должно быть выполнено для всех случайных объектов. Тем самым U должно содержаться в множестве неслучайных объектов, так что множество неслучайных объектов должно быть максимальным по включению пренебрежимым множеством.

Отсюда следует, что наш замысел — во всех нетривиальных случаях — неосуществим: максимального по включению пренебрежимого множества, как правило, не существует — хотя бы потому, что одноэлементные под-

множества X пренебрежимы. (Мы по существу повторили — в более точных терминах — аргументацию, приведенную в предыдущем пункте по поводу необходимости ограничить класс допустимых подмножеств Ω при определении типичности.)

К счастью, как обнаружил П. Мартин-Лёф ([Мар 66], [Мар 66а]), оказывается, что во многих случаях среди нулевых (= пренебрежимых) множеств можно выделить подкласс эффективно нулевых множеств таким образом, что: (1) нулевые множества, появляющиеся в теории вероятностей, как правило, оказываются эффективно нулевыми; (2) объединение всех эффективно нулевых множеств — эффективно нулевое (и, следовательно, оно является максимальным по включению эффективно нулевым множеством). Понятие эффективно нулевого множества спасает нашу программу, и мы можем определить случайный элемент как элемент, не лежащий ни в каком эффективно нулевом множестве (или, что то же самое, не лежащий в максимальном эффективно нулевом множестве). Можно сказать также, что эффективно нулевые множества — это «эффективные тесты на случайность» и что случайным мы называем объект, выдерживающий все такие тесты.

При таком определении случайности для любого свойства P, которым могут обладать или не обладать элементы пространства X, следующие утверждения будут равносильны:

(1) множество тех элементов, для которых не выполнено свойство P, эффективно нулевое;

(2) свойство P выполнено для всех случайных элементов.

(Это происходит благодаря тому, что — как говорилось выше — объединение всех эффективно нулевых множеств является эффективно нулевым.)

Точное определение эффективно нулевого множества и формулировки соответствующих утверждений (для равномерного бернуллиева распределения) будут даны ниже в этой же главе (§ 1.4); подробному изложению количественного подхода к случайности посвящена глава 2. Как видно из сказанного, количественный подход близок к обычной практике теории вероятностей. Замечательно, что он оказывается равнообъемным с другим подходом — сложностным, отождествляющим сложность с хаотичностью.

1.3.2. Хаотичность. Стремясь понять, что означает, что последовательность сложно устроена, А. Н. Колмогоров избирает такой путь. Сперва определяется некая выражаемая числом характеристика сложности двоичного слова. Затем (бесконечная) последовательность нулей и единиц объявляется случайной, если сложности ее начальных отрезков растут так быстро, как только возможно. (Хаотичность в описанном смысле соответствует случайности по равномерному бернуллиеву распределению; сложностной подход, однако, применим и к другим распределениям, см. главу 3.)

Чем отличаются простые объекты от сложных? Почему две последовательности нулей и единиц одинаковой длины могут быть различными по сложности? Почему последовательность из 1000 нулей кажется проще «случайно выбранной» последовательности длины 1000? Предлагаемый ответ таков: простые объекты — это те, которые могут быть коротко описаны. Например, слова «тысяча нулей» можно рассматривать как описание последовательности из 1000 нулей — описание гораздо более короткое, чем запись из 1000 подряд идущих нулей. Поэтому эта последовательность проста. А «случайно выбранная» последовательность длины 1000 такого простого описания не имеет, поэтому она сложна. Мы будем рассматривать в качестве описаний не тексты на русском языке, а двоичные слова (конечные последовательности нулей и единиц). Размером описания будем считать его длину. *Сложностью объекта* (объектами также будут двоичные слова) будем называть минимальный размер его описания — надо лишь уточнить, что такое «описание».

Ясно, что возможных «способов описания» может быть много, поэтому выражение «x есть описание y» лишено смысла, если не дополнить его сло-

вами «...при данном способе описания». Мы отложим точные определение понятия «описания» до § 1.4, сказав лишь, что это определение использует теорию алгоритмов.

Возвращаясь к сложности, определенной как длина самого короткого описания, мы видим, что сложность данного слова зависит от выбора способа описания. К счастью (в этом и состоит в первую очередь открытие Колмогорова), среди способов описания можно выделить так называемые *оптимальные*. Их также много, но соответствующие им сложности отличаются сравнительно мало (не более чем на константу). Сложность при (каком-нибудь) оптимальном способе описания называется *энтропией*. После этого хаотической объявляется последовательность, у которой энтропии начальных отрезков растут с максимально возможной быстротой.

Вот, вкратце и в модернизированной терминологии, путь, намеченный А. Н. Колмогоровым в уточнение естественной идеи о том, что «случайность есть отсутствие закономерностей». (См. подробнее в § 1.4 и, далее, в главе 3.)

1.3.3. Стохастичность. Как было сказано выше в § 1.2, свойство стохастичности состоит в устойчивости частот в данной (бесконечной) последовательности и в ее подпоследовательностях, полученных «законным выбором». Краткий исторический очерк возникновения различных понятий «законного выбора» — начиная с работ Р. фон Мизеса — будет дан в последней главе. Сейчас мы ограничимся лишь одним примером, показывающим, какие следствия можно вывести из устойчивости частот в законно выбранных подпоследовательностях, считая некоторые правила выбора законными.

Итак, мы предполагаем, что в стохастической последовательности частота единиц, определенная по формуле p_N = (число единиц среди первых N членов)/N, стремится к 1/2 при $N \to \infty$. (Напомним, что в этой главе рассматривается равномерное бернуллиево распределение на Ω.) Более того, мы предположили, что это свойство справедливо не только для самой последовательности, но и для ее «законных подпоследовательностей». Будем считать, что к числу законных относятся подпоследовательности, получаемые следующим способом. Фиксируем произвольное двоичное слово X. Рассмотрим теперь подпоследовательность, которая состоит из всех членов, непосредственно следующих за словом X. (Пример: при $X = 01$ из последовательности 0010111010110... будут выбраны подчеркнутые члены, а при $X = 00$ из последовательности 00000... будут выбраны все члены, начиная с третьего.)

Считая только что описанный способ выбора законным, мы тем самым требуем от стохастической последовательности, чтобы любая ее бесконечная подпоследовательность, полученная указанным способом, имела предел частоты единиц, равный 1/2.

Из этих требований вытекает, что частота появления в стохастической последовательности $x_0x_1\ldots$ любого двоичного слова длины k одинакова и равна 2^{-k}. Более точно, пусть S — двоичное слово длины k. Рассмотрим долю тех $i < N$, при которых $x_ix_{i+1}\ldots x_{i+k-1} = S$. Утверждается, что при $N \to \infty$ эта доля стремится к $1/2^k$.

Покажем это. При $k = 1$ это свойство входит в определение стохастичности. Далее можно рассуждать индукцией по k; мы поясним индуктивный шаг примером. Покажем, что группы вида 01 составляют четвертую часть всех групп из двух цифр, встречающихся в последовательности. В самом деле, за каждым нулем (которые составляют, как мы только что видели, половину членов последовательности), идет либо 0, либо 1. При этом нули и единицы встречаются одинаково часто, иначе в подпоследовательности, составленной из членов после нулей, было бы не поровну нулей и единиц. Таким образом, на долю групп 01 приходится одна четверть. После этого, рассмотрев подпоследовательность, составленную из членов, идущих после 01, убеждаемся, что группы 010 и 011 встречаются одинаково часто, составляя — каждая — одну восьмую всех групп длины 3.

Приведенное рассуждение показывает, что все стохастические последовательности нормальны в смысле Бореля. Он называл *нормальной* последовательность, в которой «предельная частота... цифр, равно как и сочетаний из произвольного числа последовательных цифр, следует тому же закону, что и для чисел, цифры которых выбраны случайно, т. е. равна одной десятой для одной цифры, одной сотой — для сочетания двух цифр, одной тысячной — для сочетания трех цифр и т. д.» ([Бор 61, с. 61]; Борель говорит о последовательностях десятичных цифр в разложении вещественного числа, отсюда 10 вместо 2.)

1.3.4. Комментарий. Попытки определить, что такое отдельно взятый случайный объект, были предприняты, в хронологическом порядке, Р. фон Мизесом [Миз 19], А. Н. Колмогоровым [Колм 65] и П. Мартин-Лёфом [Мар 66]. Фон Мизесу принадлежит идея стохастичности: случайность бесконечной последовательности он отождествлял с ее (понимаемой, впрочем, довольно расплывчато) стохастичностью (хотя и не употреблял термина «стохастическая последовательность», а говорил «коллектив» — «Kollektiv»). (Подробнее см. ниже в гл. 6.) А. Н. Колмогорову принадлежит идея хаотичности: случайность конечной последовательности он отождествлял с отсутствием простых описаний этой последовательности, т. е. с тем, что мы называем хаотичностью. А. Н. Колмогоров не употреблял термина «хаотичность», а говорил «случайная последовательность». Перенесение представлений о хаотичности с конечных последовательностей на бесконечные (посредством измерения энтропии начальных отрезков) хотя и было достаточно очевидным для Колмогорова и его окружения, однако вызвало определенные трудности. Трудности были преодолены одновременно и независимо Л. А. Левиным (учеником А. Н. Колмогорова) и К. П. Шнорром, предложившими надлежащее понятие энтропии. См. об этом ниже п. 1.4.2 и главу 3. Идея типичности принадлежит шведскому математику П. Мартин-Лёфу (также ученику А. Н. Колмогорова): случайность бесконечной последовательности он отождествлял с ее типичностью (хотя и не употреблял термин «типическая последовательность», а говорил «случайная последовательность»).

Предложенное Мартин-Лёфом определение типической (в авторской терминологии — случайной) последовательности (мы формулируем это определение ниже в п. 1.4.1 и изучаем в главе 2) было первым в истории науки определением отдельно взятой случайной последовательности, являющимся одновременно математически строгим (в отличие от определения фон Мизеса) и адекватным (в отличие, скажем, от определения Чёрча, приводимого нами в § 6.1). Говоря об адекватности, мы имеем в виду соответствие интуитивному представлению о случайности (определение Чёрча явно приводит к слишком широкому классу последовательностей). Адекватность определения Мартин-Лёфа подтверждается, в частности, тем, что оно оказалось совпадающим по объему с другим определением случайности, в терминах хаотичности (см. п. 1.4.2 и гл. 4).

Существует и другой вариант понятия типичности для бесконечных последовательностей — типичность по Шнорру. Он будет изложен нами в п. 2.3.1. Как и другие авторы, Шнорр не употреблял термина «типическая последовательность». (Термины «типический», «хаотический», «стохастический» в принятом в данной статье смысле впервые прозвучали в широкой аудитории в сентябре 1986 г. в докладе А. Н. Колмогорова и В. А. Успенского на I Всемирном конгрессе Общества математической статистики и теории вероятностей им. Бернулли; см. [Колм Усп 87].) Шнорр же говорил «случайная (zufällig) последовательность» ([Шно 71]; в [Шно 77] он называл эти последовательности «Schnorr random»). Последовательности же типические в смысле 1.4.1 он называл (в [Шно 71]) «гиперслучайными (hyperzufällig)», а в [Шно 77] — «Martin-Löf random». Последовательности, типические в смысле Мартин-Лёфа (т. е. в смысле, удерживаемом в данной статье), образуют собственный подкласс среди всех последовательностей, типических в смысле Шнорра.

Понятие типичности по Шнорру, однако, представляется нам менее адекватно отражающим наши представления о случайности, нежели понятие типичности по Мартин-Лёфу. Тому есть три следующие причины.

1. Любой вариант понятия типичности сопряжен с соответствующим вариантом понятия пренебрежимости. Среди множеств, пренебрежимых по Мартин-Лёфу, существует наибольшее, а среди множеств, пренебрежимых по Шнорру, такового нет (см. утверждение (1) в п. 2.3.1). В то же время, как показывает обсуждение в п. 1.3.1, наличие наибольшего пренебрежимого множества можно считать полезным эффектом.

2. Шнорррово определение типичности лишено поддержки в виде совпадения по объему с определением хаотичности (а определению Мартин-Лёфа такая поддержка дается теоремой Левина — Шнорра, см. § 1.4).

3. Существует последовательность, типическая по Шнорру, но не стохастическая по Колмогорову — Лавлэнду (см. замечание в п. 6.2.1). В то же время от определения случайности разумно ожидать, что всякая случайная последовательность будет стохастической по Колмогорову — Лавлэнду.

Что касается подхода фон Мизеса, то до сих пор не удалось воплотить замысел фон Мизеса в удовлетворительное со всех точек зрения определение случайности. Одна из наиболее известных попыток в этом направлении принадлежит Чёрчу [Чёрч 40], другая — А. Н. Колмогорову [Колм 63] и Д. Лавлэнду [Лав 66а]. Эти попытки описаны ниже в § 6.1. Хотя определение Колмогорова — Лавлэнда приводит к более узкому классу стохастических последовательностей, нежели определение Чёрча, даже и это определение дает более широкий класс, чем определение Мартин-Лёфа (см. п. 6.2.4).

§ 1.4. Типические и хаотические последовательности: основные определения (для равномерного бернуллиева случая)

В этом пункте мы приведем точные определения типической и хаотической последовательности, отложив доказательства (и даже мотивировки) до глав 2 и 3. При этом мы ограничимся простейшим случаем равномерного бернуллиева распределения.

1.4.1. Типичность. Напомним определение пренебрежимого, или нулевого, подмножества множества Ω (относительно равномерного бернуллиева распределения). Через Ω_x (x — двоичное слово) обозначим множество всех (бесконечных) последовательностей, начинающихся на x: $\Omega_x = \{\omega \in \Omega \mid x$ — начало $\omega\}$. Например, если $x = \Lambda$ (пустое слово), то $\Omega_x = \Omega$. Через $l(x)$ обозначим длину слова x. Говорят, что множество $A \subset \Omega$ является *нулевым*, или *пренебрежимым*, если для всякого $\varepsilon > 0$ существует (конечная или бесконечная) последовательность двоичных слов $x_0, x_1, \ldots$, для которой:

$$A \subset \Omega_{x_0} \cup \Omega_{x_1} \cup \ldots, \tag{1}$$

$$2^{-l(x_0)} + 2^{-l(x_1)} + \ldots < \varepsilon. \tag{2}$$

Назвав $2^{-l(x)}$ *мерой* множества Ω_x, а множества вида Ω_x — *интервалами*, можно сказать, что множество A является нулевым тогда и только тогда, когда для всякого $\varepsilon > 0$ существует покрытие множества интервалами с суммой мер меньше ε.

Теперь выделим среди пренебрежимых (нулевых) множеств эффективно пренебрежимые (эффективно нулевые) по Мартин-Лёфу. Определение эффективно пренебрежимого множества отличается от данного нами выше дополнительным требованием: последовательность двоичных слов $x_0, x_1, \ldots$ должна быть вычислимой, и, более того, вычисляющая ее программа должна строиться по ε алгоритмически.

Уточнение этого определения приводит к тому, что рассматриваются не все действительные $\varepsilon > 0$, а, например, лишь рациональные, иначе непонятен смысл выражения «строиться по ε алгоритмически» (неясно, в каком

виде подавать на вход алгоритма действительное число), а с содержательной точки зрения рациональных ε достаточно. Вот что получается при этом уточнении.

Определение эффективно нулевого (эффективно пренебрежимого) множества. Множество $A \subset \Omega$ называется *эффективно нулевым*, если существует вычислимая функция X: $\langle \varepsilon, i\rangle \mapsto X(\varepsilon, i)$, определенная на парах вида ⟨положительное рациональное число, натуральное число⟩, значениями которой являются двоичные слова, причем при любом $\varepsilon > 0$

$$A \subset \Omega_{X(0,\varepsilon)} \cup \Omega_{X(1,\varepsilon)} \cup \ldots, \tag{1}$$

$$2^{-l(X(0,\varepsilon))} + 2^{-l(X(1,\varepsilon))} + \ldots < \varepsilon. \tag{2}$$

При этом не требуется, чтобы функция X была всюду определенной; если $X(\varepsilon, i)$ не определено, то соответствующее слагаемое в выражениях (1) и (2) должно пропускаться.

Эквивалентную формулировку можно дать в терминах перечислимых множеств. Множество называется *перечислимым*, если существует — возможно, никогда не завершающий работу — алгоритм, печатающий один за другим все элементы этого множества и только их; эквивалентное определение: множество перечислимо, если оно есть множество значений вычислимой функции; еще один вариант: множество перечислимо, если оно есть область определения вычислимой функции. Теперь определение эффективно нулевого множества можно сформулировать так: множество A является эффективно нулевым тогда и только тогда, когда существует перечислимое множество W пар вида ⟨положительное рациональное число, двоичное слово⟩, для которого выполнены такие свойства (при произвольном рациональном $\varepsilon > 0$):

$$A \subset \bigcup \{\Omega_x \mid \langle \varepsilon, x\rangle \in W\}, \tag{1}$$

$$\Sigma\, \{2^{-l(x)} \mid \langle \varepsilon, x\rangle \in W\} < \varepsilon. \tag{2}$$

Теорема Мартин-Лёфа [Мар 66а]. *Существует эффективно нулевое множество, содержащее все другие эффективно нулевые множества.*

Доказательство теоремы Мартин-Лёфа будет приведено в § 2.2. Как объяснено выше в § 1.3, эта теорема дает основание для определения типической последовательности как последовательности, не лежащей ни в одном эффективно нулевом множестве. Напомним еще раз, что все сказанное относится (пока) к равномерному бернуллиеву распределению; общий случай будет разобран в главе 2.

1.4.2. Хаотичность. Как мы видели, определение хаотичности надо начать с уточнения понятия способа описания.

Напомним, что через Σ мы обозначаем множество всех конечных и бесконечных последовательностей нулей и единиц. В качестве способов описания мы будем рассматривать вычислимые отображения f: $\Sigma \to \Sigma$. Разумеется, понятие вычислимости в этом случае нуждается в уточнении, так как аргументы и значения отображения f — конечные и бесконечные последовательности нулей и единиц — не являются, вообще говоря, конструктивными объектами. Мы отложим формальное определение до конца этого пункта. В качестве первого приближения читатель может представлять себе машину, «вычисляющую» функцию f так: $f(\omega)$ есть последовательность нулей и единиц, появляющихся на выходе машины, на вход которой подают (последовательно, начиная с первого) символы последовательности ω.

Итак, фиксируем вычислимое отображение f: $\Sigma \to \Sigma$. Будем говорить, что (конечное) двоичное слово y является *описанием* (конечного) двоичного слова x при данном f, если x есть начало (конечной или бесконечной) последовательности $f(y)$. Другими словами, «описать x» значит предъявить такое слово y, при помещении которого на вход машины на ее выходе появятся один за другим символы, составляющие слово x (а затем, возможно, и еще какие-нибудь символы).

О п р е д е л е н и е. *Сложностью* слова x при данном вычислимом отображении f называется число $KM_f(x)$, определяемое формулой

$$KM_f(x) = \min \{l(y) \mid y \text{ есть описание } x \text{ при } f\}.$$

Здесь $l(y)$ — длина двоичного слова y; минимум пустого множества считаем бесконечным.

О п р е д е л е н и е. Вычислимое отображение f называется *оптимальным*, если для любого вычислимого отображения g найдется такая константа C, что

$$KMg(x) \leqslant KMf(x) + C$$

для всех двоичных слов x.

Т е о р е м а К о л м о г о р о в а. *Существуют оптимальные вычислимые отображения.*

О п р е д е л е н и е. *Энтропией* называется сложность при произвольном оптимальном вычислимом отображении. Таким образом, для каждого оптимального отображения — своя энтропия. Но любые две энтропии KM_1 и KM_2 различаются не более чем на константу, т. е. для некоторого числа C и любого двоичного слова x выполнено неравенство

$$|KM_1(x) - KM_2(x)| < C.$$

Поэтому, допуская вольность речи, можно считать, что энтропия одна, но определена с точностью до ограниченного слагаемого.

Энтропия слова x обозначается $KM(x)$.

Для тождественного отображения, очевидно, сложность слова равна его длине. Сравнивая оптимальное вычислимое отображение с тождественным, заключаем по теореме Колмогорова, что энтропия слова превосходит его длину не более чем на константу: $KM(x) \leqslant l(x) + O(1)$. (Через $O(1)$, как всегда, обозначается ограниченное слагаемое.)

Те последовательности, для начальных отрезков которых это неравенство обращается в равенство, и называются хаотическими. Более точно. Обозначим через $(\omega)_n$ начальный отрезок длины n последовательности ω.

О п р е д е л е н и е. Последовательность ω называется *хаотической*, если существует такое C, что

$$|KM((\omega)_n) - n| \leqslant C$$

при всех n.

Т е о р е м а Л е в и н а — Ш н о р р а. *Классы хаотических и типических последовательностей совпадают.*

Доказательства сформулированных утверждений, так же как и ссылки на литературу, будут даны в главах 3, 4. Мы закончим обещанным формальным определением вычислимости отображения f: $\Sigma \to \Sigma$. Отображение f: $\Sigma \to \Sigma$, определенное на всем Σ, называется *вычислимым*, если выполнены такие свойства:

(1) если последовательность x есть начало последовательности x', то последовательность $f(x)$ есть начало последовательности $f(x')$ ($x, x' \in \Sigma$; любой элемент множества Σ мы считаем началом самого себя);

(2) значение отображения f на бесконечной последовательности равно минимальному продолжению значений отображения f на всех ее конечных началах (согласно требованию (1) таковое существует);

(3) множество всех пар конечных слов $\langle x_0, y_0\rangle$, для которых y_0 есть начало $f(x_0)$, перечислимо (это означает, напомним, что оно является множеством значений вычислимой функции).

Мотивировка этого определения (в частности, установление его связи с неформальным описанием вычислимости в терминах машин) будет дана в главе 3.

ГЛАВА 2

ЭФФЕКТИВНО НУЛЕВЫЕ МНОЖЕСТВА, КОНСТРУКТИВНЫЙ НОСИТЕЛЬ МЕРЫ И ТИПИЧЕСКИЕ ПОСЛЕДОВАТЕЛЬНОСТИ

В этой главе мы изложим предложенное П. Мартин-Лёфом [Мар 66а] определение типической последовательности.

§ 2.1. Эффективно нулевые множества, вычислимые распределения и формулировка теоремы Мартин-Лёфа

Как отмечалось в главе 1, разумная постановка вопроса о случайности требует фиксации вероятностного пространства. Мы будем рассматривать пространство Ω бесконечных последовательностей нулей и единиц. Напомним, что для каждого (конечного) двоичного слова x (множество таких слов мы обозначаем Ξ) через Ω_x обозначается множество всех бесконечных продолжений слова x. Множества Ω_x образуют базу топологии на Ω (стандартной топологии пространства Кантора). Тем самым определяется и класс борелевских подмножеств пространства Ω.

Помимо пространства объектов, для задания вероятностного пространства нужно указать распределение вероятностей. В нашем случае в качестве распределений вероятностей мы будем рассматривать такие борелевские меры μ на Ω (счетно-аддитивные меры на борелевских подмножествах пространства Ω), для которых $\mu(\Omega) = 1$. Как известно, такая мера определяется своими значениями $\mu(\Omega_x)$ на всех множествах вида Ω_x при $x \in \Xi$. Более того, если задана произвольная функция $m\colon \Xi \to \mathbb{R}$, удовлетворяющая требованиям

$$(M)\begin{cases} m(\Lambda) = 1 \ (\Lambda \text{ — пустая последовательность, т. е. последовательность длины } 0); \\ m(x0) + m(x1) = m(x) \text{ для всякого } x; \\ m(x) \geqslant 0 \text{ для всякого } x, \end{cases}$$

то существует единственная борелевская мера μ, для которой $\mu(\Omega_x) = m(x)$ при всех $x \in \Xi$. Поэтому можно определять распределение вероятностей на Ω как функцию, обладающую свойством (M).

