В.В.Вьюгин

КОЛМОГОРОВСКАЯ СЛОЖНОСТЬ И АЛГОРИТМИЧЕСКАЯ ТЕОРИЯ ИНФОРМАЦИИ

МОСКВА
2020

УДК 005.519.8(075.8)
ББК 65.290-2в6я73

Вьюгин В.В.

Колмогоровская сложность и алгоритмическая теория информации

Первые две части редназначены для первоначального знакомства с основами теории колмогоровской сложности и алгоритмической случайности. Вводятся и изучаются понятия колмогоровской сложности и случайности конечного объекта. Приведены основные понятия и теоремы колмогоровского подхода к обоснованию теории вероятностей на основе теории информации и теории алгоритмов. Третья часть содержит более сложные результаты, развивающие колмогоровский подход к обоснованию теории информации и теории вероятностей.

Для студентов и аспирантов математических и прикладных математических специальностей.

Оглавление

Введение

В начале 60-х гг. XX в. А. Н. Колмогоров предложил программу построения теории информации и теории вероятностей на принципиально новой алгоритмической основе. Первой публикацией по алгоритмической теории информации является его статья [11], где указан способ измерения сложности конечного объекта (слова). Для этого Колмогоров ввел понятие *алгоритмической сложности* $\mathrm{K}(x)$ конечного объекта x, равной длине самого короткого двоичного кода, по которому некоторый универсальный алгоритм – способ декодирования – может восстановить данный конечный объект x. Основным результатом Колмогорова была «теорема инвариантности», благодаря которой можно определить сложность $\mathrm{K}(x)$ независимо от способа декодирования. Таким образом, алгоритмическая сложность конечного объекта является внутренней характеристикой этого объекта, не зависящей от способа ее измерения. На основе понятия *сложности* вводится понятие *количества информации* $\mathrm{I}(y : x) = \mathrm{K}(x) - \mathrm{K}(x|y)$ в одном конечном объекте y о другом конечном объекте x.

Колмогоровская сложность конечного объекта равна длине самого короткого двоичного слова по которому некоторый универсальный способ декодирования может восстановить этот объект. В рамках классической теории информации энтропия Шеннона равна минимальному количеству битов, с помощью кото-

рых можно описать случайный объект в среднем, а точнее, распределение вероятностей, порождающее такие объекты.

Идеи колмогоровской сложности и созданной на ее основе алгоритмической теории информации возникли на фоне бурного развития теории информации и кодирования, которая была основана знаменитыми работами Клода Шеннона [31].

Близко к идеям колмогоровской сложности находятся идеи универсального сжатия информации и универсального прогнозирования, которые возникли в тот же период времени – в 60-70гг. 20-го столетия. При этом подходе, рассматривается эталонный класс стохастических моделей (reference class) и строится метод кодирования или прогнозирования, который сжимает информацию или прогнозирует будущие исходы не хуже чем любая модель этого класса, правда, с точностью до некоторой погрешности – регрета (избыточности кода). В этом случае, критерием эффективности универсального метода является минимизация регрета. К классу универсальных методов можно отнести методы универсального сжатия информации В.Ф. Бабкина [1], Б.М. Фитингофа [29], Р.Е. Кричевского [13], Ю.М. Штарькова [33], [34], Б.Я.Рябко [22], [23], Д. Риссанена [35] и др.

Шенноновская теория информации существенно основывается на вероятностных предположениях, что сужает область ее применимости. Алгоритмическая теория информации является попыткой распространить идеи и понятия теории информации на нестохастический случай.

Рэй Соломонов [53], [54] впервые стал рассматривать в качестве эталонного класс всех алгоритмически вычислимых моделей (распределений вероятностей) и построил универсальный предсказатель, который доказуемо предсказывал асимптотически не хуже чем любое вычислимое распределение вероятностей. Недостатком такого универсального предсказателя является отсутствие алгоритма вычисления его предсказаний.

Понятие алгоритмической сложности, введенное А.Н.Колмогоровым в [11], также основано на построении универсального метода декодирования для эталонного класса,

состоящего из всех вычислимых методов декодирования. Все эти алгоритмы интегрируются в один универсальный алгоритм с помощью универсальной функции (машины Тьюринга).

Понятие колмогоровской сложности развивает понятие энтропии Шеннона и имеет аналогичные свойства.

Следует отметить, что в классической теории информации имеет смысл рассматривать количество информации только для случайных величин ξ и η, принимающих значения $j \in J$ из некоторого множества:

$$I(\eta : \xi) = H(\xi) - H(\xi|\eta),$$

где H – энтропия Шеннона:

$$H(\xi) = -\sum_j p(\xi = j) \log p(\xi = j),$$

$H(\xi|\eta)$ – условная энтропия Шеннона.

Как видно из этого определения, для задания энтропии необходимо знать распределение вероятностей источника, генерирующего символы j, из которых составлены конечные объекты x. Таким образом, понятия *энтропии (аналога сложности)* и *количества информации* являются внутренними понятиями теории вероятностей и требуют для своего вычисления прежде всего определить вероятностное пространство, описывающее источник данных.

Как известно, вероятностные утверждения интерпретируются через статистические высказывания. Поэтому практически определение энтропии $H(\xi)$ и соответствующего понятия *количества информации* может быть использовано только лишь в применении к обширным совокупностям объектов.

Например, трудно представить себе его применение для определения количества информации, содержащейся в геноме человека (представленном в виде четырехбуквенного слова) о геноме шимпанзе. Для этого надо представить себе эти индивидуальные геномы как элементы обширных совокупностей подобных

им геномов, в которых появление каждого нуклеотида на определенном месте генома описывается некоторыми вероятностями. Каким образом можно оценить эти вероятности неизвестно.

Потребность использовать понятие *сложности* и определяемое через него понятие *количества информации* в случае индивидуальных объектов, не рассматриваемых как реализации случайных величин с определенным законом распределения, вызывает необходимость по крайней мере теоретического изучения соответствующего понятия сложности.

Колмогоровский подход основан на обратной последовательности действий. Сначала определяется понятие *сложности конечного объекта* и определяемое через него понятие *количества информации*, а затем на этом основании развивается теория статистических свойств конечных объектов. Идея Колмогорова, опубликованная в статье [11], заключалась в том, чтобы признаком случайности конечной последовательности символов x считать отсутствие в ней закономерностей, что выражается в невозможности более короткого описания этой последовательности, чем ее длина: в этом случае $\mathrm{K}(x) \approx l(x)$, где $l(x)$ – длина этой последовательности.

А. Н. Колмогоров придавал большое значение изучению понятия алгоритмической случайности *конечного объекта*. При этом понятие меры не должно входить в это определение. Основная идея Колмогорова заключалась в том, чтобы выводить стохастические свойства конечной последовательности из предположения о том, что ее сложность, при заданных ограничениях, близка к максимальному значению.

Алгоритмическая сложность, введенная Колмогоровым, впоследствии была названа *колмогоровской сложностью*, а способ формулирования вероятностных утверждений на основе понятий алгоритмической сложности и количества информации был назван *колмогоровским подходом* к обоснованию теории вероятностей. Основные понятия и идеи колмогоровского подхода для конечных объектов излагаются в главе 3.

Независимо от Колмогорова понятие алгоритмической слож-

ности было также введено Г. Чейтиным [37], [38].

В дальнейшем развитие алгоритмического подхода к теории вероятностей пошло иным путем. В качестве случайных объектов стали рассматриваться бесконечные последовательности исходов (обычно это 0 и 1). При таком подходе исчезают технические трудности, характерные при реализации колмогоровского подхода для конечных объектов. Параллельно с колмогоровским сложностным подходом Мартин-Леф предложил алгоритмическо – вероятностный подход к построению «конструктивной теории вероятностей». Мартин-Леф ввел понятие бесконечной последовательности, случайной относительно заданного вероятностного распределения. Бесконечная последовательность называется *алгоритмически случайной*, если она выдерживает любой вычислимый тест Мартин-Лефа. Определение и свойства случайных по Мартин-Лефу последовательностей излагаются в главе 4.

Как выяснилось позже, понятие случайной по Мартин-Лефу последовательности допускает эквивалентное описание в терминах модифицированных вариантов алгоритмической сложности. Л.А.Левин и К.П.Шнорр ввели в работах [15], [18], [16], [19] и [52] новые версии колмогоровской сложности – монотонную и префиксные версии колмогоровской сложности, которые отличаются от колмогоровской сложности использованием специальных методов декодирования конечных объектов.

Данные виды сложности позволяют дать определение бесконечной случайной последовательности, которое определяет тот же класс последовательностей, что и определение Мартин-Лефа. Таким образом, конструктивный и сложностной подходы к теории вероятностей совпадают. Все эти понятия и теоремы приведены в главе 5.

Независимо от Колмогорова, и даже несколько ранее, идеи построения «универсального предсказателя» были предложены Р. Соломоновым. Соломонов хотел построить меру M с эффективно вычислимыми свойствами, которая бы предсказывала не хуже любой вычислимой меры P. Первоначальные идеи Соло-

монова были не ясны, он уточнил их позже в статьях [53] и [54]. Его уточнения привели к построению универсальной предсказывающей меры, которая правда не обладала достаточными вычислимыми свойствами. Как выяснилось, построить предсказатель, который являлся бы одновременно вычислимым и универсальным, невозможно; позже Левин построил полувычислимый универсальный предсказатель. Здесь наиболее ценной является идея Соломонова об универсальности предсказателя, которая с самого начала присутствовала в его работах. Так же, как и колмогоровская сложность, универсальный предсказатель определялся с использованием универсальной машины Тьюринга, которая строится в теории рекурсивных функций. В этом заключается сходство подходов Колмогорова и Соломонова. Заметим, что Соломонов не рассматривал понятие алгоритмической сложности конечного объекта, а Колмогоров никогда не рассматривал задачу построения универсального предсказания. Позже идея универсального предсказателя получила свое уточнение в виде понятия *универсальной полумеры*, введенной Левиным в 1970 г. [10]. Это понятие изучается в главе 6.

В.Г.Вовк [59] предложил обобщение понятия универсальной полумеры – понятие предсказательной сложности. Это понятие изучается в разделе 10.