Итак, пусть на пространстве Ω задано распределение μ. Тогда обычным образом возникает понятие нулевого множества (которое уже не обязано быть борелевским). В терминах функции $m\colon x \mapsto \mu(\Omega_x)$ можно сказать так: множество $A \subset \Omega$ является нулевым (синонимы: имеет меру 0, является пренебрежимым) тогда и только тогда, когда для всякого $\varepsilon > 0$ существует последовательность $x_0, x_1, \ldots$ элементов Ξ, для которой

$$A \subset \Omega_{x_0} \cup \Omega_{x_1} \cup \ldots,$$

$$\Sigma m(x_i) < \varepsilon.$$

Следующие факты классической теории меры хорошо известны; мы приводим соответствующие рассуждения для сравнения с их конструктивными аналогами, которые нам понадобятся в дальнейшем.

(1) Если множества $A_0, A_1, \ldots$ нулевые, то и множество $A_0 \cup A_1 \ldots$ — нулевое.

В самом деле, чтобы найти покрытие множества $\bigcup A_i$ с суммой мер, меньшей ε, найдем покрытия множеств $A_0, A_1, \ldots$ с суммами мер, меньшими соответственно $\varepsilon/2$, $\varepsilon/4$, $\ldots$, и объединим их.

(2) Для всякого нулевого множества A существует нулевое множество B, содержащее A и являющееся D_δ-множеством (т. е. пересечением счетного семейства открытых множеств).

В самом деле, для каждого n выберем покрытие множества A с суммой мер, не превосходящей $1/n$; объединение всех множеств этого покрытия обозначим B_n. Тогда B_n открыто, $A \subset B_n$ и $\mu(B_n) \leqslant 1/n$. Множество $B = \bigcap B_n$ будет искомым.

Дадим теперь определение эффективно нулевого (эффективно пренебрежимого) множества по Мартин-Лёфу в пространстве Ω с распределением μ. Оно отличается от приведенного в § 1.4 лишь тем, что мы перешли от равномерного бернуллиева распределения вероятностей на Ω к произвольному.

О п р е д е л е н и е. Пусть μ — распределение вероятностей на Ω, A — подмножество пространства Ω. Множество A называется *эффективно нулевым*, если существует вычислимая функция X: $\langle \varepsilon, i\rangle \mapsto X(\varepsilon, i\rangle$, определенная на парах вида ⟨положительное рациональное число, натуральное число⟩, значениями которой являются двоичные слова, причем при любом рациональном $\varepsilon > 0$

$$A \subset \Omega_{X(\varepsilon,0)} \cup \Omega_{X(\varepsilon,1)} \cup \ldots, \tag{1}$$

$$\sum_i \mu(\Omega_{X(\varepsilon,i)}) < \varepsilon. \tag{2}$$

При этом не требуется, чтобы функция X была всюду определенной; если $X(\varepsilon, i)$ не определено, то соответствующее слагаемое в выражениях (1) и (2) должно пропускаться.

З а м е ч а н и е. Не ограничивая общности, можно считать, что если $X(\varepsilon, i)$ определено, то и $X(\varepsilon, j)$ определено при всех $j < i$; достаточно перенумеровать члены последовательности $X(\varepsilon, 0)$, $X(\varepsilon, 1)$, ... в порядке обнаружения их определенности. Это иногда бывает полезно. (Именно такой вариант определения использован в [Колм Усп 87] и [Усп Сем 87].)

Приведенное определение имеет смысл для любого распределения μ. Но для разумного определения понятия случайной последовательности на его основе нужно наложить на распределение μ дополнительные ограничения, а именно потребовать его вычислимости. Понятие вычислимости распределения нуждается в специальном определении, так как и аргументы меры (борелевские множества), и ее значения (действительные числа) не являются конструктивными объектами. Поэтому, хотя распределение определено для всех борелевских множеств, при формулировке определения вычислимости следует ограничиться лишь множествами вида Ω_x (которые задаются двоичными словами x) в качестве аргументов и требовать, чтобы значения меры на них можно было бы алгоритмически вычислять с любой заданной точностью. Приходим к такому определению.

Распределение μ называется *вычислимым*, если существует вычислимая функция M: $\langle x, \varepsilon\rangle \mapsto M(x, \varepsilon)$, определенная на всех парах вида ⟨конечная последовательность нулей и единиц, положительное рациональное число⟩ и принимающая рациональные значения, для которой $|M(x, \varepsilon) - \mu(\Omega_x)| \leqslant \varepsilon$ для всех слов x и для всех рациональных $\varepsilon > 0$.

Т е о р е м а М а р т и н-Л ё ф а. *Если распределение μ вычислимо, то существует максимальное по включению эффективно нулевое относительно μ множество (т. е. эффективно нулевое множество, содержащее любое другое эффективно нулевое множество).*

П е р е ф о р м у л и р о в к а: *объединение всех эффективно нулевых множеств есть эффективно нулевое множество.*

Как объяснено в главе 1, эта теорема дает основания для определения типической последовательности.

О п р е д е л е н и е. *Типической* (для данного вычислимого распределения) называется последовательность, не содержащаяся ни в одном эффективно нулевом множестве (= не содержащаяся в максимальном эффективно нулевом множестве). *Конструктивным носителем* вычислимого распределения называется множество всех типических последовательностей.

Пользуясь этим определением, можно сказать, что множество является эффективно пренебрежимым, если все его элементы нетипичны. Парадоксальным образом оказывается, что в отличие от классического определения нулевого множества, требующего, чтобы в множестве было «мало» элементов, наличие или отсутствие свойства «быть эффективно нулевым множеством» определяется лишь тем, типичны ли его элементы.

Пример. Вычислимая последовательность ω является типической тогда и только тогда, когда множество $\{\omega\}$ имеет положительную меру. (Напомним, что мы, говоря о типичности, рассматриваем лишь вычислимые распределения вероятностей.)

В самом деле, если $\mu(\{\omega\}) > 0$, то ω не может содержаться ни каком нулевом и тем более эффективно нулевом множестве. Поэтому ω — типическая последовательность. (Здесь не используется вычислимость последовательности ω; впрочем, можно доказать, что всякая последовательность ω, для которой $\mu(\{\omega\}) > 0$ при вычислимом μ, вычислима.)

Пусть, напротив, $\mu(\{\omega\}) = 0$. Покажем, что множество $\{\omega\}$ — эффективно нулевое (и, следовательно, ω — не типическая). Пусть $(\omega)_n$ — начало последовательности ω, имеющее длину n. Множества $\Omega_{(\omega)_n}$ убывают и в пересечении дают $\{\omega\}$, поэтому их меры стремятся к 0. Легко видеть, что в силу вычислимости последовательности ω и распределения μ эта сходимость эффективна в том смысле, что по любому рациональному $\varepsilon > 0$ можно указать такое начало $x(\varepsilon)$ последовательности ω, что $\mu(\Omega_{x(\varepsilon)}) < \varepsilon$. Теперь в качестве покрытия малой меры можно взять покрытие, состоящее из единственного множества $\Omega_{x(\varepsilon)}$. (Формально: $X(\varepsilon, 0) = x(\varepsilon)$, $X(\varepsilon, k)$ не определено при $k > 0$.) Поэтому $\{\omega\}$ — эффективно нулевое множество, что и требовалось доказать.

Из сказанного и из теоремы Мартин-Лёфа вытекает, что если вычислимая мера такова, что все одноэлементные множества имеют меру 0 (таковы бернуллиевы меры), то множество всех вычислимых последовательностей является эффективно нулевым.

Доказательству теоремы Мартин-Лёфа посвящен § 2 этой главы. А сейчас мы обсудим вопрос о том, насколько понятие случайности по Мартин-Лёфу согласуется с обычным словоупотреблением теории вероятностей. Мы уже говорили, что формулировку «для случайно взятой последовательности выполнено свойство P» принято рассматривать как сокращение для утверждения «множество всех последовательностей, для которых не выполнено свойство P, нулевое». Эту формулировку можно будет понимать буквально, если только нулевое множество последовательностей, не обладающих свойством P, окажется эффективно нулевым. Мы приходим к такому вопросу: можно ли надеяться, что нулевые множества, встречающиеся в теории вероятностей, оказываются эффективно нулевыми?

Рассмотрим для примера закон больших чисел, согласно которому для почти всех (по равномерному бернуллиеву распределению вероятностей) последовательностей $\omega_0\omega_1\ldots$ выполнено равенство $\lim(\omega_0 + \ldots + \omega_{n-1})/n = 1/2$. Другими словами, множество последовательностей, для которых указанный предел не существует или отличен от 1/2, нулевое. Мы хотим убедиться, что оно является эффективно нулевым.

В силу теоремы Мартин-Лёфа, достаточно убедиться, что для любого рационального $\varepsilon > 0$ множество S тех последовательностей $\omega = \omega_0\omega_1\ldots$, у которых

$$|(\omega_0 + \ldots + \omega_{n-1})/n - 1/2| > \varepsilon$$

для бесконечно многих n, является эффективно нулевым. Обозначим через D_n множество тех $\omega \in \Omega$, для которых

$$|(\omega_0 + \ldots + \omega_{n-1})/n - 1/2| > \varepsilon.$$

Представив S в виде $\bigcap\limits_k \bigcup\limits_{n \geqslant k} D_n$, мы видим, что S есть пересечение множеств

$E_k = \bigcup\limits_{n \geqslant k} D_n$. Так как $E_0 \supset E_1 \supset \ldots$, равенство $\mu(S) = 0$, выражающее закон больших чисел, равносильно тому, что $\mu(E_i) \to 0$. (Обычное доказательство закона больших чисел как раз и состоит в получении верхней оценки для $\mu(E_i)$, стремящейся к 0 при $i \to \infty$.) Множества E_i и будут эффективно открытыми множествами малой меры, требуемыми в определении эффективно нулевого множества; надо проверить лишь, что сходимость $\mu(E_i)$ к 0 будет эффективной (т. е. что в ε-N-определении предела можно эффективно найти N по ε). А это легко сделать, проанализировав обычное доказательство закона больших чисел, в котором мы оцениваем $\mu(D_n)$ с помощью формулы Стирлинга или как-нибудь еще и устанавливаем, что (при фиксированном ε) эта мера экспоненциально убывает с ростом n. (Подробнее см. в § 6.2.)

Аналогичные рассуждения можно применить и к другим теоремам теории вероятностей.

§ 2.2. Доказательство теоремы Мартин-Лёфа

Грубо говоря, идея доказательства теоремы Мартин-Лёфа состоит в следующем. Нам надо доказать, что объединение всех эффективно нулевых множеств — эффективно нулевое. Для этого мы хотим воспользоваться эффективным аналогом теоремы о том, что объединение счетного числа нулевых множеств — нулевое. Трудность, однако, состоит в том, что эффективно нулевых множеств слишком много — хотя бы потому, что всякое подмножество эффективно нулевого множества является эффективно нулевым. Мы преодолеем эту трудность так: среди всех эффективно нулевых множеств мы выберем некоторый подкласс (его элементы мы будем называть *GN-множествами*). При этом (1) каждое эффективно нулевое множество будет подмножеством некоторого множества этого подкласса; (2) множества из этого подкласса можно будет «перечислить» (в некотором уточняемом ниже смысле). Благодаря свойству (2) нам удастся доказать, что их объединение будет эффективно нулевым множеством; благодаря свойству (1) это объединение будет содержать все эффективно нулевые множества.

Перейдем к реализации намеченного плана. Напомним, что *G_δ-множеством* называют счетное пересечение открытых подмножеств. Вспоминая, что открытое подмножество Ω — это счетное объединение интервалов (множеств вида Ω_x для конечных слов x), мы видим, что G_δ-множества — это множества, представимые в виде

$$\bigcap_i \bigcup_j \Omega_{x(i,j)},$$

где $x(i, j)$ — при любых натуральных i, j — двоичное слово. Потребовав вычислимости функции x, мы могли бы получить определение эффективного аналога понятия G_δ-множества. Нас, однако, интересуют не все такие множества, а лишь «эффективно нулевые эффективно G_δ-множества», и даже не все они, а лишь множества, которые мы будем называть *GN-множествами*. Это множества, которые задаются вычислимыми функциями x: $\langle i, j\rangle \mapsto x(i, j)$, удовлетворяющими такому требованию:

(*GN*) для всякого i и для всякого n выполнено неравенство

$$\sum_{j<n} \mu(\Omega_{x(i,j)}) < 2^i.$$

(Как и раньше, мы не требуем, чтобы $x(i, j)$ было определено всегда; если $x(i, j)$ не определено, то соответствующее множество $\Omega_{x(i,j)}$ считаем пустым.) Требование (*GN*) отражает ту идею, что мера множеств $\bigcup\limits_j \Omega_{x(i,j)}$, пересечением которых является наше множество, должна убывать; мы сформулировали наше требование в виде условия на конечные суммы по техническим (но важным) причинам, которые станут ясны впоследствии (грубо говоря, нам нужно проверять соблюдение этого требования алгоритмич ески).

Итак, всякое множество вида $\bigcap\limits_i \bigcup\limits_j \Omega_{x(i,j)}$, где x — вычислимая (частичная) функция, удовлетворяющая требованию (*GN*), будем называть *GN-множеством*, а функцию x — *задающей* его *функцией*. В терминах перечислимых множеств определение *GN*-множества можно переформулировать так: *GN*-множествами являются множества вида $\bigcap\limits_i \bigcup\limits_x \{\Omega_x \mid \langle i, x\rangle \in W\}$, где W — перечислимое множество пар вида ⟨натуральное число, двоичное слово⟩, для которого $\mu(\Omega_{x_1}) + \ldots + \mu(\Omega_{x_n}) < 2^{-i}$ для любого i и любых $x_1, \ldots, x_n$, для которых $\langle i, x_1\rangle, \ldots, \langle i, x_n\rangle \in W$.

Лемма 1. *Всякое GN-множество является эффективно нулевым. Всякое эффективно нулевое множество содержится в некотором GN-множестве.*

Доказательство. Утверждение леммы непосредственно следует из определений; нужно заметить лишь, что в определении эффективно нулевого множества можно рассматривать лишь значения ε вида $1/2^k$ и что хотя при переходе к пределу по n строгое неравенство в (*GN*) заменяется нестрогим, это не страшно (надо уменьшить ε).

Следующая лемма утверждает, что все *GN*-множества могут быть эффективно перечислены.

Лемма 2. *Существует такая* (*частичная*) *функция H трех натуральных аргументов со значениями в Ξ* (*множестве всех двоичных слов*), *что для каждого k функция $\langle i, n\rangle \mapsto H(k, i, n)$ задает GN-множество и, кроме того, всякое GN-множество может быть получено таким образом при подходящем k.*

Доказательство. Мы укажем некоторый способ, который эффективно преобразует каждую вычислимую функцию x двух числовых аргументов со значениями в Ξ в некоторую другую функцию y (также из $\mathbb{N} \times \mathbb{N}$ в Ξ). Этот способ будет таким, что:

(а) любая функция y, полученная этим преобразованием, удовлетворяет требованию (*GN*) и, следовательно, задает некоторое *GN*-множество;

(б) если исходная функция x удовлетворяла требованию (*GN*), то полученная в результате преобразования функция y совпадает с x.

Говоря об эффективности преобразования, мы имеем в виду, что существует алгоритм, который по любой программе любой функции x дает программу некоторой функции y, обладающей указанными свойствами. (Однако разные программы для одной и той же функции после преобразования могут стать программами, вычисляющими разные функции; это возможно, если исходная функция не удовлетворяет требованию (*GN*).)

Если мы установили возможность такого преобразования, то построить H легко: надо применять это преобразование ко всем программам подряд. Точнее, чтобы найти $H(k, i, n)$, нужно взять k-ю программу, подвергнуть ее преобразованию и после этого применить преобразованную программу к $\langle i, n\rangle$.

Итак, осталось описать преобразование. Мы будем использовать такое замечание: если два действительных числа α и β заданы алгоритмами, вычисляющими их приближения с любой заданной точностью, и $\alpha < \beta$, то это обстоятельство рано или поздно обнаружится: надо только вычислять α и β все точнее и точнее и дождаться момента, когда разница между приближениями станет много больше погрешностей.

Пусть x — функция, которую нужно преобразовать. Опишем, как вычисляются значения результирующей функции y. (Эта функция будет сужением функции x на некоторое подмножество ее области определения.) Развернем параллельно вычисления $x(i, 0)$, $x(i, 1)$, ...; время от времени некоторые их этих вычислений будут заканчиваться. При появлении каждого нового значения $x(i, n)$ мы будем проверять, что оно не нарушает нужного нам неравенства, т. е. что сумма мер $\mu(\Omega_{x(i,k)})$ для всех тех k, для которых $x(i, k)$ уже определилось, меньше 2^{-i}. Такая проверка

представляет собой процесс, в ходе которого мы вычисляем значения меры μ со все увеличивающейся точностью. Этот процесс может не кончиться, если вместо проверяемого неравенства выполняется равенство, однако, если неравенство на самом деле выполнено, то рано или поздно мы об этом узнаем. На время проверки работа по вычислению значений $x(i, 0)$, $x(i, 1)$, ... приостанавливается, и если проверка никогда не кончится, то функция y будет иметь конечную область определения. То же случится, если проверка установит, что неравенство не соблюдается — в этом случае все вычисления прекращаются. Если же проверка даст положительный результат, то мы объявляем, что значения функции y на тех парах $\langle i, k\rangle$, для которых $x(i, k)$ уже вычислилось, определены и равны соответствующим значениям функции x.

Ясно, что построенная таким образом функция y всегда обладает свойством (GN); если же исходная функция x обладала этим свойством, то $y = x$. Доказательство леммы 2 закончено.

Теперь мы можем доказать, что объединение всех GN-множеств есть эффективно нулевое множество. Пусть H: $\langle k, i, n\rangle \mapsto H(k, i, n)$ — функция из леммы 2; $X_0, X_1, \ldots$ — GN-множества, которые получаются из нее при $k = 0, 1, \ldots$ Чтобы покрыть X системой интервалов с суммой мер менее 2^{-i}, покроем X_0 с мерой $2^{-i}/2$, X_1 — с мерой $2^{-i}/4$ и т. д., а затем объединим эти покрытия в одно.

Формально говоря, рассмотрим функцию z, для которой $z(i, n) = H(l(n), i + l(n) + 1, r(n))$, где через $n \mapsto \langle l(n), r(n)\rangle$ обозначено взаимно-однозначное соответствие между $\mathbb{N}$ и $\mathbb{N} \times \mathbb{N}$. Эта функция задает GN-множество, содержащее все GN-множества. Применяя лемму 1, получаем теорему Мартин-Лёфа.

§ 2.3. Различные варианты определения типичности

2.3.1. Определение типичности по Шнорру. К. П. Шнорр (см. [Шно 71], [Шно 77]) предложил модифицировать определение Мартин-Лёфа, усилив требования к эффективно нулевым множествам. Назовем множество $A \subset \Omega$ *эффективно нулевым по Шнорру*, если существует вычислимая функция X: $\langle \varepsilon, i\rangle \mapsto X(\varepsilon, i)$, определенная для всех рациональных $\varepsilon > 0$ и всех натуральных i, значениями которой являются двоичные слова, а также вычислимая функция N: $\langle \varepsilon, \delta\rangle \mapsto N(\varepsilon, \delta)$, определенная для всех рациональных $\varepsilon, \delta > 0$, значениями которой являются натуральные числа, для которых

(1) $$A \subset \bigcup_i \Omega_{X(\varepsilon, i)} \text{ при любом } \varepsilon > 0;$$

(2) $$\sum_i \mu(\Omega_{X(\varepsilon, i)}) < \varepsilon \text{ при любом } \varepsilon > 0;$$

(3) $$\sum_{i > N(\varepsilon, \delta)} \mu(\Omega_{X(\varepsilon, i)}) < \delta \text{ при любых } \varepsilon, \delta > 0.$$

Отличие от исходного определения Мартин-Лёфа состоит в том, что мы требуем «вычислимой сходимости» ряда $\sum_i \mu(\Omega_{X(\varepsilon, i)})$: мы не только должны быть уверены, что его сумма не превосходит ε, но и должны уметь эффективно указать по любому $\delta > 0$ частичную сумму ряда, отличающуюся от его суммы не более чем на δ (это делает функция N). Отметим также, что функции X и N должны быть всюду определенными (в исходном определении Мартин-Лёфа функция X могла быть частичной). При таком определении число $S(\varepsilon) = \sum_i \mu(\Omega_{X(\varepsilon, i)})$ является (при любом рациональном $\varepsilon > 0$) вычислимым действительным числом, причем эта вычислимость, как говорят, равномерна по ε (это значит, что программа, вычисляющая рациональные приближения к этому числу, может быть получена по ε алгоритмически).

Можно проверить, что добавив требование равномерной по ε вычислимости числа $S(\varepsilon) = \sum_i \mu(\Omega_{X(\varepsilon, i)})$ к определению эффективно нулевого множества из § 2.1, мы получим определение, равносильное данному только что определению эффективно нулевого по Шнорру множества.

Замечание. Здесь мы (не меняя объема определяемого понятия) несколько модифицировали оригинальное определение эффективно нулевого множества по Шнорру (total rekursive Nullmenge согласно определению 8.1 из [Шно 71]) в сторону упрощения и приближения к схеме определения эффективно нулевого множества из нашего § 2.1 (rekursive Nullmenge согласно определению 4.1 из [Шно 71]). На самом деле Шнорр рассматривает не сумму мер $S(\varepsilon)$, а меру суммы $s(\varepsilon) = \mu(\bigcup_i \Omega_{X(\varepsilon, i)})$ и вместо соблюдения условий (2) и (3) требует, чтобы (2′) $s(\varepsilon) < \varepsilon$ и (3′) $s(\varepsilon)$ было, причем равномерно по ε, вычислимым действительным числом.

Предложенная Шнорром модификация определения эффективно нулевого множества приводит к тому, что (1) *для равномерного бернуллиева распределения не существует наибольшего по включению эффективно нулевого по Шнорру множества* (и это же верно для всех обычно используемых распределений). Кроме того, назвав *типической по Шнорру* (относительно данного распределения) последовательность, не содержащуюся ни в одном эффективно нулевом по Шнорру (относительного этого распределения) множестве, мы получим более широкий класс типических последовательностей, чем у Мартин-Лёфа: (2) *существуют типические по Шнорру относительно равномерного бернуллиева распределения последовательности, не являющиеся типическими (в смысле определения Мартин-Лёфа)*.

Дадим наброски доказательства утверждений (1) и (2). (Так как утверждения этого пункта нигде далее не используются, то мы не приводим подробных доказательств.) Для начала установим, что для любого эффективно нулевого по Шнорру множества A существует вычислимая последовательность, не принадлежащая A. (Поскольку одноэлементное множество, состоящее из вычислимой последовательности, является эффективно нулевым по Шнорру, отсюда будет следовать, что не существует максимального по включению эффективно нулевого по Шнорру множества.)

При построении вычислимой последовательности, не принадлежащей множеству A, будет использовано существование покрытия множества A с вычислимой суммой мер, не превосходящей ε, всего лишь для одного значения $\varepsilon < 1$. Пусть ε $(0 < \varepsilon < 1)$ — рациональное число, $x(0)$, $x(1)$, ... — вычислимая последовательность двоичных слов, причем $\sum_i \mu(\Omega_{x(i)})$ — вычислимо сходящийся ряд с суммой не больше ε. Построим вычислимую последовательность, не принадлежащую множеству $U = \bigcup_i \Omega_{x(i)}$. Фиксируем рациональное число ε', для которого $\varepsilon < \varepsilon' < 1$. Назовем двоичное слово x *хорошим*, если доля элементов множества U среди всех продолжений слова x меньше ε', т. е. если

$$(*) \qquad \mu(U \cap \Omega_x) < \varepsilon' \cdot \mu(\Omega_x).$$

В силу вычислимой сходимости ряда левая и правая части неравенства (*) могут быть вычислены с любой точностью; вычисляя их со все возрастающей точностью, мы рано или поздно обнаружим любое хорошее x (таким образом, множество всех хороших слов перечислимо). По предположению, пустое слово является хорошим; кроме того, для всякого хорошего слова x одно из его продолжений $x0$ и $x1$ является хорошим. Поэтому мы можем найти вычислимую последовательность хороших слов, в которой каждое следующее слово является продолжением предыдущего. Бесконечная последовательность, началами которой эти слова являются, не будет принадлежать U, что и требовалось.

Скажем коротко, как можно построить типическую по Шнорру последовательность (относительно равномерного распределения), не являющуюся типической (в принятом нами смысле, т. е. по Мартин-Лёфу). При этом построении мы будем пользоваться совпадением (оно будет доказано в § 4.1) классов типических и хаотических последовательностей и построим типическую по Шнорру последовательность, имеющую малую энтропию начальных отрезков.

Представим себе, что мы предпринимаем изложенное выше построение вычислимой последовательности, не принадлежащей эффективно нулевому по Шнорру множеству A, для чего выбрали покрытие с вычислимой суммой мер, меньшей ε' (объединение интервалов покрытия обозначим через U) и строим удлиняющиеся хорошие слова. Пусть в некоторый момент нам захотелось, чтобы строимая последовательность не принадлежала также другому эффективно нулевому по Шнорру множеству B. К этому моменту у нас есть хорошее слово x, т. е. слово, для которого $\mu(U \cap \Omega_x) < \varepsilon' \cdot \mu(\Omega_x)$. Если V — объединение интервалов покрытия множества B с достаточно малой вычислимой суммой мер, то слово x будет хорошим и по отношению к $U \cup V$ (т. е. $\mu((U \cup V) \cap \Omega_x)$ будет меньше $\varepsilon' \mu(\Omega_x)$). Выбрав такое V, мы можем продолжать построение хороших слов — теперь уже по отношению к $U \cup V$ — и получить вычислимую последовательность, не принадлежащую $A \cup B$. Если же в некоторый момент нам захочется, чтобы строимая последовательность не принадлежала еще какому-нибудь эффективно нулевому по Шнорру множеству C, то можно найти покрытие W множества C настолько малой меры, что текущее хорошее слово останется хорошим при добавлении W и т. д.

Может показаться, что таким способом можно построить вычислимую последовательность, не принадлежащую никакому эффективно нулевому по Шнорру множеству (рассматривая по очереди все пары вычислимых функций X и N, удовлетворяющих требованиям (2) и (3) из определения эффективно нулевого по Шнорру множества). Однако это не так: вычислимой последовательности не получится, так как для построения необходима дополнительная информация о том, какие пары вычислимых функций обладают свойствами (2) и (3). Однако, если вводить в рассмотрение новую пару вычислимых функций X, N на достаточно далеком этапе конструкции, когда построенная часть последовательности имеет большую длину, можно добиться, чтобы объем этой дополнительной информации был мал по сравнению с длиной последовательности. В этом случае энтропия начал построенной типической по Шнорру последовательности будет мала по сравнению с их длиной, т. е. эта последовательность не будет хаотической.

Проведенные рассуждения являются наброском доказательства такого утверждения: существует типическая по Шнорру (относительно равномерного бернуллиева распределения) последовательность, для которой энтропия начального отрезка произвольной длины n не превосходит $C \cdot \log_2 n$, где C — константа, не зависящая от n.

2.3.2. Критерий типичности Соловея. Другая модификация определения Мартин-Лёфа приведена в [Чэй 87], [Чэй 87а] и названа там случайностью по Соловею (R. M. Solovay). (В этих работах говорится о случайных действительных числах, но все сказанное почти без изменений переносится на бесконечные последовательности нулей и единиц.)

Бесконечная последовательность ω называется *случайной по Соловею* относительно вычислимого распределения вероятностей μ, если не существует вычислимой (не обязательно всюду определенной) функции X, аргументами которой являются натуральные числа, а значениями — двоичные слова, для которой $\sigma = \sum_i \mu(\Omega_{X(i)}) < \infty$ и $\omega \in \Omega_{X(i)}$ при бесконечно многих i.

Очевидно, любая последовательность ω, не являющаяся типической, не случайна по Соловею: взяв покрытия множества $\{\omega\}$ с мерами меньше 1,

1/2, 1/4, ... и объединив их, получим покрытие общей меры не больше 2, для которого ω содержится в бесконечном числе множеств покрытия. Верно и обратное, как показывает следующее простое рассуждение, приводимое в [Чэй 87а], с. 36, со ссылкой на Соловея. Пусть ω принадлежит $\Omega_{X(i)}$ при бесконечно многих i. Определим U_n как множество тех последовательностей, которые принадлежат $\Omega_{X(i)}$ для n или более различных i. Тогда $\omega \in U_n$ при любом n. Легко понять, что U_n можно представить в виде объединения вычислимой последовательности непересекающихся интервалов, а мера U_n не превосходит σ/n. Тем самым можно эффективно указать покрытие множества $\{\omega\}$ семейством интервалов сколь угодно малой меры. (Эффективности этого построения не мешает возможная невычислимость числа σ, так как вместо σ можно воспользоваться любой его оценкой сверху.)

Таким образом, случайность по Соловею и типичность — это одно и то же.

2.3.3. Аксиоматический подход к определению типичности. Возможен и принципиально другой подход к определению типичности, восходящий к Дж. Майхиллу. Можно считать типичность неопределяемым понятием, а наши интуитивные представления о ней — аксиомами.

Например, к арифметике второго порядка или теории множеств можно добавить новый предикатный символ $R(\omega)$, который читается «последовательность ω случайна». Схема аксиом, отражающая наши интуитивные представления о типичности, может быть такой:

$$\forall \omega\, (R(\omega) \Rightarrow \varphi(\omega)) \Leftrightarrow (\mu(\{\omega \mid \neg \varphi(\omega)\}) = 0)$$

(μ — рассматриваемое распределение вероятностей на множестве Ω; легко проверить, что правая часть эквивалентности выразима в арифметике второго порядка). Если не накладывать никаких ограничений на формулу φ, то полученная теория будет, как легко видеть, противоречивой. Если требовать, чтобы формула φ не содержала свободных переменных (не считая ω) и не содержала символа R, то теория будет непротиворечивой (но неинтересной). Авторам неизвестно, что получится, если разрешить использование символа R в формуле φ, но не допускать свободных переменных, кроме ω. Можно заметить лишь, что получающаяся теория несовместна с аксиомой конструктивности $V = L$.

Аксиоматический подход к определению понятия случайности с использованием идей частотного подхода (см. главу 6) изучается в недавно появившейся работе М. Ламбальгена [Лам 89].

ГЛАВА 3
СЛОЖНОСТЬ, ЭНТРОПИЯ И ХАОТИЧЕСКИЕ ПОСЛЕДОВАТЕЛЬНОСТИ

В этой главе мы изложим сложностной подход к определению понятия случайности, исходящий из естественной идеи о том, что «случайность есть отсутствие закономерностей». Уточнение этой идеи стало возможным после того, как А. Н. Колмогоров [Колм 65] определил понятие энтропии конечного объекта. (В работе [Колм 65] то, что мы называем энтропией, названо сложностью; термин «энтропия» вместо «сложность» появился в [Колм 69].) Сразу же после этого возникло желание определить случайность (бесконечной) последовательности в терминах энтропии ее начальных отрезков.

Однако на этом пути возникли трудности (см. [Мар 66], а также, например, [Яко 70] и [Зво Лев 70]), которые были преодолены в 1973 г. Л. А. Левиным [Лев 73] и К. П. Шнорром [Шно 73]. Для определения случайности бесконечной последовательности в терминах энтропии начальных отрезков пришлось перейти от простой колмогоровской энтропии (введенной в [Колм 65]) к так называемой монотонной энтропии. После этого, определяя хаотическую последовательность как последовательность, энтропия начальных

отрезков которой растет с максимально возможной скоростью (подробнее см. ниже), мы приходим к классу хаотических последовательностей, совпадающему с классом типических (см. главу 2) последовательностей. Это совпадение дает основания считать этот класс уточнением интуитивного представления о «случайной» последовательности.

§ 3.1. Вычислимые отображения

Дадим определение понятия вычислимого отображения множества Σ конечных и бесконечных последовательностей нулей и единиц в себя. Естественный путь сделать это состоит в применении общих конструкций теории f_0-пространств в смысле Ю. Л. Ершова [Ерш 72]; см. об этом в [Шень 84]. Мы, однако, приведем определение вычислимости лишь для отображений пространства Σ. Вначале мы изложим это определение на более абстрактном языке, а затем дадим его более конкретную интерпретацию. (Читатель может выбрать тот из вариантов, который ему ближе.)

Введем на Σ порядок, считая, что $x \leqslant y$, если последовательность x является началом последовательности y. Будем говорить, что x и y *сравнимы*, если $x \leqslant y$ или $y \leqslant x$. Для каждой конечной последовательности x через Σ_x обозначим множество всех ее (конечных и бесконечных) продолжений. (Ранее мы рассматривали множество Ω_x, состоявшее только из бесконечных продолжений.) Рассматривая семейство всех Σ_x как базу топологии, мы превращаем Σ в топологическое пространство (содержащее Ω в качестве подпространства). Отметим, что пространство Σ не является хаусдорфовым (T_2-пространством); оно не является даже T_1-пространством, а является всего лишь T_0-пространством. Вычислимые отображения Σ в себя будем искать среди всюду определенных непрерывных в указанной топологии функций. Легко проверить, что непрерывность всюду определенного отображения $f\colon \Sigma \to \Sigma$ равносильна выполнению двух условий:

(а) если $x, y \in \Sigma$, $x \leqslant y$, то $f(x) \leqslant f(y)$;

(б) значение f на бесконечной последовательности равно минимальному продолжению значений f на всех конечных ее началах: $f(x) = \sup \{f(x_0) \mid x_0$ конечно, $x_0 \leqslant x\}$. (Отметим, что условие (а) гарантирует, что точная верхняя грань в (б) существует.) Таким образом, всякое непрерывное отображение f пространства Σ в себя задается своими значениями на конечных последовательностях. Свяжем с ним множество F тех пар конечных последовательностей $\langle p, q\rangle$, для которых $q \leqslant f(p)$. По этому множеству, очевидно, однозначно восстанавливается f: именно,

$$f(x) = \sup \{q \mid \exists p\, (p \leqslant x \text{ и } \langle p, q\rangle \in F)\}.$$

Соответствующие друг другу F и f будем называть *сопряженными*. Легко проверить, что отношение сопряжения является взаимно-однозначным соответствием между семейством всех непрерывных всюду определенных отображений Σ в себя и семейством всех подмножеств $F \subset \Xi \times \Xi$, обладающих такими свойствами ($p, q, p', q', q_1, q_2 \in \Xi$):

(1) $\langle p, \Lambda\rangle \in F$ для всех $p \in \Xi$;

(2) $\langle p, q\rangle \in F$, $p' \geqslant p$, $q' \leqslant q \Rightarrow \langle p', q'\rangle \in F$;

(3) $\langle p, q_1\rangle \in F$, $\langle p, q_2\rangle \in F \Rightarrow q_1$ и q_2 сравнимы.

Вычислимыми мы будем называть те непрерывные всюду определенные отображения Σ в себя, для которых сопряженное множество пар перечислимо (является множеством значений вычислимой функции, или, что эквивалентно, является множеством всех пар, печатаемых при работе некоторой программы).

Итак, окончательное определение таково. Пусть F — перечислимое множество пар двоичных слов (конечных последовательностей цифр 0 и 1),

обладающее свойствами (1)—(3). Рассмотрим функцию $f\colon \Sigma \to \Sigma$, задаваемую формулой

$$f(x) = \sup\{q \mid \exists p\,(p \leqslant x \text{ и } \langle p, q\rangle \in F)\}.$$

Получаемые таким образом функции (все они оказываются непрерывными) называются *вычислимыми отображениями* Σ в Σ. Подчеркнем, что согласно этому определению вычислимые отображения являются всюду определенными; аналогом ситуации «$f(x)$ не определено» служит ситуация «$f(x)$ — пустая последовательность».

Можно дать и другое, возможно, более наглядное определение вычислимого отображения пространства Σ в себя. Представим себе вычислительную машину, имеющую вход и выход. На вход можно подавать нули и единицы (например, нажимая одну из клавиш «0» и «1»), на выходе машина может выдавать (например, печатать на ленте) также только нули и единицы. Таким образом, работа машины состоит в получении на входе некоторой конечной или бесконечной последовательности нулей и единиц и выдаче на выходе некоторой конечной или бесконечной последовательности нулей и единиц. (Отметим, что в нашем понимании процесс работы машины продолжается неограниченно во времени. При этом машина может проводить вычисления и печатать символы на выходе параллельно с ожиданием очередного символа на входе.)

Вообще говоря, последовательность нулей и единиц, появляющаяся на выходе данной машины, может зависеть не только от входной последовательности, но и от того, в какие моменты времени были поданы на вход ее символы. Мы, однако, будем рассматривать только те машины, в которых выходная последовательность зависит только от входной последовательности, но не от тех моментов, в которые появились символы входной последовательности. Если машина такова, то ей соответствует некоторое отображение Σ в себя (сопоставляющее с каждой входной последовательностью соответствующую выходную). Такие отображения и будут вычислимыми отображениями Σ в Σ.

Например, машина может игнорировать вход и печатать на выходе знаки двоичного разложения числа π. Это означает, что постоянное отображение, значение которого на любой последовательности равно двоичному разложению π, является вычислимым. (Вообще постоянное отображение вычислимо тогда и только тогда, когда его значением является вычислимая последовательность нулей и единиц.) Другой пример вычислимого отображения — тождественное: соответствующая машина копирует входную последовательность на выход.

Использованные в приведенном только что определении машины являются разновидностью «машин с оракулом» (см. [Родж 72, § 9.2]).

§ 3.2. Теорема Колмогорова. Монотонная энтропия

Понятие монотонной энтропии было введено Л. А. Левиным (см. [Лев 73]) и независимо, но в несколько другом варианте, К. П. Шнорром [Шно 73]. (Впоследствии Шнорр в [Шно 77] отказался от своего первоначального варианта и стал работать с понятием из [Лев 73].) Это понятие представляет собой модификацию предложенного А. Н. Колмогоровым определения энтропии (сложности) конечного объекта. Мы не будем излагать определение Колмогорова, отсылая читателя к оригинальной статье [Колм 65] или к [Колм 83а], [Усп Сем 87], а изложим сразу определение монотонной энтропии.

Пусть $f\colon \Sigma \to \Sigma$ — произвольное вычислимое всюду определенное отображение. Пусть y — произвольный элемент Σ. Если $y \leqslant f(x)$ для некоторого $x \in \Sigma$, то будем называть x *описанием* объекта y. *Сложность объекта* y

относительно f мы определим как

$$\inf \{l(x) \mid x \text{ есть описание } y\} = \inf \{l(x) \mid y \leqslant f(x)\}.$$

Здесь через $l(x)$ обозначена длина x; как обычно, мы считаем, что $\inf(\varnothing) = +\infty$. Сложность y относительно f мы обозначим через $KM_f(y)$. Будем говорить, что отображение f не хуже отображения g, если $KM_f(y) \leqslant KM_g(y) + O(1)$, т. е. если существует такая константа C, что $KM_f(y) \leqslant KM_g(y) + C$ для всех $y \in \Xi$ (сокращенная запись $KM_f \overset{+}{\leqslant} KM_g$).

Теорема Колмогорова. *Существует оптимальное* (т. е. *не худшее любого другого*) *отображение.*

Доказательство. Идея доказательства очень проста. Рассмотрим все возможные вычислимые отображения Σ в себя. Их счетное число. Пронумеруем их. Оптимальным будет отображение, построенное в соответствии со следующим правилом: если x есть описание объекта y относительно n-го отображения, то пара $\langle n, x\rangle$ будет описанием объекта y относительно нашего оптимального отображения. Таким образом, описание (для строимого нами отображения) состоит из двух частей — сначала указывается номер (произвольного) вычислимого отображения, а затем описание относительно него. Другими словами, значением нашего отображения (назовем его f) на паре $\langle n, x\rangle$ будет значение n-го отображения на x.

Почему это отображение будет оптимальным? В самом деле, рассмотрим произвольное отображение g. Если x — описание объекта y относительно g, а n — номер отображения q, то пара $\langle n, x\rangle$ будет описанием того же объекта относительно f. Таким образом, объем дополнительной информации, необходимой для перехода от g к f (а именно, объем числа n) зависит лишь от g (но не от y).

В приведенном рассуждении следующие моменты нуждаются в уточнении:

(1) надо пронумеровать все вычислимые отображения, причем так, чтобы построенное нами отображение f было вычислимым;

(2) отображение f должно быть определено не на парах слов, а на словах; таким образом, необходимо какое-то «кодирование» пары слов одним словом;

(3) это кодирование должно быть таким, чтобы длина кода пары $\langle n, x\rangle$ превосходила длину x не более чем на константу (зависящую только от n).

Такие уточнения действительно возможны. Для нумерации всех вычислимых отображений используется «универсальный алгоритм» (умеющий применять любую программу к любому исходному данному). Пару $\langle n, x\rangle$ (где n — натуральное число, x — двоичное слово) можно кодировать словом 0^n1x (здесь 0^n — n нулей подряд). При этом кодировании однозначно восстанавливаются и число n (длина начального отрезка из нулей), и слово x, причем — как нам и требовалось — длина кода пары $\langle n, x\rangle$ превосходит длину x всего лишь на $n + 1$.

Дадим теперь более формальное определение оптимального способа кодирования — сначала на языке множеств пар, а затем на языке машин со входами и выходами.

На языке множеств пар. Нам потребуется перечислимое множество W троек $\langle n, p, q\rangle$, где n — натуральное число, p и q — двоичные слова, обладающее такими свойствами:

(1) при каждом n множество $W_n = \{\langle p, q\rangle \mid \langle n, p, q\rangle \in W\}$ задает вычислимое отображение Σ в себя (т. е. обладает упомянутыми выше в определении вычислимого отображения свойствами (1)—(3));

(2) среди множеств W_n встречаются все множества, задающие вычислимые отображения.

(То, что множество W с такими свойствами существует, иногда выражают так: множество всех вычислимых отображений Σ в себя перечислимо.)

Пусть такое множество W уже построено. Тогда оптимальным будет отображение, задаваемое множеством пар

$$U = \{\langle 0^n 1p, q\rangle \mid \langle n, p, q\rangle \in W\} \cup \{\langle p, \Lambda\rangle \mid p \in \Xi\}.$$

Множество U действительно задает отображение из Σ в себя, т. е. обладает свойствами (1)—(3). Это следует из того, что каждое из множеств W_n обладает этими свойствами и из того, что слово $0^m 1p$ может быть продолжением слова $0^n 1q$, лишь если $m = n$ и p — продолжение q.

Отображение f, задаваемое множеством U, является оптимальным. Если g — любое другое вычислимое отображение, G — задающее его множество пар и n — номер G, т. е. $W_n = G$, то $KM_f(y) \leqslant KM_g(y) + n + 1$ для любого $y \in \Sigma$. В самом деле, если x — описание y относительно g, т. е. $y \leqslant g(x)$, то $\langle x, y\rangle \in G$, $\langle n, x, y\rangle \in W$, $\langle 0^n 1x, y\rangle \in U$ и $0^n 1x$ — описание y относительно f, имеющее длину всего лишь на $n + 1$ большую по сравнению с x. (Мы предполагаем, что y конечно; если y бесконечно, необходимо провести аналогичные рассуждения относительно любого его конечного начала y').

Итак, осталось построить множество W. Известно, что существует универсальное перечислимое множество V троек $\langle n, p, q\rangle$, т. е. такое множество, что среди его сечений

$$V_n = \{\langle p, q\rangle \mid \langle n, p, q\rangle \in V\}$$

встречаются все перечислимые множества. Разумеется, среди сечений будут и «плохие» — не обладающие свойствами (1) — (3) и не задающие вычислимых отображений. Мы преобразуем множество V так, чтобы после преобразования все сечения стали бы «хорошими», а те сечения, которые и раньше были хорошими, не изменились. После этого мы получим искомое множество W. Преобразование множества V состоит в следующем: перечисляя элементы V, мы будем «устранять противоречия» и «заполнять пробелы». Устранение противоречий состоит в следующем. Если очередной элемент $\langle n, p, q\rangle$ противоречит какому-то уже включенному в W элементу $\langle m, r, s\rangle$ в том смысле, что $m = n$, r сравнимо с p, а q не сравнимо с s (напомним, что мы называем два слова *сравнимыми*, если одно из них есть начало другого), то $\langle n, p, q\rangle$ «вычеркивается» (не включается в перечисление W). «Заполнение пробелов», напротив, добавляет в преобразованное множество некоторые элементы, которых раньше не было. Именно, если в множество W включена тройка $\langle n, p, q\rangle$, то вместе с ней мы включаем все тройки $\langle n, p', q'\rangle$, где $p \leqslant p'$, $q' \leqslant q$.

Легко проверить, что описанные преобразования действительно делают «хорошими» все сечения, не меняя тех, которые и так были хорошими.

На языке машин со входами и выходами. Перенумеруем все машины — или все программы — описанного типа. Оптимальная машина будет действовать следующим образом. Она будет ждать появления на входе единицы, считая поступающие нули. Как только (и если) появится единица, начинается имитация n-й машины, где n — число нулей, предшествующих первой единице. Все дальнейшие знаки (после первой единицы) будут рассматриваться как вход для n-й машины. Тем самым мы добьемся того, что если x — описание y относительно n-й машины, то $0^n 1x$ будет описанием y относительно построенной «оптимальной» машины. Таким образом, при переходе от n-й машины к «оптимальной» описание удлиняется не более чем на $n + 1$, что и требовалось.

Это описание, однако, содержит серьезный пробел. Дело в том, что наши машины (или программы) должны удовлетворять требованию корректности, состоящему в том, что результат их работы (напечатанная на выходе последовательность) зависит только от поступивших на вход символов (но не от моментов времени, в которые они поступили). Чтобы построенная нами оптимальная машина (программа) была корректна в этом смысле, нужно ими-

тировать только корректные машины (программы). Если примененный нами язык программирования таков, что не все написанные на нем программы корректны в этом смысле, то нам понадобится отличать корректные программы от некорректных. А это (при любом естественном выборе языка программирования) невозможно. Таким образом, наша конструкция кажется неосуществимой. К счастью, ее можно спасти. Хотя мы и не можем отличать корректную программу от некорректной, можно выделить некоторый класс K программ, для которого:

(1) по всякой программе можно эффективно определить, принадлежит ли она классу K;

(2) все программы из K корректны;

(3) для любой корректной программы существует эквивалентная (задающая то же отображение) программа из класса K.

Имея такой класс, достаточно имитировать только принадлежащие ему программы. Но построить этот класс, пожалуй, не проще, чем рассматривать множества пар, задающие вычислимые отображения. Мы этого делать не будем и предлагаем читателю либо поверить нам, что это возможно, либо разобраться в приведенном выше рассуждении на языке множеств пар.

Пусть f и g — два оптимальных отображения. Соответствующие сложности KM_f и KM_g связаны неравенствами $KM_f \leqslant KM_g + C_1$ и $KM_g \leqslant KM_f + C_2$ для некоторых C_1 и C_2. Другими словами, KM_f и KM_g отличаются на ограниченное слагаемое: $KM_f(x) = KM_g(x) + O(1)$. Удобно фиксировать некоторое оптимальное отображение f, назвать сложность $KM_f(x)$ *монотонной энтропией* последовательности x и обозначать ее $KM(x)$ (без указания на f). Следует помнить, однако, о том, что фиксированный нами способ описания f ничем не лучше любого другого, так что по существу функция KM определена лишь с точностью до ограниченного слагаемого. Отметим еще, что хотя наше определение применимо и к бесконечным последовательностям — при этом конечную энтропию имеют вычислимые последовательности и только они,— мы всюду далее будем рассматривать лишь энтропию двоичных слов (конечных последовательностей нулей и единиц). Поскольку мы не будем касаться других алгоритмических вариантов энтропии (кроме монотонной), то всюду в дальнейшем под *энтропией* мы понимаем монотонную энтропию (если не оговорено противное).