Идеи универсального предсказания индивидуальной последовательности предвосхитили появившуюся позже в 1990-х годах «теорию машинного обучения» (Machine Learning), которая имеет более прикладную направленность, чем алгоритмическая случайность (см. [49] и [8]).

Близкие определения случайности с использованием теории мартингалов позже привели к новому теоретико-игровому обоснованию теории вероятностей и финансовой математики, предложенному Вовком и Шейфером [51].

Элементы вероятностной теории информации излагаются в части I книги. В этой части рассматриваются основные понятия и свойства классической теории информации, такие как, энтропия, количество взаимной информации, кодирование вероятност-

ных источников. В главе 2 обсуждаются методы универсального сжатия информации. Алгоритм Лемпеля–Зива изучается в разделе 2.1.

В части II книги представлены основные понятия и утверждения алгоритмической теории информации.

Часть III посвящена различным приложениям алгоритмической теории информации к теории вероятностей. В частности, в разделе 7 приводится «сложностное» доказательство закона повторного алгоритма, предложенное в работе [2], в разделе 8.1 проводится алгоритмический анализ эргодической теоремы Биркгофа для случайных по Мартин-Лефу последовательностей, предложенный в статье [7].

В части IV изучаются проблемы предсказуемости конечных и бесконечных последовательностей. В главе 10 рассматривается задача последовательного прогнозирования конечных последовательностей, вводится и изучается предсказательная сложность. В главе 11 изучаются (α, β)-стохастические по Колмогорову конечные последовательности. В части V вводится классификация и изучаются свойства бесконечных последовательностей, как носителей информациию, которые могут быть получены в комбинациях стохастических и алгоритмических процессов. В частности, в главе 12 изучаются алгоритмически-инвариантные свойства бесконечных последовательностей.

В настоящее время теория колмогоровской сложности представляет собой один из разделов математики, по которому издаются монографии и проводятся международные конференции. Первая обзорная статья по колмогоровской сложности и случайности была опубликована в 1971 г. Звонкиным и Левиным [10]. Изложение колмогоровской концепции случайности и новые результаты в области алгоритмической теории информации были представлены в 1981 г. в обзоре Вьюгина [4]. Здесь впервые были приведены доказательства новых результатов Левина, опубликованных (без доказательства) в 1970-х годах в статьях [15], [16], [18]. В 1990 г. был опубликован обзор Успенского и др. [27].

За рубежом общепринятым источником в области колмого-

ровской сложности является монография Ли и Витаньи [48]. В настоящее время наиболее полное изложение теории колмогоровской сложности представлено в монографии Успенского, Верещагина, Шеня [27]. Для более глубокого изучения предмета рекомендуются лекции П. Гача [43].

Части I и II составлены на основе курса «Колмогоровская сложность и ее приложения», прочитанного автором на факультете прикладной математики и управления Московского физико-технического института в 2011–2021 гг. Все главы этой части снабжены задачами. Часть III содержит более сложные вопросы и может быть использована для использования в более продвинутых специальных курсах.

Часть I книги основана на монографии [40] и была добавлена при обновлении 2021г. Часть II данного учебного пособия (главы 3–6) представляет собой расширенное изложение обзорной работы [4], которая была дополнена рядом результатов и доказательств из работы [61] и монографии [27]. При этом содержание раздела 6.2 основано на материале из статьи [54] и ее изложении в монографии [48]. Часть III основана на материале статей [2], [7] и [61]. Часть IV основана на материале статей [3], [5] и [62].

Часть I

Элементы вероятностной теории информации

Глава 1

Кодирование

Задан алфавит $A = \{a_1, \ldots, a_k\}$. Под источником понимаем механизм генерации слов, состоящих из букв алфавита A. Мы будем рассматривать два типа источников: вероятностные источники и источники типа "черный ящик".

Вероятностный источник представлен распределением вероятностей на множестве всех букв алфавита A. Вероятностный источник выдает букву a_i алфавита A с заданной вероятностью p_i. Здесь $p_i \geqslant 0$ для всех i и $\sum_{i=1}^{k} p_i = 1$. Методы кодирования будут учитывать распределение вероятностей соответствующего источника. Свойства вероятностных источников описываются шенновской теорией информации и будут изучаться в главе 1.

Под "черным ящиком" понимаем механизм генерации букв неизвестной нам природы. Это значит, что мы должны использовать методы кодирования, которые не основываются на какой-либо модели генерации данных. Эти методы будут изучаться в части II посвященной алгоритмической теории информации.

Промежуточное место занимают универсальные методы сжатия информации. Соответствующий алгоритм сжатия не использует никаких данных о природе источника букв. Однако алгоритм эффективен только для определенных классов (вероятностных) источников. Алгоритм универсального сжатия информации, предложенный Зивом и Лемпелем, будет рассмотрен в

разделе 2.1. Доказано, оптимальность алгоритма в случае стационарного эргодического источника.

1.1. Коды

Будем кодировать слова в конечном алфавите $A = \{a_1, \ldots, a_k\}$. Под словом понимается произвольная последовательность букв $x = x_1 x_2 \ldots x_n$, где $x_i \in A$ при $1 \leqslant i \leqslant n$, длина слова $l(x) = n$. Обозначим A^* – множество всех слов в алфавите A. Для удобства вводим пустое слово λ. Обозначаем A^n – множество всех слов длины n, составленных из букв алфпвита A, посредством A^* обозначим множество всех слов, составленных из букв алфавита A.

Слова $x = x_1 x_2 \ldots x_n$ и $y = y_1 y_2 \ldots y_m$ можно записывать одно после другого: $xy = x_1 x_2 \ldots x_n y_1 y_2 \ldots y_m$ – конкатенация слов x и y. Слово y продолжает слово x, а слово x является началом или префиксом слова y, если $y = xz$ для некоторого слова z, обозначаем $x \subseteq y$. Если $z \neq \lambda$, то пишем $x \subset y$. Два слова x и y несравнимы, если они не продолжают друг друга. Слово y называется подсловом слова x, если $x = uyv$ для некоторых слов u и v.

Пусть $I = \{0, 1\}$ – бинарный алфавит. Будем кодировать слова из A^* двоичными словами. Код – это функция $C : A \to \{0, 1\}^*$, $C(A)$ – множество кодовых слов. Функция C каждой букве из A сопоставляет слово из 0 и 1. Слово $x = x_1 x_2 \ldots x_n$ кодируем побуквенно словом $p = C(x) = C(x_1)C(x_2) \ldots C(x_n)$. Код C – однозначный, если $C(x) \neq C(y)$, при $x \neq y$ для всех $x, y \in A^*$. Код C – однозначно декодируемый, если существует функция $D : \{0, 1\}^* \to A^*$ (декодер) такая, что $D(C(x)) = x$ для всех $x \in A^*$.

Множество слов X называется безпрефиксным, если любые два различных слова из X не продолжают друг друга (несравнимы): $x \not\subseteq y$ для любых $x, y \in X$ таких, что $x \neq y$.

Код C называется безпрефиксным, если множество всех кодовых слов $C(A)$ является безпрефиксным. Легко видеть, что

имеет место следующее утверждение.

Предложение 1.1. *Всякий безпрефиксный код является однозначно декодируемым.*

Легко построить простейший безпрефиксный код.

Предложение 1.2. *Для любого алфавита A можно построить безпрефиксный код $C(x)$ такой, что $l(C(x)) \leqslant \log|A|+1$ для всех $x \in A$.*[1]

Доказательство. Выберем k так, чтобы $2^k \leqslant |A| < 2^{k+1}$. Общее число всех двоичных последовательностей длины $k+1$ равно 2^{k+1}. Все они не продолжают друг друга, поэтому их достаточно, чтобы установить взаимно-однозначное соответствие между некоторым помножеством всех двоичных строк длины $k+1$ и буквами из A. □

Описание всех беспрефиксных кодов дается в следующей ниже теореме Крафта.

Теорема 1.1. *Беспрефиксный код с длинами кодовых слов $l_1, \ldots, l_k$ ($k = |A|$) может быть построен тогда и только тогда, когда $\sum\limits_{i=1}^{k} 2^{-l_i} \leqslant 1$. Это неравенство называется неравенством Крафта.*

Доказательство. Пусть заданы числа $l_1 \leqslant l_2 \leqslant \ldots \leqslant l_k$, для которых выполнено неравенство Крафта $\sum\limits_{i=1}^{k} 2^{-l_i} \leqslant 1$. Построим соответствующий код. Выберем попарно несравнимые двоичные последовательности с этими длинами. Пусть первая из них состоит из одних нулей и имеет длину l_1. После этого, ввиду требования безпрефиксности, имеется $2^{l_2-l_1}$ последовательностей длины l_2 (продолжений выбранного кодового слова длины l_1), которые нельзя использовать в качестве кодовых слов длины l_2.

[1] Здесь и далее log обозначает логарифм по основанию 2, ln – натуральный логарифм, $|A|$ – число элементов множества A.

Однако, так как $2^{l_2} > 2^{l_2-l_1}$, найдется слово длины l_2, которое можно использовать в качестве второго кодового слова. После этого, имеется $2^{l_3-l_1} + 2^{l_3-l_2}$ "запрещенных" к использоваию кодовых слов длины l_3. Так как из неравенства Крафта следует, что $2^{l_3} > 2^{l_3-l_1} + 2^{l_3-l_2}$,мы внось найдем какое-нибудь несравнимое с ранее выбранными кодовое слово длины l_3. Продолжаем этот процесс выбора кодовых слов k раз. Код будет построен.

Допустим, что существует безпрефиксный код с длинами кодовых слов $l_1, l_2, \ldots, l_k$. Пусть $l^* = \max_{1 \leqslant i \leqslant k} l_i$. Из свойства безпрефиксности следует, что множество всех двоичных последовательностей длины l^* можно представить в виде объединения попарно непересекающихся множеств последовательностей (продолжающих кодовые слова), состоящих из $2^{l^*-l_1}, 2^{l^*-l_2}, \ldots, 2^{l^*-l_k}$ элементов соответственно. Так как $2^{l^*} \geqslant 2^{l^*-l_1} + 2^{l^*-l_2} + \cdots + 2^{l^*-l_k}$, получаем неравенство Крафта. □

Приводимая ниже теорема МакМиллана показывает, что неравенство Крафта является характеристических свойством однозначно декодируемых кодов.