З а м е ч а н и е. В [Колм Усп 87], [Усп Сем 87] дается другое определение монотонной энтропии, эквивалентное нашему в том смысле, что разница между энтропиями в смысле обоих определений ограничена. Именно, *способами описания* там названы перечислимые отношения $R \subset \Xi \times \Xi$, для которых $\langle x_1, y_1\rangle \in R$, $\langle x_2, y_2\rangle \in R$, $x_1 \leqslant x_2 \Rightarrow y_1$ и y_2 сравнимы. После этого сложностью слова y относительно способа описания R назван $\min\{l(x) \mid \langle x, y\rangle \in R\}$, а энтропией названа сложность относительно оптимального способа описания (такого, для которого сложность минимальна — с точностью до константы). Эквивалентность этого определения и используемого нами легко установить, используя следующие соображения: (1) если f — вычислимое отображение Σ в Σ, то сопряженное с ним множество пар является способом описания в смысле [Колм Усп 87]; (2) если R — способ описания в смысле [Колм Усп 87], то отображение $f\colon \Sigma \to \Sigma$, определенное соотношением

$$f(u) = \sup\{y \in \Xi \mid (\exists x \in \Xi)\ (\langle x, y\rangle \in R \text{ и } x \leqslant u)\}$$

является вычислимым.

§ 3.3. Хаотические последовательности

Сейчас мы обобщим данное в главе 1 (для случая равномерного бернуллиева распределения вероятностей) определение хаотичности на случай произвольного вычислимого распределения вероятностей. (Мы называли распределение вероятностей на пространстве Ω *вычислимым*, если существует ал-

горитм, вычисляющий меру множества Ω_x всех продолжений любого заданного слова x с любой заданной точностью.)

Хаотической естественно назвать последовательность, у которой энтропия начальных отрезков растет (с увеличением длины отрезка) с максимально возможной быстротой. Слова «с максимально возможной» требуют, разумеется, уточнения.

Л е м м а. *Пусть P — вычислимое распределение вероятностей на пространстве Ω. Тогда существует такая константа C, что для любого двоичного слова x выполнено неравенство* $KM(x) \leqslant -\log_2 P(\Omega_x) + C$. (Напомним, Ω_x — множество всех продолжений слова x.)

Формулировка этой леммы подсказывает

О п р е д е л е н и е. Пусть P — вычислимое распределение вероятностей на пространстве Ω. Последовательность $\omega \in \Omega$ называется *хаотической* относительно P, если множество

$$\{(-\log_2 P(\Omega_x)) - KM(x) \mid x \text{ является конечным началом последовательности } \omega\}$$

ограничено.

Другими словами, хаотичны те последовательности, для начальных отрезков которых неравенство $KM(x) \leqslant -\log_2 P(\Omega_x) + O(1)$ превращается в равенство.

Д о к а з а т е л ь с т в о л е м м ы. С каждым двоичным словом x сопоставим некоторый отрезок V_x действительной прямой таким образом, чтобы длина отрезка V_x равнялась $P(\Omega_x)$, пустому слову соответствовал отрезок $[0, 1]$, а отрезки, соответствующие продолжениям $x0$ и $x1$ слова x, не пересекались и в сумме давали отрезок V_x, причем отрезок V_{x0} находился левее отрезка V_{x1}. Очевидно, что соответствие $x \mapsto V_x$ однозначно определяется этими требованиями. Такое соответствие может быть построено для любого распределения вероятностей на Ω; для равномерного распределения эта конструкция дает семейство I_x отрезков, состоящих из тех чисел, двоичное разложение которых начинается на x. (Мы опускаем очевидные оговорки, связанные с концами отрезков.) Эти отрезки нам также понадобятся наряду с отрезками V_x, построенными по распределению P.

Определим вычислимое отображение f так, чтобы слово y являлось описанием непустого слова x относительно f, если отрезок I_y содержится во внутренности отрезка V_x. Другими словами, наше вычислимое отображение множества Σ в себя задается множеством пар

$$\{\langle y, x\rangle \mid x = \Lambda \text{ или } I_y \text{ содержится во внутренности } V_x\}.$$

Это множество перечислимо (является множеством значений вычислимой функции). В самом деле, концы отрезков V_x могут быть вычислены с любой точностью (ибо такова мера P). Поэтому, если I_y содержится во внутренности V_x, то при вычислении концов отрезков V_x это рано или поздно выяснится. (Именно для этого необходима оговорка о внутренности; если I_y на самом деле примыкает к краю V_x, то никакая сколь угодно высокая точность не позволит установить, что это так и что I_y не вылезает за пределы V_x.) Вычисляя концы всех отрезков V_x со все возрастающей точностью и выписывая все обнаружившиеся пары $\langle y, x\rangle$, мы получим перечисление интересующего нас множества пар. Легко проверить, что это множество задает вычислимое отображение (т. е. обладает свойствами (1) — (3) § 3.1).

Осталось доказать, что для построенного отображения справедливо неравенство

$$KM_f(x) \leqslant -\log_2 P(\Omega_x) + O(1).$$

В самом деле, если $-\log_2 P(\Omega_x)$ не превосходит некоторого натурального n, то длина отрезка V_x не меньше 2^{-n}. Поэтому отрезок V_x (как и всякий отрезок такой длины) содержит отрезок I_y (для некоторого y) длины $(1/4)\cdot 2^{-n}$

строго внутри себя. Тогда y будет описанием x, причем $l(y) \leqslant n + 2$, и, таким образом, $KM_f(x) \leqslant n + 2$. Таким образом, $KM(x)$ может превосходить $-\log_2 P(\Omega_x)$ не более чем на 3 (тройка появилась здесь потому, что логарифм может быть не целым). Лемма доказана.

Мы уже упоминали, что классы хаотических и типических последовательностей (для произвольного вычислимого распределения вероятностей) совпадают. Доказательство этого утверждения приведено в следующей главе.

ГЛАВА 4
ЧТО ЖЕ ТАКОЕ СЛУЧАЙНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ?

В этой главе мы докажем, что для любого вычислимого распределения вероятностей на пространстве Ω классы типических (гл. 2) и хаотических (гл. 3) последовательностей совпадают. Это было установлено в 1973 г. независимо Л. А. Левиным [Лев 73] и К. П. Шнорром [Шно 73], [Шно 77]. (Как отмечалось в начале § 3.2, Шнорр использовал первоначально, в работе [Шно 73], другое понятие энтропии; эта разница, однако, практически не влияет на доказательства. В [Шно 77] Шнорр приводит и использует определение монотонной энтропии, эквивалентное изложенному в главе 3.)

§ 4.1. Доказательство теоремы Левина — Шнорра для случая равномерного распределения вероятностей

Теорема Левина — Шнорра. *Последовательность типична относительно равномерного бернуллиева распределения вероятностей на Ω тогда и только тогда, когда она хаотична (относительно того же распределения).*

Доказательство. Вспоминая определения типичности и хаотичности, мы видим, что необходимо доказать следующее (для произвольной последовательности $\omega \in \Omega$):

(1) если разность $n - KM((\omega)_n)$ не ограничена, то $\{\omega\}$ — эффективно нулевое множество;

(2) если $\{\omega\}$ — эффективно нулевое множество, то разность $n - KM((\omega)_n)$ не ограничена.

(Напомним, что через $(\omega)_n$ обозначается начальный отрезок длины n последовательности ω.)

Доказательство утверждения (1). Как мы видели в § 1.4, величина $n - KM((\omega)_n)$ может быть неограниченной лишь сверху. Пусть она не ограничена сверху. Тогда для каждого c существует такое начало x последовательности ω, что $KM(x) < l(x) - c$ (здесь $l(x)$ — длина слова x). Обозначим через D_c множество всех двоичных слов x, для которых $KM(x)$ меньше $l(x) - c$. Как мы видели, для каждого c существует начало последовательности ω, принадлежащее D_c. Если представлять себе двоичные слова как вершины дерева (в корне которого находится пустое слово), а бесконечную последовательность — как путь в этом дереве, начинающийся с корня, то можно сказать, что ω упирается в любое из множеств D_c. Чтобы доказать требуемое (нетипичность последовательности ω), нужно, согласно определению, доказать, что $\{\omega\}$ — эффективно нулевое множество. Это означает, что по всякому $\varepsilon > 0$ нужно эффективно указывать множество малой меры, содержащее ω.

В качестве такого множества можно взять множество P_c всех продолжений всех слов из D_c. В самом деле, это множество содержит ω. Докажем, что мера множества P_c не превосходит 2^{-c} (при равномерном бернуллиевом распределении). В самом деле, пусть $x_0, x_1, \ldots$ — минимальные элементы множества D_c (т. е. такие, что никакие их собственные начала не принадлежат D_c). Очевидно, $P_c = \bigcup \Omega_{x_i}$, и, таким образом, достаточно доказать, что $\sum 2^{-l(x_i)} \leqslant 2^{-c}$. Так как $l(x_i) \geqslant KM(x_i) + c$, то достаточно доказать, что

$\sum 2^{-KM(x_i)} \leqslant 1$. Это вытекает из следующей простой леммы (чтобы не прерывать изложение, ее доказательство мы вынесли в § 4.3).

Лемма 1. *Пусть* $x_0, x_1, \ldots$ *— попарно несравнимые двоичные слова (т. е. ни одно из них не является началом другого). Тогда* $\sum 2^{-KM(x_i)} \leqslant 1$.

Итак, мы умеем находить при любом $\varepsilon > 0$ покрытие множества $\{\omega\}$ интервалами с суммарной мерой меньше ε. Определение эффективно нулевого множества (гл. 1) требует, чтобы это покрытие было указано в виде $\Omega_{X(\varepsilon,0)}$, $\Omega_{X(\varepsilon,1)}, \ldots$, где X — вычислимая функция, причем должно выполняться неравенство $\sum P(\Omega_{X(\varepsilon,i)}) < \varepsilon$. Сравнивая желаемое с достигнутым, видим, что естественно попытаться в качестве $X(\varepsilon, 0)$, $X(\varepsilon, 1), \ldots$ взять элементы множества D_c при достаточно большом c (при котором $2^{-c} < \varepsilon$). При этом вычислимость функции X можно обеспечить благодаря следующей лемме (доказательство см. в § 4.3).

Лемма 2. *Существует алгоритм, перечисляющий все пары* $\langle c, x\rangle$, *для которых* $KM(x) < l(x) - c$.

После этого остается единственная трудность. Дело в том, что в определении эффективно нулевого множества требовалась малость суммы мер покрывающих его интервалов, а у нас доказана малость меры объединения интервалов. Это не одно и то же: различные члены последовательности $X(\varepsilon, 0)$, $X(\varepsilon, 1), \ldots$ могут оказаться сравнимыми, а соответствующие интервалы — пересекающимися. Попытавшись преодолеть эту трудность, оставив в последовательности лишь минимальные элементы, мы нарушим ее вычислимость. Выход из положения дает такая

Лемма 3. *По всякой вычислимой последовательности* $x_0, x_1, \ldots$ *двоичных слов можно эффективно построить вычислимую последовательность* $y_0, y_1, \ldots$ *попарно несравнимых двоичных слов, для которой* $\bigcup \Omega_{x_i} = \bigcup \Omega_{y_i}$. (Эффективность понимается в том смысле, что по программе, вычисляющей члены первой последовательности, можно эффективно получить программу вычисления членов второй.)

Доказательство леммы дано в § 4.3.

Применение этой леммы завершает доказательство утверждения «типичность $\Rightarrow$ хаотичность» теоремы Левина — Шнорра. Перейдем теперь к обратному утверждению («хаотичность $\Rightarrow$ типичность»).

Пусть последовательность ω не типична. Докажем, что величина n — $KM((\omega)_n)$ не ограничена. Согласно определению типичности, нетипичная последовательность ω содержится в эффективно нулевом множестве. Это означает, что мы можем эффективно указать объединение интервалов сколь угодно малой суммарной меры, содержащее ω. Мы должны использовать это для отыскания у ω начальных отрезков с небольшой энтропией. Идея тут, грубо говоря, состоит в следующем. Если мы знаем, что ω принадлежит множеству M малой меры, то можно построить вычислимое отображение, «ориентированное на описание элементов M» — отображение, при котором элементы M будут иметь короткие описания (за счет того, что элементы не из M будут иметь длинные описания или не иметь описаний вовсе). Нужно, разумеется, уточнить это высказывание — хотя бы потому, что множество M состоит из бесконечных последовательностей, а энтропия определяется для конечных. Такое уточнение дается в следующей лемме.

Лемма 4. *Пусть* A *— эффективно нулевое множество. Тогда по всякому* c *можно эффективно указать вычислимое отображение* $f\colon \Sigma \to \Sigma$, *удовлетворяющее такому требованию: всякая последовательность* $\omega \in A$ *имеет начало* x, *для которого* $KM_f(x) < l(x) - c$. (Эффективность здесь, как обычно, понимается в том смысле, что по c можно эффективно указать алгоритм перечисления сопряженного с отображением f множества пар.)

Этого, казалось бы, еще мало для завершения доказательства теоремы, так как в этой лемме для каждого c указывается свое отображение f. Но этот дефект может быть легко устранен. Рассмотрим достаточно быстро растущую

вычислимую последовательность c_0, c_1, ... натуральных чисел и построенную с помощью леммы последовательность отображений f_0, f_1, ... Соединим их в одно вычислимое отображение f (подобно тому, как мы это делали при доказательстве теоремы Колмогорова). Именно, будем считать, что $f(0^n1x) = f_n(x)$. Тогда $KM_f(y) \leqslant KM_{f_n}(y) + n + 1$. Таким образом, если $\omega \in A$, то для любого n существует начало x последовательности ω, для которого $KM_{f_n}(x) < l(x) - c_n$, и, следовательно, $KM_f(x) \leqslant KM_{f_n}(x) + n + 1 < l(x) - c_n + n + 1$. Таким образом, при подходящем выборе c_n (например, $c_n = 2n$) мы получаем, что величина $l(x) - KM_f(x)$ для начальных отрезков последовательности ω не ограничена сверху, и, следовательно, разность $l(x) - KM(x)$ также не ограничена сверху.

Нам осталось доказать лемму 4. Согласно определению эффективно нулевого множества, для всякого $\varepsilon > 0$ можно эффективно указать вычислимую последовательность двоичных слов x_0, x_1, ..., для которой $\sum P(\Omega_{x_i}) < \varepsilon$ и $A \subset \bigcup \Omega_{x_i}$. Напомним (это обсуждалось при определении эффективно нулевого множества), что последовательность x_0, x_1, ... может быть конечной, но можно считать, что она «не имеет пробелов» (если x_i определено, то x_j определено при всех $j < i$). Любая последовательность $\omega \in A$ имеет начало среди x_i и наша цель будет достигнута, если мы построим вычислимое отображение f, для которого $KM_f(x_i) < l(x_i) - c$ для всех i. Для этого мы выберем подходящим образом (как именно — будет видно) несравнимые двоичные слова y_0, y_1, ... и в качестве f возьмем отображение, относительно которого y_i является описанием x_i. (В терминах множеств пар: рассмотрим множество $\{\langle p, q\rangle \mid q = \Lambda$ или $\exists i\,(y_i$ — начало p и q — начало $x_i)\}$; оно задает искомое вычислимое отображение Σ в себя.) Так как $KM_f(x_i)$ для этого f не превосходит $l(y_i)$, нам достаточно, чтобы $l(y_i)$ было меньше $l(x_i) - c$.

Итак, наша задача состоит в следующем: есть некоторая последовательность чисел n_i (именно, $n_i = l(x_i) - c$); нужно построить последовательность двоичных слов y_i, для которых $l(y_i) = n_i$, причем все y_i должны быть попарно несравнимы.

Необходимым условием существования такой последовательности является условие $\sum 2^{-n_i} \leqslant 1$ (сумма мер непересекающихся множеств Ω_{y_i} не превосходит 1). Оказывается, что оно будет и достаточным. Ради упрощения мы докажем достаточность более сильного условия $\sum 2^{-n_i} \leqslant 1/2$.

Лемма 5. *Пусть n_i — вычислимая последовательность натуральных чисел, причем $\sum 2^{-n_i} \leqslant 1/2$. Тогда можно эффективно указать такую вычислимую последовательность y_i попарно несравнимых двоичных слов, что $l(y_i) \leqslant n_i$.*

Доказательство этой леммы будет дано в § 4.3. Применяя ее, мы видим, что для завершения доказательства леммы 4,— а с ней и всей теоремы — достаточно убедиться в том, что

$$\sum 2^{-l(x_i)+c} = 2^c \cdot \sum 2^{-l(x_i)} \leqslant 1/2;$$

так как $\sum 2^{-l(x_i)}$ равно $\sum P(\Omega_{x_i})$ и по условию не превосходит ε, то достаточно изначально взять малое ε — например, меньшее $1/2^{c+1}$.

Итак, теорема Левина — Шнорра для случая равномерной меры доказана. В следующем параграфе мы обсудим те изменения в доказательстве, которые необходимо сделать для случая произвольной вычислимой меры. А в § 4.3 мы приведем доказательства всех лемм, оставшихся пока недоказанными (леммы 1, 2, 3, 5).

Отметим в заключение любопытное обстоятельство, которое можно усмотреть из проведенного нами рассуждения. Мы доказали, что если ω не типична, то величина $n - KM((\omega)_n)$ не ограничена. Чуть уточнив рассуждение, можно установить, что эта величина не только не ограничена, но и стремится к бесконечности. (Нужно лишь при доказательстве леммы 4 изменить построение

отображения f и считать, что не только y_i является описанием x_i, но и для любого слова z слово y_iz является описанием слова x_iz. Тогда неравенство $KM_f(x) \leqslant l(x) - c$ будет справедливо не только для слов x_i, но и для любых их продолжений.)

Таким образом, для случайных ω разность $n - KM((\omega)_n)$ ограничена, а для неслучайных — стремится к бесконечности. Отсюда вытекает, что не существует последовательностей ω, для которых эта разность была бы неограниченной, но не стремящейся к бесконечности функцией.

§ 4.2. Случай произвольного вычислимого распределения

Приведенное в предыдущем пункте доказательство теоремы Левина — Шнорра легко обобщается на случай произвольного вычислимого распределения вероятностей P. Отметим вкратце, какие изменения требуется в нем сделать.

1. Мы применяли неравенство $KM(x) \leqslant l(x) + O(1)$. Вместо него надо воспользоваться неравенством $KM(x) \leqslant -\log_2 P(\Omega_x) + O(1)$, доказанным в § 3.3.

2. Множество D_c следует определять как множество тех слов x, для которых $KM(x) < -\log_2 P(\Omega_x) - c$. Его элементы, как и раньше, могут быть эффективно перечислены (здесь мы пользуемся вычислимостью распределения P), а множество всех продолжений всех его элементов имеет меру меньше 2^{-c} (по отношению к распределению P).

В остальном рассуждения в первой части доказательства теоремы Левина — Шнорра остаются неизменными. Во второй части доказательства используется следующее обобщение леммы 4.

Лемма 4а. *Пусть A — эффективно нулевое (относительно распределения P) множество. Тогда по всякому c можно эффективно указать вычислимое отображение $f\colon \Sigma \to \Sigma$, удовлетворяющее такому требованию: всякая последовательность $\omega \in A$ имеет начало x, для которого $KM_f(x) < -\log_2 P(\Omega_x) - c$.*

Ее доказательство аналогично доказательству леммы 4, нужно только в качестве n_i брать не $l(x_i) - c$, а $-\log_2 P(\Omega_{x_i}) - c$ (точнее, целые части этих чисел). После этого доказательство теоремы Левина — Шнорра завершается как в § 4.1.

§ 4.3. Доказательства лемм

Лемма 1. *Пусть $x_0, x_1, \ldots$ — попарно несравнимые двоичные слова. Тогда $\sum 2^{-KM(x_i)} \leqslant 1$.*

Доказательство. Пусть f — оптимальное вычислимое отображение, используемое при определении KM, а y_i — кратчайшие описания слов x_i. Слова y_i попарно не сравнимы (если y — общее продолжение x_i и x_j, то $f(y)$ было бы общим продолжением y_i и y_j, а они не сравнимы). Поэтому множества Ω_{y_i} не пересекаются и сумма их мер (относительно равномерного бернуллиева распределения) не превосходит 1. А мера Ω_{y_i} равна $2^{-l(y_i)}$, т. е. равна $2^{-KM(x_i)}$, так как y_i — кратчайшее описание x_i и $l(y_i) = KM(x_i)$. Лемма 1 доказана.

Лемма 2. *Существует алгоритм, перечисляющий все пары $\langle c, x\rangle$, для которых $KM(x) < l(x) - c$.*

Доказательство. Пусть f — оптимальное вычислимое отображение, F — сопряженное с ним множество пар слов:

$$F = \{\langle p, x\rangle \mid x \text{ — начало } f(p),\ p,\ x \text{ конечны}\}.$$

Согласно определению вычислимости отображения из Σ в Σ, существует алгоритм, перечисляющий все пары $\langle p, x\rangle \in F$. Будем перебирать все такие

пары и все c, проверяя каждый раз, не оказалось ли так, что $l(p) < l(x) - c$. В этом случае пара $\langle c, x\rangle$ такова, что $KM(x) < l(x) - c$, и мы можем включить ее в перечисление фигурирующего в лемме множества пар. При этом процессе мы рано или поздно обнаружим все пары из этого множества, так как для каждого x найдется такое p, что $\langle p, x\rangle \in F$ и $l(p) = KM(x)$. Лемма 2 доказана.

Лемма 3. *По всякой вычислимой последовательности $x_0, x_1, \ldots$ двоичных слов можно эффективно построить вычислимую последовательность $y_0, y_1, \ldots$ попарно несравнимых двоичных слов, для которой $\bigcup \Omega_{x_i} = \bigcup \Omega_{y_i}$.*

Доказательство. Назовем *интервалом* любое подмножество множества Ω, имеющее вид Ω_x для некоторого слова x, и простым множеством — любое конечное объединение интервалов. Легко видеть, что всякое простое множество может быть представлено в виде объединения непересекающихся интервалов, и что разность двух простых множеств является простым множеством. (Это вытекает из того, например, что простые множества — это те, принадлежность последовательности к которым определяется начальным отрезком фиксированной длины.)

Теперь легко сообразить, как переделать последовательность $x_0, x_1, \ldots$ в последовательность $y_0, y_1, \ldots$ с требуемыми свойствами. В каждый момент, получив на входе конечное число членов $x_0, \ldots, x_k$, мы будем давать на выходе такие $y_0, \ldots, y_l$, что $\Omega_{x_0} \cup \ldots \cup \Omega_{x_k} = \Omega_{y_0} \cup \ldots \cup \Omega_{y_l}$ и все y_i попарно не сравнимы. При поступлении очередного члена x_{k+1} мы должны добавить в правую часть непересекающиеся интервалы, которые в объединении дают $(\Omega_{x_0} \cup \ldots \cup \Omega_{x_{k+1}}) \setminus (\Omega_{x_0} \cup \ldots \cup \Omega_{x_k})$. Это можно сделать, так как указанное множество — простое. Легко понять, что наши преобразования сохраняют вычислимость последовательности. Лемма 3 доказана.

Лемма 5. *Пусть n_i — вычислимая последовательность натуральных чисел, причем $\sum 2^{-n_i} \leqslant 1/2$. Тогда можно эффективно указать такую вычислимую последовательность y_i попарно несравнимых двоичных слов, что $l(y_i) \leqslant n_i$.*

Доказательство. Сопоставим, как мы это уже делали в § 3.3, с каждым двоичным словом x отрезок I_x на числовой оси, состоящий из чисел, двоичная запись которых начинается на x (с обычными оговорками о концах отрезка). Все отрезки вида I_x будем называть *регулярными*. Кроме них, будем откладывать на отрезке $[0, 1]$, начиная с его левого конца, отрезки $S_0, S_1, \ldots$ длиной $2 \cdot 2^{-n_0}, 2 \cdot 2^{-n_1}, \ldots$ По условию, они не выйдут за пределы отрезка $[0, 1]$. Для каждого из построенных отрезков S_i рассмотрим наибольший регулярный отрезок I_{y_i}, в нем содержащийся. Так как в любом отрезке длины l есть регулярный отрезок длины не меньше $l/2$, то длина I_{y_i} не меньше 2^{-n_i}, т. е. $l(y_i) \leqslant n_i$. Кроме того, отрезки I_{y_i} при различных i не пересекаются (так как отрезки S_i не пересекаются) и, следовательно, слова y_i попарно несравнимы. Лемма 5 доказана.

ГЛАВА 5
ВЕРОЯТНОСТНЫЕ МАШИНЫ, АПРИОРНАЯ ВЕРОЯТНОСТЬ И СЛУЧАЙНОСТЬ

§ 5.1. Вероятностные машины

Понятие вероятностной машины является формализацией представлений о вычислительном устройстве, снабженном «датчиком случайных чисел». Такое устройство состоит из двух частей — датчика случайных чисел и детерминированной части, перерабатывающей информацию от датчика.