Теорема 1.2. *1) Длины кодовых слов произвольного однозначно декодируемого кода удовлетворяют неравенству Крафта* $\sum\limits_{i=1}^{k} 2^{-l_i} \leqslant 1$, *где* $l_1, l_2, \ldots, l_k$ *– длины кодовых слов. 2) Для произвольного набора длин кодовых слов, удовлетворяющих неравенству Крафта, можно построить однозначно декодируемый код с этими длинами кодовых слов.*

Доказательство. Часть 2) утверждения теоремы следует из теоремы 1.1.

Докажем теперь часть 1) утверждения теоремы. Пусть $A = \{a_1, \ldots, a_k\}$ и задан однозначно декодируемый код C, $l_a = l(C(a))$ при $a \in A$. Для произвольного натурального числа M

рассмотрим

$$\left(\sum_{a\in A} 2^{-l_a}\right)^M = \sum_{x_1\in A}\sum_{x_2\in A}\cdots\sum_{x_M\in A} 2^{-l_{x_1}-l_{x_2}-\ldots l_{x_k}} = \\ = \sum_{L=1}^{Ml^*} N(L)2^{-L}, \qquad (1.1)$$

где $l^* = \max_{a\in A} l_a$, $N(L) = |X(L)|$ и $X(L) = \{x_1,\ldots,x_M : \sum_{i=1}^{M} l_{x_i} = L\}$. Все слова $x_1\ldots x_M \in X(L)$ различные, поэтому их коды также различные (так как код однозначно декодируемый). Все эти коды имеют длину L, поэтому всего их не больше чем 2^L. Так как в (1.1) $N(L)2^{-L} \leqslant 1$, получаем

$$\sum_{a\in A} 2^{-l_a} \leqslant (Ml^*)^{-M} \to 1 \text{ при } M\to\infty. \qquad (1.2)$$

Поскольку левая часть (1.2) не зависит от M, получаем $\sum_{a\in A} 2^{-l_a} \leqslant 1$. □

1.2. Энтропия Шеннона и коды

В этом разделе мы будем изучать вероятностные источники и связанные с ними коды. Задан алфавит $A = \{a_1,\ldots,a_k\}$. Вероятностный источник представлен распределением вероятностей на множестве всех букв алфавита A. Вероятностный источник X выдает букву a_i алфавита A с заданной вероятностью $p_i = P\{X = a_i\}$. Здесь $p_i \geqslant 0$ для всех i и $\sum_{i=1}^{k} p_i = 1$. Методы кодирования будут учитывать распределение вероятностей соответствующего источника.

Энтропия. Энтропия источника X определяется

$$H(X) = -\sum_{i=1}^{k} p_i \log p_i.$$

Полагаем $0 \log 0 = 0$. Также пишем $H(X) = -\sum\limits_{a \in A} p(a) \log p(a) = -\sum\limits_{a \in A} P(X = a) \log P(X = a) = E_{a \sim p}[-\log p(a)]$. Понятие энтропии относится к распределению вероятностей p и не зависит от того, какие значения принимает случайная величина X. Поэтому иногда обозначаем энтропию как $H(p)$.

В дальнейшем будем предполагать, что $p(a) > 0$ для любого $a \in A$, так как если $p(a) = 0$, то мы можем исключить букву a из алфавита A.

Далее мы покажем, что энтропия источника равна минимально необходимому количеству битов, в среднем необходимых для кодирования буквы алфавита.

Простейшие свойства энтропии представлены в задачах раздела 1.4.

Пример. Пусть $A = \{a_1, a_2\}$ и X – бернуллиевская случайная величина: $P\{X = a_1\} = p$ и $P\{X = a_2\} = 1 - p$. Тогда $H(X) = H(p) = -p \log p - (1 - p) \log(1 - p)$, $0 \leqslant H(p) \leqslant 1$ для всех p.

Энтропия пары, условная энтропия. Пусть A и B – два алфавита, (X, Y) – пара случайных величин, принимающих значения в A и B соответственно. Задано распределение вероятностей этой пары $p(x, y) = P\{X = x, Y = y\}$, где $x \in A$, $y \in B$. Соответствующие маргинальные распределения $p(x) = \sum_{y \in B} p(x, y)$ на A и $p(y) = \sum_{x \in A} p(x, y)$ на B. Задана условная вероятность $p(y|x) = p\{Y = y | X = x\} = \frac{p(x,y)}{p(x)}$.

По определению энтропия пары случайных величин равна

$$H(X, Y) = -\sum_{(x,y) \in A \times B} p(x, y) \log p(x, y).$$

Введем энтропию относительно условного распределения

$$H(Y|X = x) = -\sum_{y \in B} p(y|x) \log p(y|x).$$

Условная энтропия определяется

$$H(Y|X) = \sum_{x \in A} p(x) H(Y|X = x).$$

Можно переписать эту величину подробнее

$$H(Y|X) = -\sum_{x \in A} p(x) \sum_{y \in B} p(y|x) \log p(y|x) = \\ = -\sum_{(x,y) \in A \times B} p(x,y) \log p(y|x).$$

Если $A = B$ и $X = Y$, то $p(x,x) = p(x)$ и $p(x|x) = 1$. Отсюда $H(X,X) = H(X)$ и $H(X|X) = 0$.

Следующая теорема устанавливает разложение для энтропии пары.

Теорема 1.3. $H(X,Y) = H(X) + H(Y|X)$.

Доказательство. По определению

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y) = \\ = -\sum_{x} \sum_{y} p(x,y) \log p(x) p(y|x) = \\ = -\sum_{x} \left(\sum_{y} p(x,y) \right) \log p(x) - \sum_{x} \sum_{y} p(x,y) \log p(y|x) = \\ = H(X) + H(Y|X).$$

□

Из теоремы 1.3 непосредственно следует

Следствие 1.1. $H(X) - H(X|Y) = H(Y) - H(Y|X)$.

Правило цепи. Теорему 1.3 можно обобщить следующим образом.

$$H(X_1, X_2, X_3) = H(X_3|X_2, X_1) + H(X_1, X_2) = \\ = H(X_3|X_2, X_1) + H(X_2|X_1) + H(X_1).$$

В общем случае имеет место

Следствие 1.2. $H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1)$.

Относительная энтропия, количество информации.

Пусть $p(x)$ и $q(x)$ – распределения вероятностей на алфавите A. Относительная энтропия или расхождение Кульбака–Лейблера определяется

$$D(p\|q) = \sum_{x\in A} p(x) \log \frac{p(x)}{q(x)}.$$

Предложение 1.3. *$D(p\|q) \geqslant 0$ и равенство нулю имеет место тогда и только тогда, когда $p = q$.*

Доказательство. Из вогнутости логарифма имеем

$$-D(p\|q) = \sum_{x} p(x) \log \frac{q(x)}{p(x)} \leqslant$$

$$\leqslant \log \sum_{x} p(x) \frac{q(x)}{p(x)} = \log \sum_{x} q(x) = 0. \tag{1.3}$$

Равенство в (1.3) только при $\frac{q(x)}{p(x)} = 1$ для всех x. □

Пусть $(X, Y) \sim p_{X\times Y}(x, y)$ – совместное распределение пары случайных величин и $X \sim p_X$, $Y \sim p_Y$ – соттветствуэщие маргинальные распределения. Взаимное количество информации в случайной величине X о случайной величине Y определяется как

$$I(X : Y) = D(p_{X\times Y}\|p_X \times p_Y) = D(p(x, y)\|p(x)p(y)) =$$
$$= \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$

Следующие свойства непосредственно следуют из определения:

$I(X : Y) = I(Y : X)$,

$I(X : Y) \geqslant 0$,

$I(X : Y) = 0$ тогда и только тгда, когда $p(x, y) = p(x)p(y)$, т.е. когда случайные величины X и Y независимые.

Выражение количества информации через энтропию. Удобно количество информации записывать через энтропию.

Теорема 1.4. $I(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$.

Доказательство. По определению

$$I(X:Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} =$$
$$= -\sum_{x,y} p(x,y) \log p(x) + \sum_{x,y} p(x,y) \log p(x|y) =$$
$$= H(X) - H(X|Y).$$

□

Из равенства $H(X,Y) = H(X)+H(Y|X) = H(Y)+H(X|Y)$ и теоремы 1.4 получаем симметричное выражение для количества информации

Следствие 1.3. $I(X:Y) = H(X) + H(Y) - H(X:Y)$.

Средняя длина кода. Заданы алфавит $A = \{a_1, \ldots, a_k\}$ и код C на нем. Пусть $l_i = l(C(a_i))$ – длина i-го кодового слова. Задано распределение вероятностей $p_i = p(a_i)$ на A, $1 \leqslant i \leqslant k$, H – его энтропия. Средней длиной кода называется математическое ожидание длины кодового слова $L = \sum_{i=1}^{k} p_i l_i$.

Теорема 1.5. *Пусть задано распределение вероятностей на алфавите A. Тогда*

1) Для любого однозначно декодируемого кода $L \geqslant H$.

2) Существует безпрефиксный код, для которого $L \leqslant H + 1$.

Доказательство. Пусть l_i – длина i-го кодового слова. По теореме 1.2 имеет место неравенство Крафта $c = \sum_{i=1}^{k} 2^{-l_i} \leqslant 1$. Тогда числа $q_i = 2^{-l_i}/c$, $1 \leqslant i \leqslant k$, образуют распределение вероятностей на A. Имеем $L - H = \sum_{i=1}^{k} p_i(l_i + \log p_i) = \sum_{i=1}^{k} p_i \log \frac{p_i}{q_i} - \log c = D(p\|q) - \log c \geqslant 0$. Утверждение 1) доказано.

Полагаем $l_i = \lceil -\log p_i \rceil$. Тогда $\frac{1}{2} p_i < 2^{-l_i} \leqslant p_i$ для всех $1 \leqslant i \leqslant k$. Отсюда $\sum_{i=1}^{k} 2^{-l_i} \leqslant 1$. По теореме 1.1 можно построить безпрефиксный код длины кодовых слов которого равны $l_1, \ldots, l_k$. Из $\frac{1}{2} p_i < 2^{-l_i}$ следует, что $l_i \leqslant -\log p_i + 1$ и тогда

$L = \sum_{i=1}^{k} p_i l_i \leqslant \sum_{i=1}^{k} p_i(-\log p_i + 1) \leqslant H + \sum_{i=1}^{k} p_i = H + 1$. Утверждение 2) доказано. □

Код Шеннона. Код, построенный в утверждении 2, называется кодом Шеннона. Один из способов построения кода Шеннона указан в следующем примере.

Пример. Пусть $A = \{a, b, c, d\}$ и их вероятности $p(a) = \frac{1}{2}$, $p(b) = \frac{1}{4}$, $p(c) = \frac{1}{8}$, $p(d) = \frac{1}{8}$. Разбиваем эти буквы на два подмножества примерно равной вероятности, получаем $\{a\}$ и $\{b, c, d\}$. Кодовые слова букв из первого множества будут начинаться на 0, а второго на 1. То же самое проделываем с каждым из подмножеств (если оно делится), получаем $\{b\}$ и $\{c, d\}$. Второй бит кодовых слов букв из первого множества есть 0, а из второго 1. И так далее, продолжаем, пока не дойдем до одноэлементных подмножеств. В результате получаем безпрефиксный код $C(a) = 0$, $C(b) = 10$, $C(c) = 110$, $C(d) = 111$. Объясните, почему этим способом всегда получается код Шеннона.

Пример. Код Шеннона – оптимальный в среднем, но не поточечно.

Пусть $A = \{a, b\}$ и $p(a) = 2^{-10}$, $p(b) = 1 - 2^{-10}$. Тогда кодом Шеннона является код $C(a) = 0000000000$ и $C(b) = 1$. Средняя длина этого кода $L = 10 \cdot 2^{-10} + 1 - 2^{-10} = 1 - 9 \cdot 2^{-10} \approx 0.99$. В то же время, код $C(a) = 0$ и $C(b) = 1$ имеет более короткие кодовые слова и несколько большую среднюю длину $L = 2^{-10} + 1 - 2^{-10} = 1$.

Свойство сравнительной оптимальности кода Шеннона представлено в следующем предложении.

Предложение 1.4. *Пусть C – код Шеннона и C' – некоторый однозначно декодируемый код, $l(a) = l(C(a)) = \lceil -\log p(a) \rceil$ и $l'(a) = l(C'(a))$ – длины соответствующих кодовых слов, $a \in A$. Тогда $P\{l(a) \geqslant l'(a) + c\} \leqslant 2^{-c+1}$.*

Доказательство.

$$P\{l(a) \geqslant l'(a) + c\} = P\{\lceil -\log p(a) \rceil \geqslant l'(a) + c\} \leqslant$$
$$\leqslant P\{\frac{1}{p(a)} \geqslant l'(a) + c - 1\} = P\{p(a) \leqslant 2^{-l'(a)-c+1}\} =$$
$$= \sum_{a: p(a \leqslant 2^{-l'(a)-c+1}} p(a) \leqslant \sum_{a \in A} 2^{-l'(a)-c+1} \leqslant$$
$$\leqslant 2^{-c+1} \sum_{a \in A} 2^{-l'(a)} \leqslant 2^{-c+1}.$$

Последнее неравенство использует неравенство Крафта, которое имеет место ввиду однозначной декодируемости кода C'. $\square$

Свойство асимптотической равнораспределенности. Рассмотрим важный частный случай. Дана последовательность $X_1, X_2, \ldots$ независимых случайных величин со значениями в алфавите A. По слабому закону больших чисел

$$-\frac{p(X_1 \ldots X_n)}{n} = -\frac{1}{n} \sum_{i=1}^{n} \log p(X_i) \to E_{X_1 \sim p}[p(X_1)] = H,$$

где сходимость – по вероятности, т.е.

$$P\{\left| -\frac{p(X_1 \ldots X_n}{n} - H \right| > \epsilon\} \to 0$$

при $n \to \infty$, где $H = H(X_1)$.

Для произвольного $\epsilon > 0$ рассмотрим множество

$$A_\epsilon^n = \{(x_1 \ldots x_n) : 2^{-n(H+\epsilon)} \leqslant p(x_1 \ldots x_n) \leqslant 2^{-n(H-\epsilon)}.$$

Свойства множества A_ϵ^n представлены в следующей теореме.

Теорема 1.6. *1) Для любого $\epsilon > 0$ и для всех n выполнено*

$$H - \epsilon \leqslant -\frac{1}{n} \log p(x_1 \ldots x_n) \leqslant H + \epsilon$$

при $(x_1 \ldots x_n) \in A_\epsilon^n$

2) $P(A_\epsilon^n) \geqslant 1 - \epsilon$ для всех достаточно больших n.

3) $|A_\epsilon^n| \leqslant 2^{n(H+\epsilon)}$.

4) $|A_\epsilon^n| \geqslant (1 - \epsilon)2^{n(H-\epsilon)}$ для всех достаточно больших n.

Доказательство. Свойства 1) и 2) следуют из определения множества A_ϵ^n. Для доказательства свойства 3) рассмотрим

$$1 = \sum_{x \in A^n} p(x) \geqslant \sum_{x \in A_\epsilon^n} p(x) \geqslant$$
$$\geqslant \sum_{x \in A_\epsilon^n} 2^{-n(H+\epsilon)} = |A_\epsilon^n| 2^{-n(H+\epsilon)}.$$

Отсюда $|A_\epsilon^n| \leqslant 2^{n(H+\epsilon)}$.

Для доказательства 4) заметим, что $P(A_\epsilon^n) \geqslant 1 - \epsilon$ для всех достаточно больших n. Поэтому

$$1 - \epsilon \leqslant P(A_\epsilon^n) \leqslant \sum_{x \in A_\epsilon^n} 2^{-n(H-\epsilon)} = |A_\epsilon^n| 2^{-n(H-\epsilon)}.$$

Отсюда $|A_\epsilon^n| \geqslant (1-\epsilon) 2^{n(H-\epsilon)}$ для всех достаточно больших n. $\square$

Применим эти свойства для оптимального сжатия информации. Будем приписывать кодовые двоичные слова не буквам из алфавита A, а последовательностям этих букв (блокам) длины n. Припишем каждой последовательности $(x_1 \dots x_n) \in A_\epsilon^n$ двоичное слово $t(x_1 \dots x_n)$ длины $l(t(x_1 \dots x_n)) \leqslant \log |A_\epsilon^n| + 1$ и добавим к каждой такой последовательности префикс 0. Длина такого кодового слова $t(x_1 \dots x_n)$ не превосходит $l(t(x_1 \dots x_n)) \leqslant \log |A_\epsilon^n| + 2 \leqslant n(H + \epsilon) + 2$.

Остальные блоки длины n кодируем двоичными строками длины $\leqslant \log |A^n|$ с добавленной 1 в начале строки. Тогда $l(t(x_1 \dots x_n)) \leqslant n \log |A| + 2$ для такого блока.

Обозначаем $x^n = x_1 \dots x_n$. Средняя длина этого кода

$$L = \sum_{x^n \in A^n} p(x^n) l(t(x^n)) =$$
$$= \sum_{x^n \in A_\epsilon^n} p(x^n) l(t(x^n)) + \sum_{x^n \in A^n \setminus A_\epsilon^n} p(x^n) l(t(x^n)) \leqslant$$
$$\leqslant \sum_{x^n \in A_\epsilon^n} p(x^n)(n(H+\epsilon)+2) + \sum_{x^n \in A^n \setminus A_\epsilon^n} p(x^n) n \log |A| + 2 =$$
$$= P(A_\epsilon^n)(n(H+\epsilon)+2) + P(A^n \setminus A_\epsilon^n)(n \log |A| + 2) \leqslant$$
$$\leqslant n(H+\epsilon) + 2 + \epsilon \log |A| + 2\epsilon n(H+\epsilon').$$

Таким образом, для любого $\epsilon > 0$ для всех достаточно больших n существует безпрефиксный код $t(x^n)$ для блоков букв длины n такой, что

$$E\left[\frac{l(t(x^n))}{n}\right] \leqslant H + \epsilon.$$

1.3. Энтропия стационарного процесса

Пусть A конечный алфавит. Бесконечная последовательность случайных величин $X_1, X_2, \dots$ со значениями в A называется стационарной, если для любых n и s и любых $x_1, \dots, x_n \in A$ выполнено свойство инвариантности относительно сдвига

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} =$$
$$= P\{X_{1+s} = x_1, X_{2+s} = x_{2+s}, \dots, X_{n+s} = x_n\}.$$

Удельная энтропия стационарного процесса определяется

$$H_\infty = \lim_{n \to \infty} \frac{H(X_1, X_2, \dots, X_n)}{n}.$$

Мы докажем, что этот предел существует. По свойству энтропии $H(X_1, X_2, \dots, X_n) \leqslant n \log |A|$.

Теорема 1.7. *Для любого стационарного стохастического процесса*

$$H_\infty = \lim_{n\to\infty} H(X_n|X_{n-1},\dots X_1),$$

причем предел существует.

Доказательство. По свойству условной энтропии: $H(X|Y,Z) \leqslant H(X|Y)$ (см. задачу из раздела 1.4),

$$H(X_{n+1}|X_n,\dots X_1) \leqslant H(X_{n+1}|X_n,\dots X_2).$$

По свойству инвариантности относительно сдвига имеем

$$H(X_n|X_{n-1},\dots X_1) = H(X_{n+1}|X_n,\dots X_2).$$

Отсюда

$$H(X_{n+1}|X_n,\dots X_1) \leqslant H(X_n|X_{n-1},\dots X_1).$$

Таким образом, числовая последовательность $H(X_n|X_{n-1},\dots X_1)$ не убывает. Так как она ограничена, существует ее предел $\lim_{n\to\infty} H(X_n|X_{n-1},\dots X_1)$.

Будем использовать лемму Чезаро.