Будем считать, что выдаваемая датчиком информация представляет собой последовательность нулей и единиц. Тогда датчик характеризуется распределением вероятностей, т. е. мерой на пространстве Ω бесконечных последовательностей нулей и единиц, для которой мера всего пространства равна 1.

При исследовании возможностей вероятностных машин необходимо наложить какие-то ограничения на датчик случайных чисел, т. е. на соответствующее распределение. (В противном случае может оказаться, например, что одна-единственная бесконечная последовательность нулей и единиц образует множество полной меры. Эта последовательность может содержать любую информацию — и тем самым, грубо говоря, вероятностные машины могут все!)

Естественным требованием является требование вычислимости распределения вероятностей, которому подчиняется датчик. (Вычислимость, напомним, означает, что вероятность того, что последовательность будет начинаться на слово x, может быть эффективно вычислена по x с любой заданной точностью; см. § 2.1.) Простейшим из вычислимых датчиков, по-видимому, можно считать датчик, в котором нули и единицы появляются с равной вероятностью и каждый символ появляется независимо от предыдущих. Такому датчику соответствует равномерное бернуллиево распределение вероятностей на пространстве Ω. Мы увидим, что с некоторой точки зрения этого датчика достаточно — любой датчик с вычислимым распределением вероятностей может быть «промоделирован» с его помощью.

Перейдем к точным определениям. *Вероятностной машиной* мы будем называть пару $\langle P, f\rangle$, где P — вычислимое распределение вероятностей на Ω, а f — вычислимое отображение Σ в себя. Неформальный смысл этой пары таков: из датчика с распределением P появляются нули и единицы; к возникающей последовательности применяется отображение f. Его значение и есть результат работы вероятностной машины. (Здесь мы рассматриваем вероятностные машины не как средство вычисления функций, а как средства порождения последовательностей нулей и единиц, поэтому они не имеют «входа» — если не считать датчика.) Каждая вероятностная машина задает, таким образом, случайную величину со значениями в пространстве Σ конечных и бесконечных последовательностей нулей и единиц, и можно рассмотреть ее закон распределения.

Пусть $\langle P, f\rangle$ — вероятностная машина. Свяжем с ней распределение вероятностей Q на пространстве Σ конечных и бесконечных последовательностей нулей и единиц. Именно, меру $Q(A)$ произвольного борелевского подмножества $A \subset \Sigma$ определим как $P(f^{-1}(A))$. (Строго говоря, здесь следовало бы написать $P(f^{-1}(A) \cap \Omega)$, так как в прообраз $f^{-1}(A)$ могут входить и конечные последовательности. Для краткости мы этого здесь и далее делать не будем, отождествляя распределение P на множестве Ω с распределением на Σ, совпадающим с P на Ω и равным нулю на Ξ). Отметим, что хотя распределение P сосредоточено на бесконечных последовательностях, про распределение Q этого сказать нельзя, так как отображение f может принимать конечные значения на бесконечных аргументах. Построенное таким образом распределение Q будем называть *распределением, задаваемым вероятностной машиной* $\langle P, f\rangle$.

Возникает вопрос: какие распределения на Σ задаются вероятностными машинами? Оказывается, такие распределения имеют простое описание. Обозначим через Σ_x множество всех конечных и бесконечных последовательностей, являющихся продолжением конечной последовательности x. Каково бы ни было распределение вероятностей Q на множестве Σ (определенное на борелевских подмножествах Σ),

(1) $Q(\Sigma_\Lambda) = 1$ (мера всего Σ равна 1);

(2) $Q(\Sigma_{x0}) + Q(\Sigma_{x1}) \leqslant Q(\Sigma_x)$ для всякого двоичного слова x.

Неравенство (2) не является равенством, если мера Q множества $\{x\}$ отлична от нуля. Стандартное теоретико-мерное рассуждение показывает, что распределение однозначно задается своими значениями на множествах вида Σ_x, причем единственное, что от этих значений требуется — это соблюдение условий (1) и (2): если q — произвольная функция на двоичных словах с неотрицательными значениями, для которой $q(\Lambda) = 1$ и $q(x0) + q(x1) \leqslant q(x)$ для лю-

бого слова x, то существует и единственно распределение вероятностей Q на пространстве Σ, для которого $Q(\Sigma_x) = q(x)$ при всех x.

Таким образом, вопрос можно поставить так: какими свойствами должна обладать функция q, чтобы соответствующее ей распределение Q задавалось некоторой вероятностной машиной? Чтобы дать ответ на этот вопрос, введем понятие полувычислимого снизу действительного числа. Назовем число $x \in \mathbb{R}$ *полувычислимым снизу*, если существует монотонно возрастающая вычислимая последовательность рациональных чисел, сходящаяся к x. Всякое вычислимое число полувычислимо снизу: если x_n — его приближение с точностью до $1/n$, $y_n = x_n - 1/n$, а $z_n = \max(y_1, \ldots, y_n)$, то последовательность $z_1, z_2, \ldots$ и будет искомой. Эквивалентные определения полувычислимости снизу: «x есть точная верхняя грань перечислимого множества рациональных чисел»; «множество всех рациональных чисел, меньших x, перечислимо». Аналогичным образом определяется полувычислимость сверху. Очевидно, число x полувычислимо снизу тогда и только тогда, когда число $(-x)$ полувычислимо сверху. Отметим также, что если x полувычислимо снизу и сверху, то x вычислимо: перечисляя рациональные числа, сходящиеся к x с двух сторон, мы можем дождаться момента, когда разница между ними будет сколь угодно мала.

Будем говорить, что функция q, аргументами которой являются двоичные слова, а значениями — действительные числа, полувычислима снизу, если все ее значения $q(x)$ полувычислимы снизу, и соответствующие программы могут быть эффективно найдены по x. Более точно — и чуть-чуть в других терминах,— q *полувычислима снизу*, если существует такая вычислимая функция $\langle x, n\rangle \mapsto \bar{q}(x, n)$ с рациональными значениями, что для каждого слова x и для любого натурального n значение $\bar{q}(x, n)$ определено и последовательность $\bar{q}(x, 0), \bar{q}(x, 1), \ldots$ возрастает и сходится к $q(x)$.

Эквивалентное определение в терминах перечислимых множеств: функция q полувычислима снизу, если множество пар

$\{\langle x, r\rangle \mid x$ — двоичное слово, r — рациональное число, меньшее $q(x)\}$

перечислимо.

Оказывается, что полувычислимость снизу функции q необходима и достаточна для того, чтобы соответствующее распределение вероятностей задавалось вероятностной машиной.

Т е о р е м а. *Пусть Q — произвольное распределение на Σ, $q(x) = Q(\Sigma_x)$ для всех слов x. Тогда следующие свойства равносильны*:

(1) *существует вероятностная машина, задающая распределение Q*;

(2) *существует вероятностная машина $\langle P, f\rangle$, задающая распределение Q, причем P — равномерное бернуллиево распределение*;

(3) *функция q полувычислима снизу.*

Д о к а з а т е л ь с т в о. Поскольку из (2) очевидно следует (1), достаточно доказать (1) $\Rightarrow$ (3) и (3) $\Rightarrow$ (2). Выше мы уже упоминали эквивалентность условий (1) и (2), говоря, что достаточно ограничиться вероятностными машинами с простейшим датчиком случайных чисел.

Покажем сначала, что из (1) следует (3), т. е. что для любой вероятностной машины $\langle P, f\rangle$ функция q: $x \mapsto P(f^{-1}(\Sigma_x))$ полувычислима снизу. В самом деле, прообраз $f^{-1}(\Sigma_x)$ можно представить в виде объединения множеств Σ_y по всем $y \in \Xi$, являющимся описаниями x относительно f: $f^{-1}(\Sigma) = \bigcup \{\Sigma_y \mid x \leqslant f(y)\}$. Множество таких y перечислимо; расположив его в последовательность $y_0, y_1, \ldots$, мы видим, что $q(x) = \lim P(\Omega_{y_0} \cup \ldots \ldots \cup \Omega_{y_n})$ при $n \to \infty$. Число $P(\Omega_{y_0} \cup \ldots \cup \Omega_{y_n})$ вычислимо: надо выбрать среди y_i те, которые не являются продолжениями других, и сложить меры соответствующих Ω_{y_i} (здесь мы пользуемся тем, что сумма вычислимых чисел вычислима). Таким образом, $q(x)$ есть предел возрастающей последовательности вычислимых действительных чисел; нетрудно переделать ее в вычислимую последовательность рациональных чисел (взяв при

ближения снизу с достаточной точностью и обеспечив возрастание заменой-n-го числа на максимум из первых n членов). Эта последовательность, как легко видеть, окажется вычислимой, и, таким образом, число $q(x)$ полувычислимо снизу. Все наши построения могут быть эффективно выполнены при заданном x, и, таким образом, функция q полувычислима снизу. Импликация (1) $\Rightarrow$ (3) доказана.

Более сложно доказательство импликации (3) $\Rightarrow$ (2). Назовем *эффективно открытыми* в пространстве Σ те его подмножества, которые представимы в виде $\bigcup \{\Sigma_s \mid s \in S\}$, где S — некоторое перечислимое множество двоичных слов. Программу, его перечисляющую, будем называть *программой* соответствующего эффективно открытого множества. (У одного и того же множества могут быть, конечно, разные программы). Если f — вычислимое отображение пространства Σ в себя, а $T_y = f^{-1}(\Sigma_y)$ для любого $y \in \Xi$, то легко проверить, что:

(а) T_y для любого $y \in \Xi$ является эффективно открытым множеством, некоторую программу можно эффективно получить по y;

(б) $T_\Lambda = \Sigma$; для любого y множества T_{y0} и T_{y1} представляют собой непересекающиеся подмножества множества T_y.

Обратно, если для каждого y задано множество T_y, причем выполнены условия (а) и (б), то существует единственное вычислимое отображение f пространства Σ в себя, для которого $f^{-1}(\Sigma_y) = T_y$.

Если $\langle P, f\rangle$ — вероятностная машина, а T_y — семейство эффективно открытых множеств, соответствующее отображению f, то мера Q, задаваемая этой вероятностной машиной, может быть описана так: $Q(\Sigma_y) = P(T_y)$. Таким образом, нам нужно доказать, что для заданной перечислимой снизу функции q можно построить семейство множеств T_y, удовлетворяющее условиям (а) и (б), для которого $P(T_y) = q(y)$ при любом y.

Эти множества следует строить индукцией по длине y. Вначале положим $T_\Lambda = \Sigma$. Затем следует воспользоваться следующей леммой.

Лемма. *Пусть задано эффективно открытое множество X и два полувычислимых снизу неотрицательных действительных числа r и s, причем $r + s \leqslant P(X)$. Тогда можно эффективно указать два непересекающихся эффективно открытых подмножества Y и Z множества X, для которых $P(Y) = r$, $P(Z) = s$.* (Говоря об эффективности, мы имеем в виду, что существует алгоритм, аргументами которого являются программа для X и программы вычисления возрастающих приближений к r и s, а выходом — программы для Y и Z.)

Применяя эту лемму к множеству T_Λ и к числам $q(0)$ и $q(1)$, получаем множества T_0 и T_1. Затем, применяя ее к множеству T_0 и к числам $q(00)$ и $q(01)$, строим множества T_{00} и T_{01} и так далее: множества T_{x0} и T_{x1} получаются применением леммы к множеству T_x и числам $q(x0)$ и $q(x1)$.

Набросок доказательства леммы. Будем строить эффективно открытые множества Y и Z так. Множество X представляет собой объединение множеств Σ_x по всем x из некоторого перечислимого множества S. Эти множества мы будем «распределять» между Y и Z. Поскольку r и s вычислимы снизу, существуют возрастающие вычислимые последовательности двоично рациональных чисел, сходящиеся к r и s. Их очередные члены мы будем называть «текущими приближениями» к r и s. Получив очередной интервал из множества X, мы будем распределять его между множествами Y и Z так, чтобы (1) меры уже построенных частей множеств Y и Z не превосходили текущих приближений к числам r и s соответственно и (2) эти меры были равны текущим приближениям, если возможно (т. е. если мера рассмотренных интервалов из множества X достаточна). Как только очередной интервал из множества X распределен полностью, переходим к следующему интервалу. Нетрудно проверить, что построенные таким образом множества Y и Z удовлетворяют требованиям леммы.

На этом доказательство сформулированной выше теоремы заканчивается.

Итак, мы имеем характеризацию образов вычислимых мер при вычислимых отображениях. Часто функции q, определенные на всех двоичных словах, принимающие неотрицательные действительные значения и обладающие свойствами $q(\Lambda) = 1$, $q(x0) + q(x1) \leqslant q(x)$ для всех x, называют *полумерами* и отождествляют с соответствующими распределениями на пространстве Σ. При этом отождествлении распределениям, задаваемым вероятностными машинами, соответствуют полувычислимые снизу полумеры.

§ 5.2. Априорная вероятность

Как мы видели в § 5.1, каждой вероятностной машине соответствует некоторое распределение на Σ. Следующая теорема показывает, что среди этих распределений существует «наибольшее».

Т е о р е м а. *Существует такое распределение M на Σ, соответствующее некоторой вероятностной машине, что для любого распределения Q на Σ, соответствующего некоторой вероятностной машине, найдется такое c, что $Q(A) \leqslant c \cdot M(A)$ для любого (борелевского) множества $A \subset \Sigma$.* (Обозначение: $Q \lesssim M$.)

Д о к а з а т е л ь с т в о. Как следует из теоремы § 5.1, достаточно рассматривать вероятностные машины, использующие равномерный датчик случайных чисел (т. е. с равномерной бернуллиевой мерой в качестве первой компоненты). Построим машину, работающую следующим образом. Вначале она, используя датчик случайных чисел, выбирает случайно некоторое натуральное число n. Способ этого выбора может быть любым; необходимо лишь, чтобы вероятность выбора каждого натурального числа была ненулевой. (Например, можно считать, что n — это число появлений решетки до первого герба.) Затем моделируется работа вероятностной машины вида «равномерный датчик, n-ое вычислимое отображение», т. е. все дальнейшие результаты, получаемые от датчика случайных чисел, образуют последовательность, подаваемую на вход n-го вычислимого отображения. (О нумерации вычислимых отображений множества Σ в себя мы уже говорили выше в § 3.2.)

Распределение, соответствующее описанной машине, обозначим через M. Пусть Q — распределение на Σ, соответствующее произвольной вероятностной машине. Можно считать, что эта машина есть $\langle P, f\rangle$, где P — равномерное распределение на Ω, f — вычислимое отображение множества Σ в себя. Тогда имеется ненулевая вероятность p того, что построенная нами «универсальная машина» будет моделировать машину $\langle P, f\rangle$: именно, p не меньше вероятности появления номера отображения f при случайном выборе числа n. Поэтому $M(A) \geqslant p \cdot Q(A)$ и $Q(A) \leqslant (1/p) \cdot M(A)$, что и требовалось.

Более формально, можно для построения искомой меры использовать вероятностную машину $\langle P, f\rangle$, где P — равномерное распределение, а f — оптимальное вычислимое отображение, построенное при доказательстве теоремы Колмогорова в § 3.2. Это отображение было таким, что $f(0^n 1x) =$ = (значение n-го вычислимого отображения на x). Отсюда следует, что $f^{-1}(A)$ содержит все последовательности вида $0^n 1\alpha$, где α принадлежит прообразу A при n-м вычислимом отображении множества Σ в себя. Поэтому $M(A) \geqslant 2^{-n-1} \cdot Q(A)$, где n — номер отображения, задающего распределение Q. Теорема доказана.

С л е д с т в и е. *Среди всех полувычислимых снизу полумер существует наибольшая с точностью до мультипликативной константы.*

Отметим, что это следствие утверждает несколько меньше, чем теорема: в нем устанавливается соотношение между мерами множеств вида Σ_x (откуда, вообще говоря, не следует аналогичное соотношение для мер всех множеств). Это следствие может быть доказано и непосредственно. Именно,

все полувычислимые снизу полумеры могут быть эффективно перечислены (перечисляем все полувычислимые снизу функции и модифицируем их так, чтобы превратить их в полумеры, не меняя функции в тех случаях, когда она и так была полумерой). Затем берем произвольный вычислимый ряд $\sum p_i$ с положительными членами и единичной суммой и рассматриваем функцию $m = \sum p_n \cdot (n\text{-я функция в перечислении})$.

Построенную при доказательстве теоремы вероятностную машину можно назвать универсальной в том смысле, что если какая-то машина с положительной вероятностью выдает последовательность из некоторого множества $A \subset \Sigma$, то и эта «универсальная» машина также выдает последовательность из A с положительной вероятностью.

Если M_1 и M_2 — два распределения на Σ, для которых выполнено утверждение теоремы, то при некоторых c_1 и c_2 неравенства $M_1(A) \leqslant c_1 M_2(A)$, $M_2(A) \leqslant c_2 M_1(A)$ выполнены для всех (борелевских) множеств A. Поэтому M_1 и M_2 отличаются на ограниченный и отделенный от нуля множитель.

Выберем и зафиксируем какое-нибудь распределение M, для которого выполнено утверждение теоремы. Будем называть его *априорной вероятностью*.

§ 5.3. Априорная вероятность и энтропия

Таким образом, для каждого двоичного слова x определены две его характеристики: его (монотонная) энтропия $KM(x)$ и априорная вероятность $M(\Sigma_x)$ множества всех его продолжений, которую мы будем в дальнейшем обозначать через $m(x)$. Можно сказать, что первая из них определяет, насколько трудно описать объект x (или его продолжения), а вторая — насколько вероятно, что объект x (или его продолжение) получится случайно. Оказывается, что между этими характеристиками имеется тесная связь.

Т е о р е м а. *Справедливы неравенства*

$$-\log_2 m(x) \leqslant KM(x) + O(1); \tag{1}$$

$$KM(x) \leqslant -\log_2 m(x) + O(\log_2 l(x)). \tag{2}$$

Д о к а з а т е л ь с т в о. Первое из этих неравенств почти очевидно. В самом деле, пусть f — оптимальное отображение, используемое при определении монотонной энтропии. Рассмотрим вероятностную машину $\langle P, f\rangle$, где P — равномерное распределение вероятностей на Ω. Если $KM(x) = n$, то существует слово y длины n, для которого $f(y)$ начинается на x. Тогда для любой последовательности $\omega \in \Omega$, начинающейся на y, последовательность $f(\omega)$ принадлежит Σ_x. Поэтому множество $f^{-1}(\Sigma_x)$ содержит все продолжения y и имеет меру не меньше 2^{-n}. Остается воспользоваться максимальностью априорной вероятности и перейти к логарифмам.

Второе неравенство более сложно. Грубо говоря, проблема состоит в том, что прообраз множества Σ_x может быть «раздробленным» и состоять из большого числа множеств вида Σ_t, где все t довольно длинные, но благодаря их количеству объединение $\bigcup \Sigma_t$ имеет достаточно большую меру. Из-за этого трудно рассчитывать на то, что неравенство $KM(x) \leqslant -\log_2 m(x) + O(1)$ будет выполнено для всех x (как показал П. Гач [Гач 83], это неравенство и в самом деле неверно), и нам придется добавить в правую часть $O(\log_2 l(x))$. Отметим, что «типичное» (для большинства слов данной длины) значение $KM(x)$ (а также $-\log_2 m(x)$) по порядку величины есть $l(x)$, так что эта логарифмическая добавка, как правило, мала по сравнению с остальными членами неравенства.

Основная часть доказательства заключена в следующей лемме. Назовем последовательность действительных чисел $p_0, p_1, \ldots$ *полувычислимой снизу*, если существует такая вычислимая функция p: $\langle i, n\rangle \mapsto p(i, n)$ рациональными значениями, определенная на всех парах натуральных

чисел, что при каждом i последовательность $p(i, 0)$, $p(i, 1)$, ... монотонно возрастает и сходится к p_i.

Лемма. *Пусть дана полувычислимая снизу последовательность неотрицательных действительных чисел* $p_0, p_1, \ldots$, *причем* $\sum p_i < \infty$, *и произвольная вычислимая последовательность слов* $x_0, x_1, \ldots$ *Тогда* $KM(x_i) \leqslant -\log_2 p_i + O(1)$.

(Отметим, что в условиях леммы число p_i никак не связано со словом x_i; единственное, что у них общее — это номер!)

Прежде чем доказывать лемму, убедимся, что из нее вытекает неравенство (2) теоремы. Для этого сопоставим с каждым словом x величину $m(x)/l(x)^2$. Ряд $\sum\limits_x m(x)/l(x)^2$ сходится. В самом деле, сумма $m(x)$ по всем словам x данной длины n не превосходит 1, так как они не сравнимы и соответствующие подмножества в Σ не пересекаются, поэтому $\sum\limits_x m(x)/l(x)^2 \leqslant \sum\limits_n 1/n^2$ (группируем слагаемые, соответствующие словам одинаковой длины). Расположим теперь все двоичные слова в произвольном вычислимом порядке в последовательность $x_0, x_1, \ldots$ и через p_i обозначим $m(x_i)/l(x_i)^2$. Тогда последовательность p_i полувычислима снизу (в силу полувычислимости снизу функции m). Применяя лемму, получаем, что $KM(x_i) \leqslant -\log_2 (m(x_i)/l(x_i)^2) + O(1) \leqslant -\log_2 m(x_i) + 2\log_2 l(x_i) + O(1)$, откуда и вытекает требуемое неравенство.

Доказательство леммы. Пусть сначала числа p_i имеют вид 2^{-n_i}, где $n_0, n_1, \ldots$ — вычислимая последовательность целых чисел. Не ограничивая общности, можно предположить, что $\sum p_i$ не превосходит $1/2$ (умножение всех p_i на константу поглощается членом $O(1)$). Вспоминая лемму 5 из § 4.3, мы можем построить вычислимую последовательность $y_0, y_1, \ldots$ несравнимых двоичных слов, для которых $l(y_i) = n_i$. Рассмотрев теперь вычислимое отображение f, для которого $f(y_i) = x_i$, мы видим, что $KM_f(x_i) \leqslant l(y_i) = n_i = -\log_2 p_i$, откуда $KM(x_i) \leqslant -\log_2 p_i + O(1)$.

Если числа p_i вычислимы (равномерно по i, т. е. программа, вычисляющая их приближения, может быть алгоритмически получена по i), то можно заменить каждое из них на число вида $1/2^k$, отличающееся не более чем в 2 раза, и свести дело к уже разобранному случаю.

Рассмотрим теперь общий случай произвольной полувычислимой снизу последовательности p_i. (Именно этот общий случай нам необходим при доказательстве неравенства (2) теоремы.) Здесь мы применим следующий трюк. Вычисляя приближения снизу к p_i, мы будем констатировать все обнаруживаемые соотношения вида $1/2^k < p_i$ (для всевозможных k и i) и все появившиеся при этом числа $1/2^k$ включать в последовательность. Эта последовательность вычислима, суммируема (ее сумма не более чем вдвое превосходит $\sum p_i$, так как $\sum \{2^{-k} \mid 2^{-k} < p\} \leqslant 2p$) и для каждого p_i наилучшее обнаруженное приближение к нему будет отличаться от самого p_i не более чем в 2 раза.

Более формально. Рассмотрим перечислимое множество пар $\langle i, k\rangle$, для которых $2^{-k} < p_i$. Расположим его в последовательность $\langle i(0), k(0)\rangle$, $\langle i(1), k(1)\rangle$, ... Имеет место неравенство $\sum\limits_s 2^{-k(s)} < \infty$, так как $\sum\limits_s 2^{-k(s)} = \sum\limits_i \left(\sum\limits_{i(s)=i} 2^{-k(s)}\right) \leqslant \sum\limits_i (2p_i) < \infty$. Поэтому можно найти вычислимую последовательность несравнимых двоичных слов $y(0), y(1), \ldots$, для которых $l(y(s)) \leqslant k(s) + c$ (при некотором фиксированном c). Построим теперь вычислимое отображение f, для которого $f(y(s)) = x_{i(s)}$. Тогда $KM_f(x_i) \leqslant k(s) + c$ для любых i и s, для которых $i(s) = i$. Пусть i фиксировано. Выберем наименьшее k среди тех, для которых $2^{-k} < p_i$. Тогда $p_i/2 \leqslant 2^{-k} < p_i$. Найдется такое s, что $i(s) = i$, $k(s) = k$. При этом s выполнено

неравенство $k(s) \leqslant -\log_2 p_i + 1$, поэтому $KM_f(x_i) \leqslant -\log_2 p_i + 1 + c$, что и требовалось доказать.

Замечание. На самом деле неравенство (2) доказанной в этом параграфе теоремы может быть несколько усилено:

$$KM(x) \leqslant -\log_2 m(x) + O(\log_2(-\log_2 m(x))).$$

§ 5.4. Априорная вероятность и случайность

В предыдущем параграфе было показано, что монотонная энтропия довольно близка к логарифму априорной вероятности; неудивительно поэтому, что обе эти величины могут служить для характеризации случайности. О первой из них шла речь в главе 4; здесь мы приведем характеризацию случайности (т. е. типичности, или, что эквивалентно, хаотичности) в терминах априорной вероятности.

Теорема. *Пусть P — вычислимое распределение вероятностей на пространстве бесконечных последовательностей нулей и единиц. Тогда для любой последовательности $\omega \in \Omega$ следующие свойства равносильны:*

(1) *ω типична (= хаотична) относительно распределения P;*

(2) *величина $-\log_2 P(\Omega_{(\omega)_n}) - (-\log_2 m((\omega)_n))$ ограничена [напомним, $(\omega)_n$ — начало ω длины n].*

Эта теорема отличается от теоремы Левина — Шнорра лишь заменой $KM((\omega)_n)$ на $-\log_2 m((\omega)_n)$. Условие (2) можно переформулировать так: величина $P(\Omega_x)/M(\Sigma_x)$ ограничена и отделена от 0 для всех x, являющихся началами последовательности ω. (Мы пишем $P(\Omega_x)$, но $M(\Sigma_x)$, так как распределение P задано на множестве Ω только бесконечных последовательностей, а распределение M — на множестве Σ конечных и бесконечных последовательностей.) Отметим, что частное $P(\Omega_x)/M(\Sigma_x)$ ограничено в силу максимальности меры M (аналогичным образом в случае монотонной энтропии выполнялось неравенство $KM(x) \leqslant -\log_2 P(\Omega_x) + O(1)$). Таким образом, надо доказать: ω типична тогда и только тогда, когда $P(\Omega_{(\omega)_n})/M(\Sigma_{(\omega)_n})$ отделено от нуля.

Доказательство. 1. Пусть $P(\Omega_{(\omega)_n})/M(\Sigma_{(\omega)_n})$ не отделено от нуля. Докажем, что ω не типична относительно распределения P. Рассмотрим для каждого рационального $\varepsilon > 0$ множество тех слов x, для которых $P(\Omega_x)/M(\Sigma_x) < \varepsilon$. Так как априорная полумера полувычислима снизу, а распределение P вычислимо, то множество таких x перечислимо. Обозначим его через A_ε. Легко проверить, что множество $S_\varepsilon = \bigcup \{\Omega_x \mid x \in A_\varepsilon\}$ имеет малую меру (относительно распределения P). Именно, $P(S_\varepsilon) \leqslant \varepsilon$. В самом деле, пусть $x_0, x_1, \ldots$ — все элементы A_ε, не являющиеся продолжениями других элементов A_ε. Тогда $S_\varepsilon = \bigcup \Omega_{x_i}$ и $P(S_\varepsilon) = \sum P(\Omega_{x_i}) \leqslant \varepsilon \cdot \sum M(\Sigma_{x_i}) \leqslant \varepsilon \cdot 1 = \varepsilon$, так как множества Σ_{x_i} не пересекаются. Если отношение $P(\Omega_{(\omega)_n})/M(\Sigma_{(\omega)_n})$ не отделено от 0, то последовательность ω лежит в любом из множеств S_ε, и, следовательно, не типична. (Здесь надо воспользоваться леммой 3 из § 4.3 подобно тому, как это делалось в § 4.1.)

2. Перейдем теперь к обратному утверждению. Ясно, что оно вытекает из соответствующего утверждения теоремы Левина — Шнорра. В самом деле, если последовательность ω не типична, то, согласно этой теореме, разность $-\log_2 P((\omega)_n) - KM((\omega)_n)$ не ограничена сверху, а $-\log_2 m((\omega)_n) \leqslant KM((\omega)_n) + O(1)$. Тем не менее дадим прямое доказательство.

Пусть последовательность ω не типична. Тогда для каждого $\varepsilon > 0$ существует вычислимая последовательность $X(\varepsilon, 0), X(\varepsilon, 1), \ldots$ двоичных слов, для которой $\omega \in \bigcup_i \Omega_{X(\varepsilon,i)}$ и $\sum P(\Omega_{X(\varepsilon,i)}) < \varepsilon$ при всех $\varepsilon > 0$. Рассмотрим теперь меру, которая вне множества $\bigcup_i \Omega_{X(\varepsilon,i)}$ будет нулевой,

а внутри него будет превосходить меру P в $1/\varepsilon$ раз. Более точно, рассмотрим меру P_ε на множестве Σ, для которой $P_\varepsilon(\Sigma_u) = (1/\varepsilon)\cdot P(\Omega_u \cap \bigcup_i \Omega_{X(\varepsilon,i)})$. (Строго говоря, это равенство не задает распределения вероятностей, так как определенная с его помощью мера всего Σ не равна 1, а лишь не превосходит 1. Чтобы исправить положение, положим $P_\varepsilon(\Sigma) = 1$; тем самым мы получим распределение вероятностей на $\Omega \bigcup \{\Lambda\}$, которое мы также будем обозначать P_ε.) Теперь возьмем достаточно быстро стремящуюся к нулю последовательность ε_s положительных рациональных чисел и сходящийся ряд $\sum k_s$ с не очень быстро убывающими членами, положив, к примеру, $\varepsilon_s = 1/2^s$ и $k_s = 1/s^2$. Рассмотрим теперь распределение $Q = c\cdot\sum k_s\cdot P_{\varepsilon_s}$, где константа c подобрана так, чтобы мера всего пространства равнялась 1. Мы докажем, что для нашей нетипической последовательности ω величина $P(\Omega_{(\omega)_n})/Q(\Sigma_{(\omega)_n})$ и тем более меньшая (с точностью до постоянного множителя) величина $P(\Omega_{(\omega)_n})/M(\Sigma_{(\omega)_n})$ не отделены от нуля. В самом деле, для любого s последовательность ω имеет начало, равное $X(\varepsilon_s, i)$ при некотором i (иначе $\omega \notin \bigcup \Omega_{X(\varepsilon_s,i)}$, что противоречит предположению). Обозначим это начало через x. Очевидно, $Q(\Sigma_x) \geqslant c\cdot k_s\cdot P_{\varepsilon_s}(\Sigma_x) = c\cdot k_s\cdot(1/\varepsilon_s)\,P(\Omega_x)$. Поэтому $P(\Omega_x)/Q(\Sigma_x) \leqslant \varepsilon_s/ck_s$. Но последовательность ε_s/ck_s стремится к 0 при $s \to \infty$. Теорема доказана.

По существу мы повторили рассуждение из доказательства теоремы Левина — Шнорра с некоторыми упрощениями, связанными с переходом от энтропии к априорной вероятности.

Обратим в заключение внимание на то, что наше доказательство устанавливает несколько больше того, что мы хотели: мы фактически доказали, что если ω не типична, то $P(\Omega_{(\omega)_n})/M(\Sigma_{(\omega)_n})$ стремится к 0. (В самом деле, в проведенной в конце доказательства оценке можно заменить слово x на любое его продолжение.) Таким образом, в формулировке теоремы можно вместо «ограничена» сказать «не стремится к бесконечности»». (Ср. с аналогичным замечанием в конце § 4.1 по поводу теоремы Левина — Шнорра.)

ГЛАВА 6

ЧАСТОТНЫЙ ПОДХОД К ОПРЕДЕЛЕНИЮ СЛУЧАЙНОСТИ

§ 6.1. Подход Мизеса. Уточнения Чёрча и Колмогорова — Лавлэнда

Основоположником частотного подхода является немецкий математик и механик Рихард фон Мизес, предложивший его в [Миз 19], [Миз 28]. Нужно иметь в виду, однако, что Мизес предлагал считать понятие случайной последовательности (в терминологии Мизеса — «коллектива»), основным понятием теории вероятностей. С нашей сегодняшней точки зрения понятие случайной последовательности дополняет классическую, «теоретико-мерную» теорию вероятностей, так что первичным является понятие распределения вероятностей, а лишь потом мы определяем понятие случайного — относительно этого распределения — объекта. Напротив, для Мизеса первичным было понятие коллектива, а распределение вероятностей было характеристикой коллектива. Кроме того, подход Мизеса не был чисто математическим — с точки зрения современной теоретико-множественной математики. Это проявляется в том, что его определение коллектива использует неформально описанное понятие «допустимого правила выбора», а существование коллективов обосновывается — среди прочего — ссылкой на существование игорных домов.

Предупредив об этом, попытаемся изложить точку зрения Мизеса. Пусть дана последовательность нулей и единиц, получаемая в результате бросания симметричной монеты (нуль — решетка, единица — герб). Эмпи-

рический факт состоит в том, что доля единиц в начальном отрезке длины N этой последовательности стремится к $1/2$ при $N \to \infty$. Этот факт можно интерпретировать и так: в игре в орлянку с симметричной монетой, в которой мы при выпадении герба получаем копейку, а при выпадении решетки — теряем, в среднем мы ничего не выигрываем и не проигрываем.

Более того, опыт игорных домов показывает, что никакая «система игры», говорящая нам, в каких случаях делать ставку, а когда пропускать игру, не позволяет достичь выигрыша. Такую систему игры можно рассматривать как «правило выбора», согласно которому из последовательности выбирается подпоследовательность тех значений, на которые делаются ставки. Для случайной последовательности предел доли единиц в выбранной согласно такому правилу подпоследовательности также равен $1/2$. При этом, разумеется, допустимы не все правила («системы игры»); вот пример очевидно недопустимого правила: «выбрать те члены, которые равны 1» («ставить в тех случаях, когда выпадет герб»).

Аналогичные свойства выполнены и для несимметричной монеты: в этом случае доля единиц стремится к некоторому p ($0 < p < 1$), и это же верно для любой подпоследовательности, выбранной с помощью допустимого правила. Число p и называется *вероятностью* выпадения герба.

Итак, можно считать, что общий замысел фон Мизеса таков. Рассматриваются так называемые правила выбора, каждое из которых позволяет перейти от произвольной последовательности к некоторой ее (конечной или бесконечной) подпоследовательности. Некоторые правила выбора объявляются *допустимыми*. Всякая подпоследовательность, полученная допустимым правилом выбора из данной последовательности, объявляется *законной* подпоследовательностью (причем сама последовательность также считается своей законной подпоследовательностью). Бесконечная последовательность нулей и единиц является случайной (по Мизесу) для бернуллиева распределения с вероятностью единицы, равной p, при выполнении следующего условия: для любой бесконечной законной подпоследовательности предел доли единиц в ее начальных отрезках существует и равен p.

Мы изложили в самых общих чертах точку зрения Мизеса. Позднее А. Чёрч [Чёрч 40] предложил уточнить понятие допустимого правила выбора следующим образом. Прежде всего, решение о том, делать ли ставку, должно приниматься на основе результатов уже сделанных бросаний. Другими словами, допустимое правило выбора представляет собой множество A, элементами которого являются двоичные слова; это правило выбирает из последовательности $x_0x_1 \ldots$ подпоследовательность из тех членов x_n, для которых $x_0x_1 \ldots x_{n-1} \in A$ (сохраняя тот порядок, в котором они шли в исходной последовательности).

Но этого требования мало: для любой последовательности $x_0x_1 \ldots$ можно рассмотреть множество A, содержащее все такие слова $x_0 \ldots x_{n-1}$, для которых $x_n = 1$. Соответствующее правило, очевидно, выберет подпоследовательность из одних единиц. Мизес, вероятно, возразил бы на это, что это правило недопустимо, так как допустимое правило должно быть фиксировано до начала игры. Однако неясно, как это требование может быть сформулировано математически, и Чёрч заменил его требованием разрешимости множества A — требованием существования алгоритма, который по любому двоичному слову определяет, принадлежит ли оно множеству A.

Впоследствии Колмогоров в [Колм 63] и независимо Лавлэнд в [Лав 66а, с. 499] обобщили понятие допустимого правила выбора. Это допущение связано с изменением схемы игры, разрешающим игроку самому выбирать порядок просмотра членов последовательности. Представим себе, что члены последовательности написаны на лицевых сторонах карточек, выложенных перед игроком вверх изнанкой. Эти карточки переворачивают одну за другой. При этом игрок имеет право указать на произвольную еще не перевернутую карточку и попросить ее перевернуть (не делая ставки), а также сде-

лать ставку на любую еще не перевернутую карточку — в этом случае он выигрывает копейку, если на ней окажется единица, и проигрывает копейку, если окажется нуль. При такой схеме игры в подпоследовательность включаются те карточки, на которые была сделана ставка (в том порядке, в котором они переворачивались). Отметим, что самоё понятие подпоследовательности здесь несколько расширяется по сравнению с традиционным: порядок следования членов в подпоследовательности может быть иным, чем в исходной последовательности.

Говоря формально, допустимое по Колмогорову — Лавлэнду правило выбора задается вычислимыми функциями F и G, аргументами которых являются двоичные слова, значениями F являются натуральные числа, а значениями G — истина и ложь. (Функция F определяет, какую карточку надо переворачивать следующей, а функция G — надо ли делать ставку.) Заметим, что функции F и G могут быть частичными. Опишем применение правила выбора, задаваемого функциями F и G, к последовательности $x_0x_1\ldots$ Вначале мы определяем последовательность натуральных чисел $n_0, n_1, \ldots$ так: $n_0 = F(\Lambda)$ (Λ — пустая последовательность), $n_1 = F(x_{n_0})$, $n_2 = F(x_{n_0}x_{n_1})$ и т. д.; построение кончается, как только хотя бы одно из значений $F(x_{n_0}x_{n_1}\ldots x_{n_k})$ и $G(x_{n_0}x_{n_1}\ldots x_{n_k})$ оказывается неопределенным или если $F(x_{n_0}x_{n_1}\ldots x_{n_k})$ встречается среди $n_0, n_1, \ldots, n_k$. Затем среди чисел n_k отбираются те, для которых $G(x_{n_0}x_{n_1}\ldots x_{n_{k-1}})$ истинно; соответствующие им x_{n_k} и образуют выбранную подпоследовательность (в порядке возрастания k).

Пусть p — произвольное (не обязательно вычислимое) число из интервала $(0, 1)$. Последовательность ω называется *стохастической по Чёрчу* (или *случайной по Мизесу — Чёрчу*) относительно бернуллиева распределения вероятностей с вероятностью единицы, равной p, если частота единиц в ее начальных отрезках стремится к p и это же верно для всех бесконечных последовательностей, получаемых из ω с помощью допустимого по Чёрчу правила выбора. Аналогичным образом, заменяя допустимость по Чёрчу на допустимость по Колмогорову — Лавлэнду, получаем определение *стохастической по Колмогорову — Лавлэнду* (или *случайной по Мизесу — Колмогорову — Лавлэнду*) последовательности.

Свойства определенных таким образом понятий случайности обсуждаются в следующих параграфах этой главы; этот параграф мы закончим следующим простым наблюдением, относящимся к случаю симметричной монеты. Расширим права игрока, разрешив ставить не только на единицы, но и на нули. Это значит, что перед очередным бросанием монеты (в схеме Чёрча) или переворачиванием карточки (в схеме Колмогорова — Лавлэнда) игрок имеет право сделать одно из трех заявлений: «ставлю на 0», «ставлю на 1» и «пас». В последнем случае он остается при своих, в первых двух он либо выигрывает копейку (если угадал), либо проигрывает (если не угадал). Последовательность будет называться *стохастической*, если средний выигрыш игрока (т. е. выигрыш, деленный на число ставок) стремится к 0 для любой системы игры.

Легко проверить, что такое расширение прав игрока не меняет класса стохастических (по Чёрчу или по Колмогорову — Лавлэнду) последовательностей. В самом деле, пусть есть система игры S в новом смысле (разрешено ставить на 0 и на 1). Рассмотрим две системы игры S_0, S_1 в старом смысле: система S_0 получается, если делать ставки в тех случаях, когда S предписывает ставить на 0, а S_1 получается, если делать ставки в тех случаях, когда S предписывает ставить на 1. Ясно, что выигрыш стратегии S равен разности между выигрышами стратегий S_0 и S_1, откуда легко следует высказанное утверждение.

В о п р о с. Дана произвольная последовательность, стохастическая по Колмогорову — Лавлэнду. Применим к ней правило, допустимое по

Колмогорову — Лавлэнду. Будет ли полученная последовательность стохастична по Колмогорову — Лавлэнду? (Обсуждение этого вопроса можно найти в [Шень 82].)

§ 6.2. Соотношения между различными определениями. Конструкция Вилля. Теорема Мучника. Пример Ламбальгена

6.2.1. Соотношения между различными вариантами стохастичности. Дав определения различных вариантов стохастичности в рамках частотного подхода, естественно поставить вопрос об их соотношениях с типичностью (и, тем самым, с хаотичностью). Эти сравнения имеют смысл, естественно, лишь для бернуллиевых мер (иначе не определено понятие стохастичности), у которых вероятность появления единицы p является вычислимым действительным числом (иначе не определены понятия типичности и хаотичности). В этом случае справедлива

Т е о р е м а. (а) *Всякая типическая последовательность является стохастической по Колмогорову — Лавлэнду*;

(б) *всякая стохастическая по Колмогорову — Лавлэнду последовательность стохастична по Чёрчу*;

(в) *обратные утверждения к* (а) *и* (б) *неверны.*

В этом параграфе мы изложим схему доказательства утверждений (а)—(в), а также приведем некоторые конструкции и результаты, имеющие, на наш взгляд, самостоятельный интерес. Для простоты мы ограничимся случаем равномерного бернуллиева распределения ($p = 1/2$).

Утверждение (б) непосредственно вытекает из определения. Чтобы доказать утверждение (а), вспомним обсуждение закона больших чисел в § 2.1. Там — для фиксированного рационального $\varepsilon > 0$ — мы рассматривали множества D_n тех последовательностей, у которых частота единиц в начальном отрезке длины n отличается от $1/2$ более чем на ε. Стандартная оценка (с помощью формулы Стирлинга, теоремы Муавра — Лапласа и т. п.) показывает, что равномерная бернуллиева мера множеств D_n экспоненциально убывает с ростом n (при фиксированном ε). Каждое из множеств D_n легко представить в виде объединения конечного числа непересекающихся интервалов (соответствующих словам длины n, в которых доля единиц отличается от $1/2$ более чем на ε), а множество $E_k = \bigcup\limits_{n \geqslant k} D_n$ — в виде объединения вычислимой последовательности интервалов; суммарная их мера будет сколь угодно мала, если k достаточно велико. Поэтому множество $\bigcap\limits_k E_k$ — эффективно нулевое и любая принадлежащая ему последовательность нетипична. Поскольку любая последовательность, у которой предел частот не существует или не равен $1/2$, попадает в $\bigcap\limits_k E_k$ (при каком-то ε), то у любой типической последовательности предел частот равен $1/2$.

Нужно еще установить, что для любого допустимого по Колмогорову — Лавлэнду правила выбора R результат его применения к любой типической последовательности либо конечен, либо имеет предел частот $1/2$. Для этого нужно рассмотреть (при фиксированном ε) множества D_n^R тех последовательностей, применение к которым правила R дает последовательность не менее чем из n членов и частота единиц среди ее первых n членов отличается от $1/2$ более чем на ε. Мера этого множества не превосходит меры множества D_n (она может быть меньше за счет того, что результатом применения правила может быть последовательность длины меньше n). Кроме того, множество D_n^R можно представить в виде объединения перечислимого семейства интервалов (если $\omega \in D_n^R$, то это можно обнаружить по конечному начальному отрезку последовательности ω). Далее рассуждения повторяют изложенные выше.

Замечание. Здесь существенно то, что мы пользуемся определением типичности по Мартин-Лёфу, а не по Шнорру, поскольку мы не приводим оценки скорости сходимости ряда из мер интервалов, составляющих D_n^R; не всякая типическая по Шнорру последовательность стохастична по Колмогорову — Лавлэнду, поскольку типическая по Шнорру последовательность с логарифмическим ростом энтропии начальных отрезков, упоминавшаяся в п. 2.3.1, не может быть стохастичной по Колмогорову — Лавлэнду, как следует из формулируемой ниже теоремы Мучника.

6.2.2. Пример Вилля. Переходя к доказательству утверждения (в), начнем с изложения конструкции, восходящей к Виллю (J. Ville), который показал, что существует стохастическая по Чёрчу (относительно равномерной бернуллиевой меры) последовательность, в которой любой начальный отрезок содержит не меньше нулей, чем единиц. (На самом деле конструкция Вилля не использует разрешимости множеств, задающих допустимые по Чёрчу правила выбора, и была построена еще до определения Чёрча. В ее изложении мы следуем [Лавл 66].) Множество последовательностей, у которых любой начальный отрезок содержит не меньше нулей, чем единиц, является эффективно нулевым. (То, что оно нулевое, следует из т. н. «закона повторного логарифма»; анализ доказательства показывает, что оно является эффективно нулевым.) Поэтому построенная Виллем последовательность не является типической. Тем самым пример Вилля показывает, что существует стохастическая по Чёрчу последовательность, не являющаяся типической, т. е. что по крайней мере одно из утверждений, обратных к (а) и (б), неверно.

Итак, изложим конструкцию Вилля. Нам нужно, чтобы после применения любого из допустимых по Чёрчу правил выбора получалась «сбалансированная» последовательность (последовательность, в которой частота единиц стремится к 1/2). Для ясности начнем с модельного примера: пусть допустимое правило выбора только одно.

Будем строить искомую последовательность постепенно. Пусть уже построен некоторый начальный отрезок $x_0 \ldots x_k$. Обратимся к правилу выбора: оно определяет, будет ли включен следующий член в выбираемую подпоследовательность. Этот член мы выберем равным нулю или единице в зависимости от того, четно или нечетно число уже включенных (если он включается правилом) или невключенных (если он не включается) к рассматриваемому моменту членов. Таким образом, выбранная подпоследовательность будет иметь вид 010101...— так же как и подпоследовательность из оставшихся невыбранными членов. При этом любой начальный отрезок построенной последовательности представляет собой «смесь» двух начальных отрезков последовательностей вида 010101...— и, следовательно, содержит не меньше нулей, чем единиц.

Пусть теперь имеется конечное число правил выбора $R_1, \ldots, R_l$. Тогда каждый новый член характеризуется двоичным вектором длины l, определяющим, какие из правил $R_1, \ldots, R_l$ захотят включить его в выбираемую ими подпоследовательность. Наша последовательность разбивается, таким образом, на 2^l подпоследовательностей, каждая из которых соответствует одному из 2^l значений упомянутого двоичного вектора. Мы будем выбирать очередные члены так, чтобы в каждой из этих 2^l подпоследовательностей нули и единицы чередовались. Тогда любой начальный отрезок будет содержать не меньше нулей, чем единиц. Применяя i-е правило выбора, мы получаем последовательность, представляющую собой «смесь» 2^{l-1} последовательностей вида 010101... (тех, для которых соответствующий двоичный вектор имеет единицу на i-м месте). Эта смесь, очевидно, сбалансирована.