Лемма 1.1. *Для любой числовой последовательности* $a_1, a_2, \dots$, *если* $a_n \to a$ *при* $n \to \infty$, *то* $\frac{1}{n}\sum_{i=1}^n a_i \to a$.

По правилу цепи

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

По лемме Чезаро предел

$$H_\infty = \lim_{n\to\infty} \frac{H(X_1,\dots,X_n)}{n} = \lim_{n\to\infty} H(X_n|X_{n-1},\dots,X_1)$$

существует. □

Пример. При изучении частотных характеристик литературных текстов используется модель стационарных процессов.

Пусть $\alpha_1, \alpha_2, \ldots$ – стационарный процесс, который описывает процесс порождения литературного произведения. Измеряются основные характеристики такого процесса:
$H_1 = H(\alpha_1)$,
$H_2 = H(\alpha_2|\alpha_1) = H(\alpha_1\alpha_2) - H(\alpha_1)$,
$H_3 = H(\alpha_3|\alpha_2, \alpha_1) = H(\alpha_1, \alpha_2, \alpha_3) - H(\alpha_1, \alpha_2) - H(\alpha_1)$,
..
$H_n = H(\alpha_n|\alpha_{n-1}, \ldots, \alpha_1) = H(\alpha_1, \ldots, \alpha_n) - H(\alpha_1, \ldots, \alpha_1) - \cdots - H(\alpha_1)$,
..

По теореме 1.7 $H_n \searrow H_\infty$ при $n \to \infty$.

К.Шеннон [31] проводил опыты по измерению энтропии естественного английского языка и получил следующие значения:

$H_1 \approx 4.76$, $H_2 \approx 4.03$, $H_3 \approx 3.32$, $H_4 \approx 3.1$, ... , $H_6 \approx 1.9 \longleftarrow H_\infty \approx 1.3$.

Вычисление частот всех блоков достаточно большой длины затруднительно. Поэтому для оценки энтропии использовались игровые и психологические методы.

1.4. Задачи и упражнения

1. Доказать предложение 1.1.

2. Доказать, что не существует кода C такого, что $l(C(x)) < \lfloor \log |A| \rfloor$.[2]

3. Пусть $A = \{a_1, \ldots, a_k\}$ – алфавит, $p_i = P\{X = a_i\}$. Доказать, что а) $H(X) \leqslant \log |A|$, равенство достигается когда $p_i = \frac{1}{k}$ при $1 \leqslant i \leqslant k$.

б) $H(X) = 0$ тогда и только тогда, когда существует $a \in A$ такое, что $p(a) = 1$.

в) $H(P)$ – вогнутая по P.

г) $D(p\|q)$ – выпуклая по p и q.

4. Доказать, что $H(X, X) = H(X)$, $H(Y|X) \leqslant H(Y)$, $I(X : X) = H(X)$.

[2] $\lfloor r \rfloor$ обозначает целую часть вещественного числа r.

5. Доказать, что $H(Y|X) = 0$ тогда и только тогда Y есть функция от X: $Y = g(X)$.

6. Пусть $g : A \to A$ – произвольная функция. Доказать, что

а) $H(X) \leqslant H(g(X)$. Для каких g имеет место равенство.

б) $H(X|g(Y)) \geqslant H(X|y)$.

в) $I(X : Y) \geqslant I(X : g(Y))$

7. Доказать, что $H(X) = \log |A| - D(p\|\mu)$, где $X \sim p$ – распределена по p, μ – равномерное распределение на A: $\mu(x) = \frac{1}{|A|}$.

8. Докажите, что $H(X|Y) \leqslant H(X)$ и $H(X|Y, Z) \leqslant H(X|Y)$.

Глава 2

Универсальное сжатие информации

Рассмотренные выше алгоритмы кодирования используют для построения кода распределение вероятностей источника. В этой главе мы рассмотрим алгоритм универсального сжатия информации, который не использует в своей работе никаких предположений об источнике данных. Тем не менее, для доказательства оптимальности этого алгоритма необходимо принять предположение о том, что данные генерируются некоторым стационарным эргодическим процессом. При этом, знание конкретного распределения вероятностей этого процесса не требуется.

Стационарные эргодические процессы – максимально широкий класс процессов, для которых выполнены вероятностные законы.

2.1. Алгоритм Зива–Лемпеля

В этом разделе мы изучим алгоритм Зива–Лемпеля универсального сжатия информации [63], [64], [40].

На вход кодирующему алгоритму подается слово, составленное из букв конечного алфавита. Алгоритм LZ читает это слово

слева направо. В процессе чтения алгоритм производит разбиение входного слова на подслова, разделенные запятыми (парсинг), и одновременно по этим подсловам формирует кодирующую последовательность.

Декодирующий алгоритм восстанавливает исходную последовательность букв проходя и читая слева направо кодирующую последовательность.

Мы приведем один из вариантов алгоритма LZ. Параметр алгоритма: W – длина окна. Вход алгоритма: строка букв $x = x_1 x_2 \dots x_n$

LZ-алгоритм.

WHILE $i \leqslant n$

Пусть подстрока $x^{i-1} = x_1 \dots x_{i-1}$ уже обработана на предыдущих шагах и представлена в виде набора подстрок, разделенных запятой.

Находим наибольшее k такое, что $\exists j (i - 1 - W \leqslant j \leqslant i - 1)$ и подстрока длины k, начинающаяся с x_j, т.е. $x_j x_{j+1} \dots x_{j+k-1}$ совпадает с подстрокой $x_i x_{i+1} \dots x_{i+k-1}$ ($x_{j+s} = x_{i+s}$ при $0 \leqslant s \leqslant k - 1$).

Выделяем подстроку $x_i x_{i+1} \dots x_{i+k-1}$ запятыми и кодируем ее тройкой (F, P, L), где $F = 1$ – индикатор типа кодирования, $P = i - j$ – координата начала от i влево, где j – максимальное из существующих, $L = k$ – длина подстроки.

Если такое k не найдется, то кодируем x_i парой (F, C), где $F = 0$ и $C = x_i$.

END

Таким образом, кодирующая последовательность состоит из пар и троек (F, C) и (F, P, L). Легко построить декодирующий алгоритм.

Для простоты мы не учитываем конец входного слова – считаем, что входной поток букв никогда не заканчивается. Чтобы учесть конец слова, можно дополнительно проверять при поиске k условие $i + k - 1 \leqslant n$.

Пример. Входное слово ABBABBABBBAABABA – слово в алфавите $\{A, B\}$, длина окна $W = 4$. Алгоритм производит

парсинг и кодирует:

A B B A B B A B B B A A B A B A

A,B,B,A B B A B B,B A,A,B A,B A

(0.A),(0,B),(1,1,1),(1,3,6),(1,4,2),(1,1,1),(1,3,2),(1,2,2)

Например, первые две буквы A и B ранее не встречались, поэтому кодируем их парами (0.A) и (0,B). Третья буква B встречается раньше (соседняя буква B слева), а ее продолжение BA раньше не встречалось, поэтому кодируем эту B тройкой (1,1,1). Двигаясь вправо по входному слову обнаруживаем, что подстроку A B B A B B можно отложить от третьей слева буквы слова (а ее продолжение уже нельзя), поэтому кодируем ее тройкой (1,3,6) и т.д.

Оптимальность LZ-сжатия. Полный анализ алгоритма LZ технически сложен. Мы рассмотрим некоторую упрощенную схему: входное слово бесконечно влево и вправо. Точнее, задан стационарный эргодический процесс

$$\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$$

Алгоритм начинает сжатие с буквы X_0, при этом, $\dots, X_{-2}, X_{-1}$ – известная история, длина окна не ограничена.

Под стационарностью мы понимаем инвариантность относительно сдвига

$$P\{X_i = a_1 \dots X_{i+k} = a_k\} = P\{X_{i+s} = a_1 \dots X_{i+k+s} = a_k\}$$

для любых i, k, $a_1, \dots a_k$ и $-\infty < s < \infty$.

Под эргодичностью понимаем следующее свойство: частота любой комбинации $a_1, \dots a_k$ на отрезке длины n почти всюду сходится к вероятности этой комбинации при $n \to \infty$.[1]

[1]Мы также предполагаем, что вероятность любой комбинации $a_1, \dots a_k$, $k \geqslant 1$, положительна

Рассмотрим упрощенный вариант алгоритма. Так как любое подслово стационарной и эргодической последовательности встречается ранее, при заданном n будем последовательно кодировать блоки входного слова длины n.

Изучим асимптотическое поведение длины кодового слова блока длины n при $n \to \infty$. Введем случайную величину

$$R_n(X_0, X_1, \ldots, X_{n-1}) = \\ = \max_{j>0}\{X_{-j}X_{-j+1}\ldots X_{-j+n-1} = X_0X_1\ldots X_{n-1}\}.$$

Из свойства эргодичности эта величина конечная почти всюду.

Согласно алгоритму LZ мы будем кодировать слово $X_0^n = X_0X_1\ldots X_{n-1}$ тройкой $(1, j, n)$. Закодируем эту тройку двоичной последовательностью длины $L_n(X_0^n) = \log R_n + 2\log\log R_n + 3$, где $R_n = R_n(X_0, X_1, \ldots, X_{n-1}) = j$. Про способы кодирования натуральных чисел см. разделы 3.1.1 и 5.1 далее.

В дальнейших рассуждениях решающую роль играет лемма Каца. Дадим необходимые определения.

Пусть A – счетный алфавит и $\ldots, U_1, U_0, U_1, \ldots$ – стационарный процесс с значениями в A. Тогда при $u \in A$ рассмотрим величину

$$Q_u(i) = P\{U_{-i} = u,\ U_j \neq u \text{ при } -i < j < 0 | U_0 = u\}.$$

Эта величина представляет собой условную вероятность того, что наблюдаемая в нулевой момент времени буква u в ближайшем прошлом наблюдалась i шагов тому назад.

Выше была определена случайная величина R_n, в частности, $R_1(u) = \min_{j>0} U_{-j} = u$. Тогда $E[R_1(U)|U_0 = u] = \sum_{i=1}^{\infty} iQ_u(i)$ – среднее время ближайшего появления наблюдаемой буквы u в прошлом. Обозначим также $p(u) = P\{U_0 = u\}$. Мы предположили, что $p(u) > 0$ для всех $u \in A$.[2]

Лемма 2.1. $E[R_1(U)|U_0 = u] = \frac{1}{p(u)}$.