Перейдем теперь к случаю счетного числа правил выбора. В этом случае каждому очередному члену последовательности соответствует уже не двоичный вектор, а бесконечная последовательность $u_0u_1 \ldots$, в которой $u_i = 1$, если i-е правило включает рассматриваемый член в подпоследователь-

ность, и $u_i = 0$ в противном случае. Другими словами, каждому члену соответствует некоторый путь в двоичном дереве. Мы, однако, будем вводить в рассмотрение не все правила сразу, а по одному, т. е. использовать не весь путь, а лишь некоторый его начальный отрезок. Более точно. Выберем и зафиксируем достаточно быстро растущую последовательность $n_0 < n_1 < \ldots$ натуральных чисел. (Пусть, например, $n_i = 2^{2^i}$.) На каждом шаге построения одна из вершин дерева будет активной. Искать активную вершину мы будем, начав с корня дерева и двигаясь вдоль пути, соответствующего уже построенному участку последовательности. При этом мы остановимся, как только попадем в вершину, которая была активной менее n_i раз (где i — ее высота). Другими словами, с каждым членом строимой двоичной последовательности мы сопоставим двоичное слово x наименьшей длины, удовлетворяющее таким условиям: (1) m-я буква x равна 1 или 0 в зависимости от того, выбирает или не выбирает m-е правило рассматриваемый член; (2) слово x встречалось менее $n_{l(x)}$ раз среди слов, сопоставленных с предыдущими членами. Таким образом, вся последовательность разбивается на счетное число конечных подпоследовательностей: каждому двоичному слову длины i (каждой вершине двоичного дерева уровня i) соответствует подпоследовательность не более чем из n_i членов; их номера — это номера тех шагов построения, на которых вершина x была активной. При этом подпоследовательность, соответствующая слову (вершине) x, начинается лишь после того, как все подпоследовательности, соответствующие началам x, закончились.

Осталось описать, как выбираются члены искомой стохастической по Чёрчу последовательности. Мы будем строить ее так, чтобы в каждой подпоследовательности, соответствующей произвольному двоичному слову x, нули и единицы чередовались, начиная с нуля. Такое построение возможно, так как к моменту выбора очередного члена уже известно, в какую подпоследовательность он попадет. В результате этого любой начальный отрезок последовательности будет содержать не меньше нулей, чем единиц. Осталось показать, что любое правило выбора выберет конечную или сбалансированную последовательность.

Пусть $y_0y_1\ldots$ — бесконечная последовательность, полученная применением i-го правила выбора. Рассмотрим произвольный начальный отрезок этой подпоследовательности и слова (вершины дерева), сопоставленные входящим в него членам. Эти слова могут быть не длиннее i (число таких членов ограничено — не превосходит $2^0n_0 + \ldots + 2^in_i$ — и мы можем их не рассматривать), либо быть длиннее i. Во втором случае на i-м месте этих слов стоит 1, так как i-е правило включило соответствующие члены в подпоследовательность. Рассмотрим самое длинное слово x из числа сопоставленных с членами выбранного начального отрезка. Пусть его длина равна k. (Как уже говорилось, можно считать, что $k > i$.) Тогда общее число использованных слов не превосходит $1 + 2 + \ldots + 2^k < 2^{k+1}$. Так как подпоследовательности, соответствующие каждому из них, содержат либо поровну нулей и единиц, либо на один нуль больше, то разница между числом нулей и единиц в выбранном начальном отрезке не превосходит 2^{k+1}. Оценим теперь длину нашего отрезка — установим, что она велика. В самом деле, раз мы использовали слово x, то, следовательно, слово x', получающееся из x отбрасыванием последнего символа, использовано полностью — оно встретилось n_{k-1} раз. На i-м месте слова x' стоит единица (мы считаем, что $k > i$), и, следовательно, все члены, соответствующие слову x', включены в рассматриваемую подпоследовательность. Таким образом, ее длина не менее $n_{k-1} = 2^{2^{k-2}}$, и отклонение частоты единиц от 1/2 не превосходит $(2^{k+1})/(2^{2^{k-2}})$ и стремится к нулю.

Итак, мы построили стохастическую по Чёрчу последовательность, любой начальный отрезок которой содержит не меньше нулей, чем единиц. (Приписав к ней вначале 0, можно построить стохастическую по Чёрчу по-

следовательность, в которой каждый начальный отрезок содержит больше нулей, чем единиц.) Отметим, что это построение никак не использует алгоритмическую природу рассматриваемых правил выбора — важно лишь, что их счетное число. Кроме того, важна схема выбора — в частности то, что все правила просматривают последовательность в одном и том же порядке. Поэтому наше построение неприменимо к схеме выбора, предложенной Колмогоровым — Лавлэндом.

Анализ конструкции Вилля позволяет установить, что существуют стохастические по Чёрчу последовательности, энтропия начальных отрезков которых длины n имеет порядок $O(\log_2 n)$. В самом деле, изложенная конструкция позволяет для любого счетного семейства правил $R_1, R_2, \ldots$ построить последовательность ω, сбалансированную относительно всех правил семейства. При этом, если задан алгоритм, перечисляющий эти правила (т. е. дающий по i программу, соответствующую i-му правилу), то последовательность ω получится вычислимой. Если бы существовал алгоритм, перечисляющий все допустимые по Чёрчу правила выбора, то получилась бы вычислимая стохастическая по Чёрчу последовательность, что, конечно, невозможно. Так что построение стохастической по Чёрчу последовательности не может быть алгоритмическим и должно использовать дополнительную информацию о том, какие программы задают допустимые по Чёрчу правила выбора (т. е. разрешимые множества). Однако объем этой информации может быть невелик по сравнению с длиной строимого участка последовательности, если числа n_s (см. конструкцию) растут достаточно быстро с ростом s. Таким способом можно построить стохастическую по Чёрчу последовательность, энтропия начальных отрезков которой длины n растет логарифмически (не превосходит $c \cdot \log_2 n$ для некоторой константы c); тем самым мы снова приходим к примеру стохастической по Чёрчу последовательности, не являющейся хаотической и типической.

6.2.3. Теорема Мучника. Существование стохастической по Чёрчу последовательности с логарифмическим ростом энтропии было отмечено (без доказательства) в [Колм 69]. Там же было сформулировано утверждение о том, что существуют стохастические по Колмогорову — Лавлэнду последовательности с логарифмическим ростом энтропии. Как недавно установил Ан. А. Мучник, это утверждение неверно. Именно, справедлива следующая

Теорема (Ан. А. Мучник). *Если последовательность ω такова, что энтропия начального отрезка длины n не превосходит αn для некоторого $\alpha < 1$ и для всех достаточно больших n, то последовательность ω не стохастична по Колмогорову — Лавлэнду.*

Опишем — по согласованию с автором — схему доказательства этой теоремы (публикуется впервые). Пусть n — натуральное число, A — некоторое множество двоичных слов длины n. Рассмотрим такую игру. Противник выбирает произвольную последовательность нулей и единиц длины n, принадлежащую множеству A. Он записывает ее члены на лицевой стороне карточек, выложенных вверх изнанкой. Игрок переворачивает карточки по очереди, слева направо. Перед переворачиванием каждой он имеет право поставить любую сумму не более 1 копейки на нуль или единицу. Если он угадал, то он выигрывает указанную сумму, если нет — теряет. (Эта схема игры отличается от схемы Чёрча, помимо фиксированного числа партий, возможностью непрерывного изменения ставок.)

Лемма 1. *Пусть $\alpha < 1$. Тогда для каждого натурального n и для каждого множества $A \subset \{0, 1\}^n$, содержащего не более $2^{\alpha n}$ элементов, можно указать стратегию, гарантирующую Игроку в описанной игре выигрыш не менее $(1 - \alpha) \cdot n$.*

Доказательство леммы. Введем понятие *информационного капитала* Игрока (на данный момент игры). Пусть $x_0 \ldots x_{k-1}$ — значения на перевернутых к данному моменту игры карточках; M — общее количество продолжений последовательности $x_0 \ldots x_{k-1}$ ($M = 2^{n-k}$), m — коли-

чество продолжений последовательности $x_0 \ldots x_{k-1}$, принадлежащих A. Информационный капитал Игрока определим как $\log_2 (M/m)$. В начале игры он равен $(1 - \alpha)\, n$, в конце игры равен 0. Покажем, что существует стратегия Игрока, для которой можно гарантировать, что в процессе игры сумма выигрыша Игрока и его информационного капитала не убывает. Тем самым в конце игры — когда информационный капитал равен 0 — Игрок выиграет не менее $(1 - \alpha)n$.

Чтобы убедиться в существовании такой стратегии, для каждого момента игры, помимо определенных выше чисел M и m, рассмотрим также числа m_0 — количество продолжений последовательности $x_0 \ldots x_{k-1}0$, принадлежащих множеству A, и m_1 — количество продолжений последовательности $x_0 \ldots x_{k-1}1$, принадлежащих множеству A. (Очевидно, $m = m_0 + m_1$.) Числа $p_0 = m_0/m$ и $p_1 = m_1/m$ можно рассматривать как условные вероятности появления нуля и единицы после $x_0 \ldots x_{k-1}$, если считать все элементы множества A одинаково вероятными a priori; $p_0 + p_1 = 1$. Нам необходимо выбрать размер ставки на нуль из отрезка $[-1, 1]$ (отрицательные ставки соответствуют ставкам на единицу) таким, чтобы при любом исходе сумма выигрыша Игрока и его информационного капитала не уменьшилась.

Пусть x — размер ставки на нуль. В случае появления нуля к информационному капиталу добавится $-\log_2 p_0 - 1$, а к сумме добавится $x - \log_2 p_0 - 1$; в случае появления единицы к сумме добавится $-x - \log_2 (1 - p_0) - 1$. Тем самым осталось проверить, что для каждого $p \in [0, 1]$ существует такое число $x \in [-1, 1]$, что оба числа $x - \log_2 p - 1$ и $-x - \log_2 (1 - p) - 1$ неотрицательны. Поскольку для произвольных a и b существование такого $x \in [-1, 1]$, что $x - a \geqslant 0$ и $-x - b \geqslant 0$, равносильно тому, что $a \leqslant -b$ и отрезок $[a, -b]$ пересекается с $[-1, 1]$, достаточно проверить, что $\log_2 p + 1 \leqslant -\log_2 (1 - p) - 1$, т. е. что $(\log_2 p + \log_2 (1 - p))/2 \leqslant -1$ (это следует из выпуклости логарифма) и что хотя бы одно из чисел $\log_2 p + 1$ и $-\log_2 (1 - p) - 1$ принадлежит отрезку $[-1, 1]$ (первое, если $p \geqslant 1/2$; второе, если $p \leqslant 1/2$).

В следующей лемме мы переходим к стратегиям, всегда делающим ставки в максимальном размере (ставящим копейку либо на 0, либо на 1). Это ограничение будет компенсировано тем, что будет указана не одна, а несколько стратегий.

Лемма 2. *Пусть $\alpha < 1$. Тогда для каждого n и для каждого множества $A \subset \{0, 1\}^n$, содержащего не более $2^{\alpha n}$ элементов, можно указать конечный набор $S_1, \ldots, S_t$ стратегий, делающих ставки в максимальном размере, обладающий таким свойством: для всякой последовательности $x \in A$ существует стратегия S_i этого набора, обеспечивающая выигрыш на x не менее $(1 - \alpha/2)/n$. При этом количество стратегий в наборе зависит только от α (но не от n и A).*

Доказательство леммы. Рассмотрим стратегию S (с произвольными ставками), существующую по лемме 1. Рассмотрим стратегию S' со ставками, кратными $1/N$ для некоторого достаточно большого N, приближающую стратегию S с точностью $1/N$ (в любой ситуации ставки отличаются не более чем на $1/N$). Если N достаточно велико $(1/N < (1 - \alpha)/2)$, то стратегия S' будет гарантировать выигрыш не меньше $(1-\alpha/2)/n$. Теперь представим S' в виде «усреднения» $2N$ стратегий, делающих ставки в максимальном размере. Пусть, например, S' предлагает сделать ставку на нуль в размере m/N. В этом случае некоторые из стратегий $S_1, \ldots, S_t$ будут делать ставку на 0, а некоторые на 1, причем число первых будет равно $N + m$, а число последних равно $N - m$. При этом для любой последовательности x будет справедливо такое соотношение: выигрыш S' на x будет равен среднему арифметическому выигрышей $S_1, \ldots, S_{2N}$ на x. Поэтому хотя бы одна из стратегий $S_1, \ldots, S_{2N}$ будет давать выигрыш не менее $(1-\alpha/2)/n$. Лемма 2 доказана.

Перейдем теперь к изложению схемы доказательства теоремы Мучника. Пусть ω — последовательность, энтропия начальных отрезков которой длины n не превосходит αn (при любом достаточно большом n). Разрежем ее на куски $u_0, u_1, \ldots$, длины которых $n_0, n_1, \ldots$ быстро возрастают ($\omega = u_0u_1, \ldots$, $l(u_i) = n_i$). Если возрастание достаточно быстрое (положим, например, $n_i = 2^{2^i}$), то энтропия u_i меньше $\lfloor\beta \cdot l(u_i)\rfloor$ ($\lfloor z\rfloor$ — целая часть числа z) для некоторой рациональной константы $\beta < 1$ и для всех достаточно больших i (будем считать для простоты, что для всех — конечным началом можно пренебречь). Тем самым u_i принадлежит множеству A_i всех слов длины n_i, имеющих энтропию менее $\lfloor\beta n_i\rfloor$; количество элементов в этом множестве не превосходит $2^{\beta n_i}$. Применяя лемму 2, можно для каждого i указать набор стратегий $S_1, \ldots, S_t$, при этом для любого элемента $u \in A_i$ — в том числе и для u_i — хотя бы одна из стратегий S_j гарантирует выигрыш не менее $(1-\beta/2)/n_i$. Количество стратегий (t) не зависит от i, поэтому из них можно сформировать t стратегий игры с бесконечной последовательностью. Одна из этих стратегий при игре с последовательностью ω будет приводить к выигрышу, который в бесконечном числе моментов времени превосходит $\varepsilon\cdot$(число партий) для некоторого $\varepsilon > 0$. Чтобы убедиться в этом, достаточно заметить, что выигрыш на i-м участке существенно превосходит любой возможный проигрыш на предыдущих участках, а также, что если на каждом участке хоть одна стратегия успешна, то существует стратегия, успешная на бесконечном числе участков.

Приведенное рассуждение позволило бы установить, что последовательность ω не стохастична по Чёрчу, если бы построенная стратегия была вычислимой (напомним, что разрешение делать ставки и на нуль, и на единицу обсуждалось в конце § 6.1). Однако таковой она не является, поскольку невозможно эффективно по i получить список всех слов длины n_i, имеющих энтропию менее $\lfloor\beta n_i\rfloor$: эти слова можно перечислять (если слово имеет короткое описание, то это рано или поздно обнаружится), но ни в какой момент нельзя гарантировать, что обнаружены все слова.

Выход из положения, предложенный Ан. А. Мучником, состоит в следующем: надо сгруппировать участки последовательности в пары u_0u_1, $u_2u_3, \ldots$ Пусть известно, что слово u_{2n} принадлежит множеству A_{2n}, а слово u_{2n+1} принадлежит множеству A_{2n+1}. У нас есть следующие две возможности игры (по схеме Колмогорова — Лавлэнда) с последовательностью $u_{2n}u_{2n+1}$. Первая из них состоит в следующем. Не делая ставок, мы переворачиваем карточки и узнаем слово u_{2n}. Затем перечисляем множество A_{2n} до тех пор, пока не обнаружим слово u_{2n}. После этого делаем столько же шагов перечисления множества A_{2n+1}; полученную за это время часть множества A_{2n+1} обозначаем $\bar{A}_{2n+1}$ и используем вместо A_{2n+1} при построении стратегии, работающей с участком u_{2n+1}. Вторая возможность состоит в том, чтобы поступить наоборот: не делая ставок, узнать слово u_{2n+1}, дождаться его появления в перечислении множества A_{2n+1}, сделать столько же шагов перечисления множества A_{2n}, обнаружившуюся часть $\bar{A}_{2n}$ множества A_{2n} использовать при построении стратегии, работающей с участком u_{2n}. При этом хотя бы в одном из двух случаев (замена A_{2n} на $\bar{A}_{2n}$ и замена A_{2n+1} на $\bar{A}_{2n+1}$) замена окажется допустимой (если u_{2n} появится в перечислении A_{2n} раньше, чем u_{2n+1} в перечислении A_{2n+1}, то во втором случае, иначе — в первом).

Поскольку у нас имеется не одна стратегия для каждого участка, а набор из t стратегий, то этот прием приведет к $2t$ вычислимым стратегиям игры с бесконечной последовательностью, обозначаемым S_{pr} ($1 \leqslant p \leqslant t$, $r = 0$ или 1). Стратегия S_{p0} просматривает (без ставок) $u_0, u_2, u_4, \ldots$ и после каждого просмотра использует стратегию с номером p для игры с участками $u_1, u_3, u_5, \ldots$ соответственно; эта стратегия строится с использованием множеств $\bar{A}_1, \bar{A}_3, \bar{A}_5, \ldots$; напротив, стратегия S_{p1} просматривает (без ставок)

u_1, u_3, u_5, . . . и использует стратегию с номером p для игры с участками u_0, u_2, u_4, . . ., получаемую на основе множеств $\bar{A}_0$, $\bar{A}_2$, $\bar{A}_4$, . . .

Теперь уже можно утверждать, что все стратегии S_{Ir} являются вычислимыми. Для каждого n либо u_{2n} появляется раньше в перечислении A_{2n}, чем u_{2n+1} — в перечислении A_{2n+1}, либо наоборот. В первом случае стратегия S_{I1} (при некотором p), а во втором — стратегия S_{I0} (при некотором p) окажется успешной. Поскольку стратегий конечное число, то одна из них окажется успешной для бесконечного числа участков и, следовательно, последовательность ω не является стохастической по Колмогорову — Лавлэнду. Теорема Мучника доказана.

6.2.4. Пример Ламбальгена. Покажем, что утверждение, обратное к утверждению (а) п. 6.2.1, также неверно. Это можно сделать, используя метод Ламбальгена [Лам 87], [Лам 87а]. В этой работе доказано существование стохастической по Чёрчу последовательности, не являющейся типической. Как замечено в [Шень 88], эта же конструкция позволяет доказать существование стохастической по Колмогорову — Лавлэнду последовательности, не являющейся типической. Мы не приводим здесь деталей построения, но идея его очень проста. Рассмотрим вычислимую последовательность рациональных чисел p_0, p_1, . . ., сходящуюся к 1/2, и распределение вероятностей μ на Ω, соответствующее независимым испытаниям, вероятность появления 1 в i-м из которых равна p_i. Оказывается, что в этом случае всякая типическая относительно μ последовательность будет стохастична по Колмогорову — Лавлэнду относительно равномерной меры. Однако, если сходимость достаточно медленна ($\Sigma\ (p_i - 1/2)^2 = \infty$), то никакая типическая относительно μ последовательность не будет типической относительно равномерной бернуллиевой меры. Осталось положить, например, $p_i = (i + 10)^{-1/2} + 1/2$. (Подробности см. в [Шень 88].)

Осталось показать, что утверждение, обратное к утверждению (б) п. 6.2.1, также неверно: существуют стохастические по Чёрчу, но не стохастические по Колмогорову — Лавлэнду последовательности. Пример такой последовательности был построен Лавлэндом [Лав 66]. Он построил стохастическую по Чёрчу последовательность, становящуюся не стохастической (по Чёрчу) после вычислимой перестановки членов. Легко понять, что эта последовательность не может быть стохастической по Колмогорову — Лавлэнду.

Мы не приводим построения Лавлэнда, поскольку из теоремы Мучника вытекает, что стохастическая по Чёрчу последовательность с логарифмическим ростом энтропии начальных отрезков, о которой шла речь выше, не является стохастической по Колмогорову — Лавлэнду.

§ 6.3. Критерий типичности в терминах игр

В этом пункте мы покажем, что критерий типичности в терминах априорной вероятности, приведенный в § 5.4, может быть — для случая равномерной меры — интерпретирован в терминах некоторой игры, являющейся обобщением рассмотренных выше схем игр. Мы уже рассматривали игру, в которой ставки могут меняться от нуля до фиксированного максимального значения. Изменим теперь правила игры, считая, что размер ставки ограничен наличным капиталом Игрока. Более точно, будем считать, что имеющийся в распоряжении Игрока капитал он может разделить на три части: часть поставить на 0, часть поставить на 1, а часть выбросить. (Последнее кажется очевидно невыгодным для Игрока, но необходимо по причинам, которые станут ясны в дальнейшем.) Поставленная на верную цифру часть удваивается, а поставленная на неверную — так же как и выброшенная, пропадает. (Тем самым, поделив свой капитал поровну между нулем и единицей, Игрок в любом случае остается «при своих».) Начальный капитал Игрока равен 1. Просмотр членов последовательности, как и для схемы Чёрча, осуществляется в порядке их номеров.

Для игры описанного типа стратегия представляет собой правило, предписывающее так или иначе делить капитал на три части в зависимости от известных членов последовательности. Такая стратегия задается функцией $L\colon \Xi \to \mathbb{R}$, значением которой на двоичном слове x является величина капитала, который будет у Игрока после игры с начальным отрезком последовательности, равным x. (В начале игры у Игрока была копейка.) В терминах функции L стратегия описывается так: капитал (равный $L(x)$) делится на три части: $L(x1)/2$, $L(x0)/2$ и остаток, равный $L(x) - L(x1)/2 - L(x0)/2$. (Первая ставится на 1, вторая — на 0, а третья выкидывается.)

При этом следует наложить очевидные ограничения на функцию L: $L(x) \geqslant 0$ при всех x, $L(\Lambda) = 1$, $L(x0) + L(x1) \leqslant 2L(x)$. Функции, обладающие такими свойствами, находятся во взаимно-однозначном соответствии со стратегиями в игре описанного типа.

С другой стороны, функции с указанными свойствами находятся во взаимно-однозначном соответствии с полумерами (в смысле § 5.1): функции L соответствует полумера $x \mapsto L(x)/2^{l(x)}$. При этом полувычислимым снизу полумерам соответствуют полувычислимые снизу функции. Это позволяет переформулировать критерий типичности из § 5.4 следующим образом. Назовем стратегию *полувычислимой снизу*, если соответствующая ей функция L полувычислима снизу. Последовательность ω типична тогда и только тогда, когда не существует полувычислимой стратегии, выигрыш которой при игре с последовательностью ω неограничен. (В этом критерии «неограничен» можно заменить на «стремится к бесконечности»; см. обсуждение в конце § 5.4.)

Особенности приведенной конструкции связаны с тем, что она по сути дела представляет собой тривиальную переформулировку критерия типичности в терминах игр. Этим объясняется и наличие «выбрасываемой» части капитала (соответствующее положительной мере конечных последовательностей относительно априорной вероятности на Σ) и требование полувычислимости снизу для значений функции L (а, скажем, не для частных $L(x1)/L(x)$ и $L(x0)/L(x)$).

ДОПОЛНЕНИЕ
РОБКАЯ КРИТИКА В АДРЕС ТЕОРИИ ВЕРОЯТНОСТЕЙ

Начнем с примера из книги Пойа [Пойа 57, гл. XIV, п. 7, с. 329]. Пусть «... из 315672 попыток выбросить игральными костями пять или шесть очков было 106602 случая успеха. Если бы... кости были честными, то... мы должны были ожидать приблизительно $315672/3 = 105224$ успеха... Таким образом, число, полученное из наблюдений, отклоняется от ожидаемого числа на ... 1378 единиц. Говорит ли такое отклонение за или против гипотезы честных костей?»

Вопрос традиционен, и не менее традиционен ответ. «...Наше суждение зависит от решения следующей задачи: дано, что вероятность успеха равна 1/3 и что испытания независимы; найти вероятность того, что... в 315672 испытаниях число успехов будет больше чем 106601 или меньше чем 103847» (там же; последние два числа — это ожидаемое значение, равное 105524, плюс — минус отклонение в 1377 единиц, на единицу меньшее действительно наблюдавшегося). Далее легко найти, что искомая вероятность меньше $2 \cdot 10^{-7}$, «...и, таким образом, представляется крайне невероятной гипотеза честных костей» (там же, с. 330).