[2] В противном случае букву можно удалить из A.

Доказательство. Пусть $u \in A$. Определим случайное событие

$$A_{j,k} = \{U_{-j} = u,\ U_i \neq u \text{ при } -j < i < k, U_k = u\}.$$

Из определения $A_{j,k} \bigcap A_{j',k'} = \emptyset$ при $(j,k) \neq (j',k')$ и $P\left(\bigcup_{j,k} A_{j,k}\right) = 1$. Представим вероятность этого объединения в виде суммы вероятностей попарно несовместимых событий

$$1 = P\left(\bigcup_{j,k} A_{j,k}\right) = \sum_{j=1}^{\infty}\sum_{k=0}^{\infty} P(A_{j,k}) =$$

$$= \sum_{j=1}^{\infty}\sum_{k=0}^{\infty} P(U_k = u) P\{U_{-j} = u,\ U_i \neq u \text{ при } -j < i < k | U_k = u\} =$$

$$= \sum_{j=1}^{\infty}\sum_{k=0}^{\infty} P(U_k = u) Q_u(j+k) = P\{U_0 = u\} \sum_{i=1}^{\infty} i Q_u(i).$$

Здесь мы использовали стационарность процесса, также то, что имеется i различных пар (j,k) таких, что $j + k = i$. Отсюда $E[R_1(u)|U_0 = u] = \frac{1}{p(u)}$. □

Из леммы $E[R_1(u)] = \sum_{u \in A} \frac{1}{p(u)} p(u) = |A|$ – среднее время вторичного появления какой-либо буквы.

Можно распространить эту лемму на последовательности букв.

Следствие 2.1. *Пусть $\ldots, X_{-1}, X_0, X_1, \ldots$ – стационарный эргодический процесс, значения которого принадлежат конечному алфавиту. Тогда для для любых $x_0, x_1, \ldots, x_{n-1} \in A$ будет*

$$E[R_n(X_0, X_1, \ldots, X_{n-1}) | X_0 X_1 \ldots X_{n-1} = x_0 x_1 \ldots x_{n-1}] = \\ = \frac{1}{p(x_0 x_1 \ldots x_{n-1})}.$$

Доказательство. Определим стационарный эргодический процесс $U_i = (X_i, X_{i+1}, \ldots, X_{i+n-1})$, $-\infty < i < \infty$, и применим лемму 2.1. □

Обозначим

$$H_n = -\frac{1}{n} \sum_{x_0^{n-1} \in A^n} p(x_0^{n-1}) \log p(x_0^{n-1}),$$

где $x_0^{n-1} = x_0 x_1 \dots x_{n-1}$. Энтропия стационарного процесса (источника) равна

$$H_\infty = \lim_{n\to\infty} \frac{H_n}{n}.$$

Обозначим $X_0^{n-1} = X_0 X_1 \dots X_{n-1}$. Также $L_n(X_0^{n-1})$ –длина двоичной последовательности, которая кодирует тройку $(1, j, n)$.

Приведем теперь теорему об оптимальности сжатия алгоритмом LZ.

Теорема 2.1. $\lim_{n\to\infty} \frac{E[L_n(X_0^{n-1}]}{n} = H_\infty$. [3]

Доказательство. Так как L_n однозначно декодируемый код, по теореме 1.5

$$E[L_n(X_0^{n-1})] \geqslant nH_n \tag{2.1}$$

для всех n. Ранее было отмечено, что можно кодировать тройки так, что

$$L_n(X_0^{n-1}) = \log R_n(X_0^{n-1}) + 2 \log\log R_n(X_0^{n-1}) + 3. \tag{2.2}$$

Докажем, что $\limsup_{n\to\infty} \frac{E[\log R_n(X_0^{n-1})]}{n} \leqslant H_\infty$. Действитель-

[3] Код, удовлетворяющей этому условию, называется универсальным для класса всех стационарных эргодических источников.

но,

$$\frac{E[\log R_n(X_0^{n-1})]}{n} =$$

$$= \frac{1}{n} \sum_{x_0^{n-1} \in A^n} p(x_0^{n-1}) E[\log R_n(X_0^{n-1}) | X_0^{n-1} = x_0^{n-1}] \leqslant \qquad (2.3)$$

$$\leqslant \frac{1}{n} \sum_{x_0^{n-1} \in A^n} p(x_0^{n-1}) \log E[R_n(X_0^{n-1}) | X_0^{n-1} = x_0^{n-1}] = \qquad (2.4)$$

$$= \frac{1}{n} \sum_{x_0^{n-1} \in A^n} p(x_0^{n-1}) \log \frac{1}{p(x_0^{n-1})} = \frac{1}{n} H_n(X_0^{n-1}) \to H_\infty \qquad (2.5)$$

при $n \to \infty$. Здесь переход от (2.3) к (2.4) происходит по неравенству Иенсена, переход от (2.4) к (2.5) происходит по следствию 2.1 к лемме Каца.

Для второго слагаемого из (2.2) имеем

$$\frac{E[\log \log R_n(X_0^{n-1})]}{n} \leqslant \frac{1}{n} \log E[\log R_n(X_0^{n-1})] \leqslant \frac{1}{n} \log H_n(X_0^{n-1}).$$

Из существования предела (2.5) следует, что для любого $\epsilon > 0$ будет $H_n(X_0^{n-1}) \leqslant n(H_\infty + \epsilon)$ для всех достаточно больших n. Отсюда $\limsup_{n\to\infty} \frac{1}{n} \log \log R_n(X_0^{n-1}) = 0$ и по (2.1) и (2.5) $\lim_{n\to\infty} \frac{E[\log R_n(X_0^{n-1})]}{n} = H_\infty$. Отсюда следует, что $\lim_{n\to\infty} \frac{E[L_n(X_0^{n-1}]}{n} = H_\infty$. $\square$

Часть II

Алгоритмическая теория информации

Глава 3

Простая колмогоровская сложность

В этой главе дается определению простой колмогоровской сложности и доказываются ее основные свойства. Излагается финитарный подход А. Н. Колмогорова к определению случайного индивидуального конечного объекта. Показано, что стохастические свойства конечного объекта являются следствиями его сложностных характеристик.

3.1. Основные понятия теории алгоритмов

Определения колмогоровской сложности и алгоритмической случайности основываются на использовании общей теории алгоритмов, которая называется также теорией рекурсивных функций, а также ее основного результата – теоремы о существовании универсальной функции.

В этом разделе мы обсудим основные понятия теории алгоритмов и приведем идею построения универсальной функции. Классическое пособие по теории алгоритмов – монография Роджерса [21].

3.1.1. Конструктивные объекты

Алгоритмы применяются к конструктивным объектам и в качестве значений также выдают конструктивные или конечные объекты. Понятие *конструктивного объекта* является исходным в данном изложении и не будет иметь точного математического определения. Свойства конструктивных объектов подробно обсуждаются в книге [26].

Типичными примерами конструктивных объектов в нашем понимании являются слова в некотором конечном алфавите

$$A = \{a_1,\ \ldots,\ a_k\},$$

где $k \geqslant 1$. Алфавит состоит из букв $a_1,\ \ldots,\ a_k$. Буква a_i – это неделимый символ, который не будет иметь точного определения. При задании алфавита обычно задается некоторый линейный порядок на его буквах – у нас он задается простой нумерацией этих букв: $a_1 < \cdots < a_k$. В дальнейшем этот порядок используется при определении лексикографического порядка на словах алфавита A. Слово в алфавите A – это конечная последовательность букв $x = x_1 \ldots x_n$ этого алфавита, т.е. $x_i \in A$ при $i = 1,\ \ldots,\ n$. Множество всех слов в алфавите A обозначается символом A^*.

Длина слова x обозначается $l(x) = n$ и равна числу букв в этом слове. Удобно рассматривать пустое слово λ, которое не содержит ни одной буквы. Его длина равна нулю: $l(\lambda) = 0$.

Конкатенацией двух слов $x = x_1 \ldots x_n$ и $y = y_1 \ldots y_m$ называется слово $xy = x_1 \ldots x_n y_1 \ldots y_m$, длина которого равна сумме длин слов x и y. По определению $x\lambda = \lambda x = x$ для любого слова x.

Слово x является префиксом слова y, обозначается $x \subseteq y$, если $y = xz$ для некоторого слова z. Если слово z непустое, то пишем $x \subset y$. Для слова $x = x_1 \ldots x_n$ его префикс длины $m \leqslant n$ обозначаем $x^m = x_1 \ldots x_m$.

В теории информации широко используется двоичный алфавит $I = \{0, 1\}$. Символы 0 и 1 называются битами. Выделение

такого алфавита связано со способом хранения информации в памяти компьютера. Слова в алфавите I называются двоичными (бинарными) последовательностями или строками. Обозначаем множество всех двоичных слов $\Xi = \{0,1\}^*$.

Множество всех натуральных чисел $\mathcal{N} = \{1,\ 2,\ \dots\}$ также является множеством конструктивных объектов. Часто для удобства мы будем присоединять число 0 к натуральным числам. Конструктивная природа натуральных чисел связана с тем, что они представляются в памяти компьютера в виде слов в некотором алфавите. Например, натуральные числа можно представлять в виде слов в унарном алфавите $U = \{1\}$: последовательность $11\dots1$ из n единиц представляет число $n \in \mathcal{N}$. Такое представление является неэкономным – длина унарной записи числа n равна n.

Экспоненциальное уменьшение длины записи числа происходит при использовании неодноэлементного алфавита. Мы будем использовать стандартное представление натуральных чисел с помощью двоичных строк, которое определяется следующим образом. Для удобства мы включим число 0 в это соответствие.

Пусть $\mathrm{bin}(n+1) = 1\nu_{k-1}\dots\nu_0$ представляет собой запись числа $n+1$ в двоичной форме:

$$n+1 = 2^k + \nu_{k-1}2^{k-1} + \dots + \nu_1 2^1 + \nu_0 2^0.$$

Сопоставляем натуральному числу n строку $\mathrm{str}(n) = \nu_{k-1}\dots\nu_0$. Заметим, что каждая строка соответствует некоторому натуральному числу, пустая строка соответствует числу 0.

Имеет место неравенство $2^k \leqslant n+1 < 2^{k+1}$, и поэтому будет $k = \lfloor \log_2(n+1) \rfloor$. Отсюда длина строки, сопоставленной числу n, равна $l(\mathrm{str}(n)) = \lfloor \log(n+1) \rfloor \leqslant \log n + 1$ при $n \geqslant 1$ [1] .

Пример сопоставления указан в таблице.

[1] Здесь и далее $\lfloor r \rfloor$ обозначает целую часть вещественного числа r, $\log n$ обозначает двоичный логарифм n, $\ln n$ – натуральный логарифм.

Число n	Строка $\mathrm{str}(n)$	Число $n+1$	Дв. зап. $\mathrm{bin}(n+1)$
0	λ	1	1
1	0	2	10
2	1	3	11
3	00	4	100
4	01	5	101
5	10	6	110
6	11	7	111
7	000	8	1000
...	...	...	...

Отождествляем натуральное число n и его запись в виде строки $\mathrm{str}(n)$. Преимуществом такого представления по сравнением с двоичным представлением натуральных чисел является то, что оно является взаимно однозначным соответствием между множеством $\mathcal{N}$ всех натуральных чисел и множеством $\Xi = \{0,1\}^*$ всех конечных двоичных последовательностей, тогда как не всякая двоичная последовательность является двоичной записью некоторого натурального числа.

В дальнейшем будем широко использовать такое соответствие. Когда будет удобно, не будем различать натуральное число n и его запись $\mathrm{str}(n)$. С вычислительной точки зрения использование множества Ξ вместо $\mathcal{N}$ более естественно, так как алгоритмы работают со словарными представлениями натуральных чисел.

Пары строк могут кодироваться строками различным образом. Мы не можем рассматривать конкатенацию двух строк $x = x_1 \dots x_n$ и $y = y_1 \dots y_m$ как код пары (x, y), так как по ней невозможно однозначным образом разделить элементы пары. Поэтому необходимо затратить дополнительную информацию для разделения пары на ее элементы. Например, это удобно делать следующим образом. Для произвольной строки $x = x_1 \dots x_n$ обозначим $\overline{x}$ строку, в которой все биты повторены по два раза: $\overline{x} = x_1 x_1 \dots x_n x_n$. Тогда сопоставляем паре строк (x, y) последовательность $\overline{x}01y = x_1 x_1 \dots x_n x_n 01 y_1 \dots y_m$. Легко видеть, что в

этом случае существует алгоритм, который восстанавливает элементы пары x и y по последовательности $\overline{x}01y$. При этом длина кода равна $l(\overline{x}01y) = 2l(x) + l(y) + 2$.

Основываясь на той же идее, можно устроить и более экономное кодирование пар (x, y). Сопоставим паре (x, y) строку

$$\overline{\mathrm{str}(l(x))}01xy,$$

которая состоит из удвоенной двоичной записи $\overline{\mathrm{str}(l(x))}$ длины строки x, разделителя 01 и строк x и y, записанных подряд. Ясно, что по этому коду можно однозначно восстановить x и y, при этом длина кодирующей последовательности равна

$$\begin{aligned} l(x) + l(y) + 2l(\mathrm{str}(l(x)) + 2 \leqslant \\ \leqslant l(x) + l(y) + 2\log l(x) + 3. \end{aligned}$$

Можно продолжить идею такой экономии и построить кодирование пары (x, y) с помощью последовательности

$$\overline{\mathrm{str}(l(\mathrm{str}(l(x))))}01\mathrm{str}(l(x))xy.$$

Длина кодирующей последовательности равна

$$\begin{aligned} l(x) + l(y) + l(\mathrm{str}(l(x))) + 2l(\mathrm{str}(l(\mathrm{str}(l(x)))) + 2 \leqslant \\ \leqslant l(x) + l(y) + \log l(x) + \log\log l(x) + 4 \end{aligned}$$

и т.д. В одной из задач из раздела 3.6 утверждается, что величину $\log l(x)$ нельзя устранить из этих верхних оценок.

В дальнейшем под парой (x, y) двоичных строк будет пониматься строка, кодирующая эту пару одним из приведенных выше способов.

Аналогичным образом можно кодировать тройки (x, y, z), если записывать их в виде $(x, (y, z))$. И так далее.

Кроме множества Ξ, отождествленного с множеством $\mathcal{N}$, мы будем использовать множество $\mathcal{Q}$ всех рациональных чисел и множество $\mathcal{R}$ всех вещественных чисел.

Рациональные числа можно естественным образом занумеровать парами натуральных чисел (и битом знака) и тем самым двоичными строками. Не будем останавливаться на деталях такой нумерации.

Вещественные числа не являются конструктивными объектами, хотя бы потому, что их множество несчетно. Поэтому алгоритмы не будут работать непосредственно с вещественными числами. Вместо этого они будут применяться к их рациональным приближениям.

3.1.2. Алгоритмы

В качестве основной модели алгоритма будет использоваться понятие машины Тьюринга (МТ). *Машина Тьюринга* представляется в виде ленты, неограниченно расширяемой в обе стороны, и головки. Лента разделена на ячейки, в каждую из которых головка может записывать символ некоторого входного алфавита. На этом же алфавите записывается выходное слово МТ. Машина Тьюринга задается набором $(A, \Gamma, Q, \delta, q_0, q_K)$, где

- A – основной алфавит, на котором задается входное слово МТ и записывается результат;
- Γ – ленточный или рабочий алфавит, который используется для вычислений МТ, при этом $A \subseteq \Gamma$;
- Q – множество внутренних состояний или память головки; в процессе работы головка может запоминать ограниченную по объему информацию с помощью своих состояний $q \in Q$; говорят также, что МТ находится в состоянии q;
- $\delta(q, a)$ – функция переходов, где $q \in Q$ и $a \in \Gamma$; ее значения – тройки $\delta(q, a) = (q', a', M)$, где $q' \in Q$, $a' \in \Gamma$ и $M \in \{R, L, S)$, которые называются командами МТ; команды интерпретируются следующим образом: если МТ находится в состоянии q и читает на ленте букву a, то команда $\delta(q, a) = (q', a', M)$ дает указание стереть букву a, записать

вместо нее букву a', изменить текущее состояние q головки на q' и переместить головку влево, если $M = L$, сдвинуться вправо, если $M = R$, или оставаться над той же ячейкой, если $M = S$;

- q_0 – начальное состояние головки, при этом головка устанавливается над самым левым символом входного слова;
- q_K – заключительное состояние; как только головка МТ первый раз перейдет в состояние q_K, машина прекращает работы, при этом слово, которое записано на ленте, считается результатом ее работы.

Текущее состояние работы МТ описывается конфигурацией (мгновенным описанием) – словом

$$x_1 \dots x_s q a y_1 \dots y_k, \tag{3.1}$$

где q – текущее состояние головки, которая обозревает символ a на ленте; справа и слева от q находятся все символы, находящиеся на ленте. Один шаг работы МТ – это переход от одной конфигурации к непосредственно следующей конфигурации. Например, если текущее состояние МТ описывается конфигурацией (3.1) и выполняется команда $\delta(q, a) = (q', a', R)$, то происходит переход к следующей конфигурации:

$$x_1 \dots x_s q a y_1 \dots y_k \Rightarrow x_1 \dots x_s a' q' y_1 \dots y_k. \tag{3.2}$$

В этом заключается один шаг работы МТ.

МТ вычисляет некоторую функцию $\psi : A^* \to A^*$ следующим образом. Перед запуском МТ на ленте записан аргумент – входное слово x, над первой буквой которого установлена головка МТ, находящаяся в начальном состоянии q_0. Как правило, конец входного слова отмечен специальным символом – маркером конца входного слова. Если в процессе вычисления МТ переходит в заключительное состояние q_K, то в качестве значения функции $\psi(x)$ берется часть содержимого ленты – от символа, отмеченного состоянием q_K, до маркера конца выходного слова. Если одно

из этих правил нарушено или МТ никогда не приходит в состояние q_K, то значение функции ψ на аргументе x не определено.

Функция $\psi : A^* \to A^*$ называется вычислимой, если существует МТ, которая вычисляет значения этой функции.

Характерной особенностью процесса вычисления МТ является то, что на некоторых входных словах МТ может никогда не достичь конечного состояния. На таких словах соответствующая функция не определена. В этом случае вычислимая функция является частично определенной.

Описанные выше машины Тьюринга являются «специализированными». Каждая такая машина работает только с одной программой. Можно построить «универсальную» МТ, т.е. такую, которая может интерпретировать работу любой программы. Точнее, фиксируем входной алфавит A и рассматриваем все МТ, которые вычисляют функции типа $A^* \to A^*$. В частности, можно рассмотреть $A = \{0, 1\}$ и считать, что рассматриваются все вычислимые функции, аргументы и значения которых – натуральные числа.

Пусть символ # не принадлежит алфавиту A. Каждую программу для вычисления функции $\psi : A^* \to A^*$ можно закодировать словом p в алфавите A. Слово содержит закодированную последовательность команд программы МТ, вычисляющей значения функции ψ. Можно также написать программу–интерпретатор, на вход которой подаются слова $p\#x$, где p – код некоторой программы, а $x \in A^*$ – входное слово для этой программы. Интерпретатор может использовать более широкий ленточный алфавит. Схема работы программы интерпретатора следующая: на каждом шаге работы интерпретатора на ленте записана конфигурация моделируемой МТ; интерпретатор декодирует команды из слова p и выполняет необходимые переходы типа (3.2) от одной конфигурации к другой, которые предписываются этими командами. Таким образом, интерпретатор вычисляет некоторую функцию $U(p, x)$, где $p, x \in A^*$, обладающую свойством:

- для любой вычислимой функции $\psi : A^* \to A^*$ найдется

$p \in A^*$ такое, что $\psi(x) = U(p, x)$ для всех x [2] .