Как видим, схема рассуждений такова. Есть некоторое множество возможных исходов (в данном случае — множество всех возможных протоколов серий из 315672 бросаний). Есть статистическая гипотеза, т. е. некоторое распределение вероятностей на этом множестве (в данном случае это гипотеза «честных костей», согласно которой испытания независимы и вероятности выпадения всех граней равны, т. е. все протоколы равновероятны). И, нако-

нец, есть результат эксперимента, т. е. один из возможных исходов (в данном случае — некоторый протокол, зарегистрировавший 106602 успеха, т. е. выпадения пятерки или шестерки). Мы хотим выяснить, не «противоречит» ли он статистической гипотезе. Для этого предлагается следующая процедура. Выберем некоторое событие, которое имело место в эксперименте (в данном случае — событие «число успехов отклонилось от ожидаемого более чем на 1377»). Подсчитаем его вероятность (оказавшуюся в данном случае меньше $2 \cdot 10^{-7}$). Если она мала, то статистическая гипотеза дискредитируется. «Действительное появление события, которому статистическая гипотеза приписывает маленькую вероятность, является аргументом против этой гипотезы, и чем меньше вероятность, тем сильнее аргумент» (там же, с. 330).

Излагая эту аргументацию, мы обходили тонкое место, имеющееся в ней. Именно, неясно, какие события разрешается рассматривать. Почему мы вычисляли именно вероятность того, что отклонение больше или равно 1378, а не вероятность того, что оно в точности равно 1378? (Получилась бы еще меньшая вероятность!) Мы могли бы даже подсчитать вероятность того, что цифры на костях выпадут именно в том порядке, как это произошло на самом деле, получив уже астрономически малое число. Легко видеть, что последним способом можно дискредитировать гипотезу «честных костей» при любом результате испытаний!

Мы возвращаемся к вопросу, который упоминался во введении, когда мы рассматривали последовательности нулей и единиц длины 12 как протоколы бросания монеты. Мы констатировали, что гипотеза о «честности» монеты (ее симметричности и независимости испытаний) обычно отвергается, если последовательность состоит из 12 нулей — отвергается на том основании, что вероятность такого события (появления последовательности из 12 нулей) очень мала. Но столь же мала и вероятность появления любой другой последовательности!

Было бы наивно думать, что указанный пробел в аргументации мог оставаться незамеченным. Вот что пишет Пойа по поводу все того же примера (с. 329): «...Должны ли мы рассматривать отклонение 1378 как маленькое или как большое? Вероятность такого отклонения высока или низка? Последний вопрос, по-видимому, имеет смысл. Однако мы нуждаемся в разумном истолковании короткого, но важного слова «такое». Мы отвергнем статистическую гипотезу, если вероятность, вычислением которой мы интересуемся, окажется низкой. Однако вероятность того, что отклонение будет в точности равно 1378 единиц, во всяком случае очень мала — была бы очень мала даже вероятность того, чтобы отклонение точно равнялось 0. Поэтому мы должны принять в расчет все отклонения, по абсолютной величине не меньшие, чем наблюдаемое отклонение...»

Согласитесь, что это объяснение выглядит малоубедительно: почему один способ истолкования слова «такое» разумнее другого, так и остается неясным. Но сама проблема «разумного истолкования» сформулирована вполне отчетливо.

Приведем еще несколько цитат. А. Реньи пишет (от имени Б. Паскаля) в своих «Письмах о вероятности» [Рен 80, с. 153]: «Что же, собственно говоря, означает утверждение, что колода карт «тщательно перетасована»? ...При тщательном тасовании вероятности каждого из возможных порядков расположения карт должны быть одинаковы. Но как же в таком случае лишь на основании изучения того, как расположены карты, можно решить, насколько хорошо перетасована колода, ведь появление одного порядка расположения столь же вероятно, как и появление любого другого? А если на основании рассмотрения порядка нельзя решить, что колода хорошо перетасована, то имеет ли само это предположение какой-то определенный смысл?»

Возражения такого рода в адрес теории вероятностей имеют давнюю историю. Эмиль Борель в книге «Случай» [Бор 23, с. 76—77] приводит следующее выскызывание Бертрана (автора известного «парадокса Бертрана», по-

казывающего, что различные способы вычисления вероятности того, что «наудачу» выбранная хорда окружности будет стягивать больше трети окружности, приводят к разным ответам): «Плеяды [звездное скопление, в котором невооруженным глазом видно около десятка звезд; обсуждается вопрос о том, образуют ли звезды Плеяд на самом деле скопление в пространстве или их близость на небесной сфере объясняется случайным совпадением. Авт.] кажутся более близкими друг к другу, чем следует. Это утверждение заслуживает внимания, но если бы мы захотели бы выразить выводы цифрами, у нас не хватило бы данных. Каким образом можно точно определить это туманное понятие близости? Искать наименьший круг, в который заключена данная группа? Наибольшее угловое расстояние? Сумму квадратов всех расстояний?... Все эти величины в группе Плеяд будут меньше, чем можно было бы ожидать. Которая из них будет мерилом вероятности? Если три звезды образуют равносторонний треугольник, следует ли считать, что это обстоятельство, мало вероятное a priori, указывает на существование некоторой причины?».

Несмотря на распространенность такого рода возражений, общепринятого и ясного ответа на них, по-видимому, нет. Воспроизведем отрывок из указанной книги Бореля [Бор 23, с. 77—78]: «Скажем несколько слов о мысли Бертрана по поводу равностороннего треугольника, который могли бы образовать 3 звезды; она связана с вопросом о круглом числе. Если рассматривать число, взятое наудачу между 1000000 и 2000000, то вероятность, что оно равняется 1342517, равна одной миллионной; вероятность, что оно равняется 1500000, равна тоже одной миллионной. Между тем эту последнюю возможность охотно будут считать менее вероятной, чем первую; это зависит от того, что никто и никогда не представляет себе индивидуально такое число, как 1542317; оно рассматривается как тип чисел такого же вида, и если при его переписке изменяется одна цифра, то это едва замечается, и число 1324519 не различается от 1324517; читателю нужно сделать усилие, чтобы убедиться, что четыре числа, написанные в предыдущих строках — все разные.

Когда число вроде предыдущего встречается как мера угла в десятичных долях центезимальных секунд, мы не задаемся вопросом о том, какова была вероятность, чтобы данный угол равнялся именно 13°42′51″,7; ибо мы никогда не поставили бы себе такого вопроса до измерения угла. Этот угол должен, конечно, иметь какое-нибудь значение, и каково бы оно ни было с точностью до одной десятой секунды, можно было бы, измерив его, сказать, что вероятность a priori, что это значение будет именно таково, каково оно есть, равняется одной десятимиллионной и что это факт крайне удивительный...

Вопрос заключается в том, нужно ли делать те же оговорки, если установлено, что один из углов треугольника, образованного тремя звездами, имеет замечательное значение и равен, например, углу равностороннего треугольника... или половине прямого угла... Вот что можно сказать по этому поводу: следует очень опасаться склонности считать замечательным обстоятельство, не определенное в точности перед опытом, так как число обстоятельств, могущих, с различных точек зрения, показаться замечательными, очень значительно».

Мы оставим этот отрывок без комментариев, отметив лишь, что трудно представить себе, каким образом тот субъективный факт, что гипотеза высказывалась кем-то до опыта, может быть учтен в рамках какой-либо объективной математической теории.

Еще один вопрос, естественно возникающий при обсуждении применений теории вероятностей, состоит в следующем. Допустим, статистическая гипотеза так или иначе выбрана. Ну и что? Мы привыкли к тому, что цель науки — предсказание. Но теория вероятностей ничего не предсказывает достоверно: все ее предсказания сами имеют вероятностный характер. Этот «порочный круг лежит очень глубоко: так как достоверность невозможна, то

все, что A [теория вероятностей — Авт.] может утверждать, должно быть сформулировано в терминах вероятности» (Литлвуд [Лит 78, с. 58]).

Мы говорили о трудностях, возникающих при попытке объяснить, как теория вероятностей применяется к явлениям реального мира. Попытаемся указать теперь, как можно пытаться их преодолеть, следуя [Шень 83].

Применение теории вероятностей происходит в два этапа. На первом этапе мы оцениваем согласие той или иной статистической гипотезы с результатами наблюдений. Проиллюстрированное выше на примере в начале дополнения правило отбора статистических гипотез, согласно которому «действительное появление события, которому статистическая гипотеза приписывает малую вероятность, является аргументом против этой гипотезы» ([Пойа 57, гл. XIV, п. 7, с. 330]), по-видимому, может быть сделано более корректным, если рассматривать не все события, а лишь «просто описываемые». Ясно, что событие «выпало 1000 гербов подряд» описывается проще, чем равновероятное событие «выпала последовательность A», где A — «случайная» последовательность из тысячи гербов и решеток. Эта разница, быть может, и является причиной того, что наблюдения этих событий по-разному сказываются на нашем отношении к гипотезе «честной монеты». Для уточнения понятия «просто описываемого события» может оказаться полезным введенное Колмогоровым понятие энтропии (или, как иногда говорят, сложности) конструктивного объекта (о нем см. выше в главе 3).

Выбрав ту или иную статистическую гипотезу — согласующуюся, как мы решили, с результатами наблюдений,— мы переходим ко второму этапу — делаем те или иные выводы из принятой гипотезы. По-видимому, здесь приходится признать, что теория вероятностей ничего не предсказывает, а может лишь давать рекомендации такого рода: если (вычисленная на основе принятой статистической гипотезы) вероятность события A превосходит вероятность события B, то реальность события A надлежит учитывать в большей степени, чем реальность события B.

Из этого можно вывести такой принцип: события, вероятности которых малы, можно не учитывать. В уже упоминавшейся книге [Бор 23] Борель пишет: «...в Париже меньше миллиона взрослых людей; газеты ежедневно извещают о странных случаях или несчастьях, постигших одного из них; жизнь была бы невозможной, если бы каждый постоянно опасался за себя в отношении всех приключений, о которых можно прочесть в отделе разных происшествий; а это равносильно тому, чтобы сказать, что практически нужно пренебречь вероятностями, которые меньше одной миллионной... Часто боязнь дурного приводит к еще худшему. Чтобы уметь различить худшее, надо хорошо знать вероятности различных явлений» (с. 159—160).

Иногда правило отбора статистических гипотез и правило их применения объединяют в формулировке «события с малой вероятностью достоверно не происходят»; вот что пишет, например, Борель: «Не следует бояться применить слово достоверность для обозначения вероятности, которая отличается от единицы на достаточно малую величину» ([Бор 61], с. 7). Мы предпочитаем, однако, разделять эти этапы, так как на первом из них существенную роль играет простота описания маловероятного события, в то время как на втором она, по-видимому, безразлична. (Впрочем, можно ожидать, что интересные для нас события — именно в силу этого интереса — имеют простое описание.)

СПИСОК ЛИТЕРАТУРЫ

Асарин Е. А. (Asarin E. A.)

[Аса 86] (совместно с Покровским А. В.) Применение колмогоровской сложности к анализу динамики управляемых систем // Автоматика и телемеханика.— 1986.— № 1.— С. 25—33.

[Аса 86a] Individual random continuous functions // Тезисы докладов Первого всемирного конгресса общества математической статистики и теории вероятностей им. Бернулли. / Отв. ред. Ю. В. Прохоров — Т. 1.— М.: Наука, 1986.— С. 450.

[Аса 87] О некоторых свойствах Δ-случайных по Колмогорову конечных последовательностей // Теория вероятностей и ее применения.— 1987.— Т. XXXII, вып. 3.— С. 556—558.

[Аса 87а] О некоторых свойствах случайных в алгоритмическом смысле конечных объектов // Доклады АН СССР.— 1987.— Т. 295, № 4.— С. 782—785.

Борель Э.

[Бор 23] Случай.— Пер. с франц.— М., Петроград: Госиздат, 1923.— 215 с. (Современные проблемы естествознания. Кн. 8.)

[Бор 61] Вероятность и достоверность.— Пер. с франц.— М.: Физматгиз, 1961.— 119 с.

Витаньи П., Ли М.

[Вит Ли 88] Колмогоровская сложность: двадцать лет спустя // Успехи математических наук.— 1988.— Т. 43, вып. 6 (264).— С. 129—166.

Вовк В. Г. (Vovk V. G.)

[Вовк 85] Алгоритмическая теория информации и задача прогнозирования // Сложностные проблемы математической логики / Под ред. М. И. Кановича.— Калинин, 1985.— С. 21—24.

[Вовк 86] О понятии бернуллиевости // Успехи математических наук.— 1986.— Т. 41, вып. 1 (247).— С. 185—186.

[Вовк 86а] Об одном варианте понятия случайности по Мизесу — Чёрчу // IV Всесоюзная конференция «Применение методов математической логики». Тезисы докладов. Секция «Алгоритмика трудных задач».— Таллинн, 1986.— С. 95—97.

[Вовк 86б] On use of the algorithmic notions of randomness and simplicity in the mathematical statistics // Тезисы докладов Первого Всемирного Конгресса Общества математической статистики и теории вероятностей им. Бернулли. / Отв. ред. Ю. В. Прохоров — Т. 1.— М.: Наука, 1986.— С. 456.

[Вовк 87] Закон повторного логарифма для случайных по Колмогорову, или хаотических, последовательностей // Теория вероятностей и ее применения.— 1987.— Т. XXXII, вып. 3.— С. 456—468.

[Вовк 87а] Об одном критерии случайности // Доклады АН СССР.— 1987.— Т. 294, № 6.— С. 1298—1302.

[Вовк 88] О законе повторного логарифма Колмогорова — Стаута // Математические заметки.— 1988.— Т. 44, № 1.— С. 27—37.

Вьюгин В. В. (V'jugin V. V.)

[Вью 81] Алгоритмическая энтропия (сложность) конечных объектов и ее применение к определению случайности и количества информации. // Семиотика и информатика.— М.: ВИНИТИ, 1981.— Вып. 16 (второй выпуск за 1980 г.).— С. 14—43.

[Вью 86] Some estimates for nonstochastic sequences. // Тезисы докладов Первого Всемирного Конгресса Общества математической статистики и теории вероятностей им. Бернулли. / Отв. ред. Ю. В. Прохоров — Т. 1.— М.: Наука, 1986.— С. 455.

[Вью 87] О дефекте случайности конечного объекта относительно мер с заданными границами их сложности. // Теория вероятностей и ее применения.— 1987.— Т. XXXII, вып. 3.— С. 558—563.

Гач П. (Gacs P.)

[Гач 74] О симметрии алгоритмической информации // Доклады АН СССР.— 1974.— Т. 218, № 6.— С. 1265—1267.

[Гач 83] On the relation between descriptional complexity and algorithmic complexity // Theoretical Computer Science.— 1983.— V. 22.— P. 71—93.

[Гач 83а] Every sequence is Turing-reducible to a random sequence // Information and Control.— 1986.— V. 70.— P. 186—192.

Дэвид А. П. (David A. P.)

[Дэв 85] Calibration-based empirical probability // The Annals of Statistics.— 1985.— V. 13, № 4.— P. 1251—1273.

Ершов Ю. Л.

[Ерш 72] Вычислимые функционалы конечных типов // Алгебра и логика.— 1972.— Т. 11, № 4.— С. 367—437.

З в о н к и н А. К., Л е в и н Л. А.

[Зво Лев 70] Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов // Успехи математических наук.— Т. 25, вып. 6 (156).— С. 85—127.

К о л м о г о р о в А. Н. (K o l m o g o r o v A. N.)

[Колм 63] On tables of random numbers // Sankhya. The Indian Journal of statistics. Ser. A.— 1963.— V. 25, part 4.— P. 369—376. (Русский перевод: О таблицах случайных чисел. // Семиотика и информатика.— М.: ВИНИТИ, 1982.— Вып. 18 (второй выпуск за 1981 г.) — С. 3—13.)

[Колм 65] Три подхода к определению понятия «количество информации» // Проблемы передачи информации.— 1965.— Т. 1, вып. 1.— С. 3—11. (Перепечатано в [Колм 87, с. 213—223].)

[Колм 69] К логическим основам теории информации и теории вероятностей // Проблемы передачи информации.— 1969.— Т. 5, вып. 3.— С. 3—7. (Перепечатано в [Колм 87, с. 232—237].)

[Колм 83] Комбинаторные основания теории информации и исчисления вероятностей // Успехи математических наук.— 1983.— Т. 38, вып. 4.— С. 27—36. (Перепечатано в [Колм 87, с. 238—250].)

[Колм 83a] On logical foundations of probability theory // Probability Theory and Mathematical Statistics. Proceedings of the Fourth USSR — Japan Symposium, held at Tbilisi, USSR, August 23—29, 1982. / K. Ito, J.V. Prokhorov, eds.— Berlin et al.: Springer-Verlag, 1983. (Lecture Notes in Mathematics. V. 1021.) — P. 1—5. (Русский перевод: О логических основаниях теории вероятностей. // А. Н. Колмогоров. Теория вероятностей и математическая статистика.— М.: Наука, 1986.— 534 с.— С. 467—471.) [Данный текст представляет собой реконструкцию доклада А. Н. Колмогорова на указанной конференции, сделанную по — весьма некачественной — магнитофонной записи А. К. Звонкиным, А. А. Новиковым и А. Шенем без участия А. Н. Колмогорова.]

[Колм 87] Теория информации и теория алгоритмов.— М.: Наука, 1987.— 304 с.

К о л м о г о р о в А. Н., У с п е н с к и й В. А.

[Колм Усп 87] Алгоритмы и случайность. // Теория вероятностей и ее применения.— 1987.— Т. XXXII, вып. 3.— С. 425—455.

Л а в л э н д Д. (L o v e l a n d D.)

[Лав 66] A new interpretation of the von Mises'concept of random sequence // Zeitschrift f. mathematische Logik und Grundlagen d. Mathematik.— 1966.— Bd 12, № 3.— S. 279—294.

[Лав 66a] The Kleene hierarchy classification of recursively random sequences // Transactions of the American Mathematical Society.— 1966.— V. 125, № 3.— P. 497—510.

Л а м б а л ь г е н М. (L a m b a l g e n M. von)

[Лам 87] Von Mises'definition of random sequences reconsidered // The Journal of Symbolic Logic.— 1987.— V. 52, № 3.— P. 725—755.

[Лам 87a] Random sequences. Academisch Proefschrift.— Amsterdam.— 1987.

[Лам 89] The axiomatization of randomness.— Preprint.— University of Amsterdam, 1989.

Л е в и н Л. А. (L e v i n Leonid A.)

[Лев 73] О понятии случайной последовательности // Доклады АН СССР.— 1973.— Т. 212, № 3.— С. 548—550.

[Лев 74] Законы сохранения (невозрастания) информации и вопросы обоснования теории вероятностей // Проблемы передачи информации.— 1974.— Т. 10, вып. 3.— С. 30—35.

[Лев 76] Равномерные тесты случайности // Доклады АН СССР.— 1976.— Т. 227, № 1.— С. 33—35.

[Лев 76a] О различных мерах сложности конечных объектов (аксиоматическое описание) // Доклады АН СССР.— 1976.— Т. 227, № 4.— С. 804—807.

[Лев 76б] О принципе сохранения информации в интуиционистской математике // Доклады АН СССР.— 1976.— Т. 227, № 6.—1293—1296.

[Лев 77] Об одном конкретном способе задания сложностных мер // Доклады АН СССР.— 1977.— Т. 234, № 3.— С. 536—539.

[Лев 80] A concept of independence with applications in various fields of mathematics.— April 1980.— Massachusets institute of technology. Laboratory for computer science. Technical Report 235. (РЖМат 1981, № 1, реферат 1В827.)

Литлвуд Дж.

[Лит 78] Математическая смесь.— 4-ое изд.— М.: Наука, 1978.— 143 с.

Мартин-Лёф П. (Martin-Löf P.)

[Мар 66] О понятии случайной последовательности // Теория вероятностей и ее применения.— 1966.— Т. XI, № 1.— С. 198—200.

[Мар 66а] The definition of random sequences // Information and control.— 1966.— V. 9, № 6.— P. 602—619.

[Мар 68] On the notion of randomness // Intuitionism and proof theory. / Eds. Kino A. et al.— N. Y., 1968.— P. 73—78. (Русский первод: О понятии случайности // Сложность вычислений и алгоритмов / Под ред. В. А. Козмидиади, А. Н. Маслова и Н. В. Петри.— М.: Мир, 1974.— С. 364—369.)

[Мар 70] Notes on constructive mathematics.— Stockholm: Almqvist, Wiksell, 1970.— 109 p. (Русский первод: Очерки по конструктивной математике,— М.: Мир, 1975.— 136 с.)

Мизес Р. фон (Mises R. von)

[Миз 19] Grundlagen der Wahrscheinlichkeitsrechnung // Mathematische Zeitschrift.— 1919.— Bd. 5.— S. 52—89.

[Миз 28] Wahrscheinlichkeit, Statistik und Wahrheit.— Wien: J. Springer, 1928. (Русский перевод: Вероятность и статистика.— М.— Л.: Государственное издательство, 1930.— 250 с.)

[Миз 41] On the foundations of probability and statistics // Annals of Mathematical Statistics.— 1941.— V. 12.— P. 191—205.

Пойа Д.

[Пойа 57] Математика и правдоподобные рассуждения. Том. I. Индукция и аналогия в математике. Том II. Схемы правдоподобных умозаключений.— М.: Издательство иностранной литературы, 1957.— 535 с.

Реньи А.

[Рен 80] Трилогия о математике. Диалоги о математике. Письма о вероятности. Дневник.— Записки студента по теории информации. / Перевод с венгерского.— М.: Мир, 1980.

Роджерс Х.

[Род 72] Теория рекурсивных функций и эффективная вычислимость.— М.: Мир, 1972.— 624 с.

Успенский В. А., Семенов А. Л.

[Усп Сем 87] Теория алгоритмов: основные открытия и приложения.— М.: Наука, 1987.— 286 с. (Библиотечка программиста. № 49.)

Чёрч А. (Church A.)

[Чёрч 40] On the concept of a random sequence // Bulletin of the American Mathematical Society.— 1940.— V. 46, № 2.— P. 130—135.

Чэйтин Г. (Chaitin G.)

[Чэй 66] On the length of programs for computing finite binary sequences // Journal of the Association for Computing Machinery.— 1966.— V. 13.— P. 547—569.

[Чэй 87] Incompleteness theorems for random reals // Advances in Applied Mathematics.— 1987.— V. 8.— P. 119—146.

[Чэй 87а] Algorithmic information theory.— Cambridge, 1987.— X + 175 pp.

Шень А.

[Шень 82] Частотный подход к определению понятия случайной последовательности // Семиотика и информатика.— М.: ВИНИТИ, 1982.— Вып. 18 (второй выпуск за 1981 г.).— С. 14—42.

[Шень 83] К логическим основам применения теории вероятностей // Семиотические аспекты формализации интеллектуальной деятельности. Школа-семинар. Г. Телави,

29 октября — 6 ноября 1983 г. Тезисы докладов и сообщений.— М.: ВИНИТИ, 1983.— 253 с.— С. 144—146.

[Шень 83а] Понятие (α, β)-стохастичности по Колмогорову и его свойства // Доклады АН СССР.— 1983.— Т. 271, № 6.— С. 1337—1340.

[Шень 84] Алгоритмические варианты понятия энтропии // Доклады АН СССР.— 1984.— Т. 276, № 3.— С. 563—566.

[Шень 89] О соотношениях между различными алгоритмическими определениями случайности // Доклады АН СССР.— 1988.— Т. 302, № 3.— С. 548—552.

Шнорр К. П. (Schnorr C. P.)

[Шно 71] Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie.— Berlin etc: Springer-Verlag.— 1971.— IV + 212 s. (Lecture notes in mathematics. V. 218.)

[Шно 73] Process complexity and effective random tests // Journal of computer and system sciences.— 1973.— V. 3, № 4.— P. 376—378.

[Шно 77] A survey of the theory of random sequences // Logic, foundations of mathematics and computability theory. / Eds. R. E. Butts, J. Hintikka.— Dordrecht: D. Reidel.— 1977.— X + 406 pp.— P. 193—211.

Якобс К. (Jacobs K.)

[Яко 70] Turing-Maschinen und zufällige 0—1-Folgen // Selecta mathematica.— V. 2.— Berlin, New York etc: Springer-Verlag, 1970.— S. 141—167. (Русский перевод: Машины Тьюринга и случайные 0—1-последовательности. // Машины Тьюринга и рекурсивные функции.— М.: Мир, 1972.— 264 с.— С. 183—215.)

Московский государственный
университет им. М. В. Ломоносова
Научный Совет по комплексной
проблеме «Кибернетика» АН СССР
Институт проблем передачи
информации АН СССР

Поступила в редакцию
10 июля 1989 г.