Функция $U(p, x)$, обладающая этим свойством, называется *универсальной функцией* для всех частично определенных вычислимых функций типа $A^* \to A^*$.

Учитывая отождествление произвольной пары конечных последовательностей (x, y) с последовательностью $\overline{x}01y$ (или, более экономно, с последовательностью $\overline{\mathrm{str}(l(x))}01xy$), можно рассматривать функцию $U(p, x, y)$, универсальную для всех вычислимых функций $B(x, y)$ от двух аргументов: для любой такой вычислимой функции $B(x, y)$ существует такое p, что $B(x, y) = U(p, x, y)$ для всех x, y. Аналогичным образом можно рассматривать функции, универсальные для всех вычислимых функций от любого заданного числа аргументов.

В дальнейшем мы будем использовать некоторые эффективные (алгоритмические) свойства множеств слов заданного алфавита.

Множество конструктивных объектов называется перечислимым, если либо оно пусто, либо является множеством значений некоторой вычислимой функции.

Предложение 3.1. *Область определения любой вычислимой функции $f(x)$ является перечислимым множеством.*

Доказательство. Допустим, что область определения функции $f(x)$ – непустое множество. Построим алгоритм для вычисления функции $g(n)$, перечисляющей область определения функции $f(x)$. Предварительно определим $g(0) = a$, где a – какой-либо элемент из области определения функции f.

Запускаем процесс одновременного вычисления всех значений $f(x)$ на всех возможных входах x, делая на каждом шаге нашего процесса один шаг вычисления значения $f(x)$ только для одного из таких x. Если на шаге n нашего моделирования значение $f(x)$ впервые определилось, полагаем $g(n) = x$. В противном случае полагаем $g(n) = g(n - 1)$. □

[2] Здесь имеется в виду, что обе части этого равенства определены или не определены одновременно.

Заметим, что тривиальным образом верно и утверждение, обратное к предложению 3.1, а именно: произвольное перечислимое множество C является областью определения вычислимой функции

$$\xi(a) = \begin{cases} 1, \text{ если } a \in C, \\ \text{неопределено в противном случае.} \end{cases}$$

Пусть

$$W_p = \{x : U(p, x) \text{ определена}\}.$$

Из определения и предложения 3.1 следует, что

- W_p – перечислимое множество для любого p;
- для любого перечислимого множества C найдется такое p, что $C = W_p$;
- множество $\{(p, x) : x \in W_p\}$ перечислимо.

Эти свойства означают, что имеется алгоритм, который «равномерно» перечисляет все перечислимые множества.

Множество конструктивных объектов C называется разрешимым, если его характеристическая функция

$$\xi(a) = \begin{cases} 1, \text{ если } a \in C, \\ 0, \text{ если } a \notin C \end{cases}$$

является вычислимой.

Легко видеть, что всякое разрешимое множество, а также его дополнение являются перечислимыми. Обратное утверждение неверно. Соответствующие примеры строятся с помощью универсальной функции $U(p, x)$. Область определения функции $U(p, x)$ называется универсальным множеством.

Предложение 3.2. *Универсальное множество*

$$\{(p, x) : U(p, x) \textit{ определено}\} \tag{3.3}$$

перечислимо, но не разрешимо.

Доказательство. Множество (3.3) перечислимо как область определения вычислимой функции.

Вторую часть утверждения теоремы докажем методом от противного. Допустим, что множество (3.3) разрешимо, т.е. функция

$$\xi(p,x) = \begin{cases} 1, \text{ если } U(p,x) \text{ определено,} \\ 0 \text{ в противном случае} \end{cases}$$

является вычислимой. Тогда функция

$$\theta(p) = \begin{cases} 0, \text{ если } \xi(p,p) = 0, \\ \text{неопределено, если } \xi(p,p) = 1, \end{cases}$$

также является вычислимой. Поэтому существует такое q, что $\theta(p) = U(q,p)$ для всех p.

Изучим, что происходит при $p = q$. Если $U(q,q)$ определено, то $\xi(q,q) = 1$. В этом случае значение $\theta(q)$ не определено. Выполнено $\theta(q) = U(q,q)$, значит, и $U(q,q)$ не определено. Получаем противоречие.

Пусть значение $U(q,q)$ не определено. Тогда $\xi(q,q) = 0$. В этом случае $\theta(q) = 0$, т.е. это значение определено. Одновременно $\theta(q) = U(q,q)$, т.е. значение $U(q,q)$ также определено. Опять получаем противоречие.

Значит, исходное предположение о том, что функция $\xi(p,x)$ вычислимая, или то же самое, что множество (3.3) разрешимое, неверно. □

Заметим, что предложение 3.2 эквивалентно тому, что не существует алгоритма, который решает вопрос о том, остановится ли произвольная программа p на входе x или нет. Задача построения такого алгоритма называется «проблемой остановки».

3.2. Определение колмогоровской сложности

А. Н. Колмогоров в статье [11] предложил измерять сложность конечного объекта x при заданном конечном объекте y длиной

самой короткой последовательности p (программы для x), состоящей из 0 и 1, по которой некоторой способ декодирования B может восстановить x, используя в качестве дополнительной информации слово y. Математически это записывается следующим образом:

$$\mathrm{K}_B(x|y) = \min\{l(p) : B(p,y) = x\},$$

где $l(p)$ – длина последовательности $p \in \{0,1\}^*$, а $B(p,y)$ – некоторая вычислимая функция. Мы считаем, что $\min \emptyset = \infty$. Называем функцию $\mathrm{K}_B(x|y)$ мерой сложности относительно способа декодирования $B(p,y)$.

Учитывая то, что мы можем кодировать любые конструктивные объекты двоичными строками, можно предполагать, что $x, y \in \{0.1\}^*$.

Сформулированное выше определение сложности зависит от вычислимой функции $B(p,y)$ – способа декодирования конечных объектов [3] . Однако использование основного результата теории алгоритмов – теоремы о существовании универсальной функции – позволило Колмогорову определить сложность независимо от способа декодирования $B(p,y)$.

Имеет место основная теорема теории алгоритмической сложности – теорема инвариантности.

Теорема 3.1. *Существует такая вычислимая функция $A(p,y)$, что для любой вычислимой функции $B(p,y)$ имеет место неравенство*

$$\mathrm{K}_A(x|y) \leqslant \mathrm{K}_B(x|y) + c, \tag{3.4}$$

где c – некоторая константа, не зависящая от x и y.

Доказательство. Пусть $U(q,p,y)$ – функция, универсальная для всех вычислимых функций $B(p,y)$ от двух аргументов.

[3] В отличие от теории информации, при определении колмогоровской сложности рассматриваются способы декодирования, для которых могут не существовать соответствующие способы кодирования. Кроме этого, такие способы декодирования могут быть не всюду определенными функциями.

Определим «универсальный» способ декодирования:

$$A(\overline{q}01p, y) = U(q, p, y)$$

для всех $p, q, y \in \{0,1\}^*$. Для всех остальных входов, не имеющих вида $(\overline{q}01p, y)$, значение функции A не определено.

Пусть $B(p, y)$ – произвольная вычислимая функция, представляющая некоторый способ декодирования. По определению универсальной функции для некоторого q имеет место равенство $B(p, y) = U(q, p, y)$ для всех p и y. Допустим, что p – самый короткий код для строки x при способе описания B и дополнительной информации y. Для него выполнено $B(p, y) = x$. Тогда

$$A(\overline{q}01p, y) = U(q, p, y) = B(p, y) = x.$$

Сравниваем длины кратчайших кодов для x при способах декодирования A и B:

$$\mathrm{K}_A(x|y) \leqslant \mathrm{K}_B(x|y) + 2l(q) + 2.$$

Соответствующая константа c имеет вид $c = 2l(q) + 2$. Теорема доказана. □

В доказательстве теоремы 3.1 используется следующая схема универсального декодирования:

$$q : p, y \longrightarrow x,$$
$$A : \overline{q}01p, y \longrightarrow x.$$

Здесь $\overline{q}01p$ – архив, который содержит программу q декодирования x по p и дополнительной информации y.

Функция $A(p, y)$, определенная в доказательстве теоремы 3.1, называется *оптимальным способом декодирования.*

По теореме 3.1 для любых двух оптимальных способов декодирования A_1 и A_2 для всех x и y выполнено

$$|\mathrm{K}_{A_1}(x|y) - \mathrm{K}_{A_2}(x|y)| \leqslant c, \tag{3.5}$$

где c – некоторая константа (зависящая от A_1 и A_2).

Мы также записываем (3.4) в виде [4]

$$\mathrm{K}_A(x|y) \leqslant \mathrm{K}_B(x|y) + O(1),$$

а (3.5) – в виде

$$\mathrm{K}_{A_1}(x|y) = \mathrm{K}_{A_2}(x|y) + O(1).$$

Фиксируем одну такую оптимальную функцию $A(p, y)$ и обозначим

$$\mathrm{K}(x|y) = \mathrm{K}_A(x|y).$$

Назовем функцию $\mathrm{K}(x|y)$ *условной колмогоровской сложностью* слова x при известном y.

Определим *безусловную колмогоровскую сложность*

$$\mathrm{K}(x) = \mathrm{K}(x|\lambda)$$

конечного объекта x. В этом случае оптимальный способ декодирования для восстановления x не использует никакой дополнительной информации.

Отметим некоторые простейшие свойства колмогоровской сложности. Первое из них – колмогоровская сложность строки не превосходит с точностью до константы ее длины:

$$\mathrm{K}(x|y) \leqslant l(x) + O(1).$$

Это неравенство имеет место, так как можно рассмотреть тривиальный способ декодирования $B(p, y) = p$ для всех p, y. Для этого способа декодирования выполнено $\mathrm{K}_B(x|y) = l(x)$, и поэтому по теореме 3.1

$$\mathrm{K}(x|y) \leqslant \mathrm{K}_B(x|y) + O(1) = l(x) + O(1).$$

[4] В дальнейшем неравенство $f(x_1, \ldots, x_n) \leqslant g(x_1, \ldots, x_n) + O(1)$ означает, что существует константа c такая, что неравенство $f(x_1, \ldots, x_n) \leqslant \leqslant g(x_1, \ldots, x_n) + c$ выполнено для всех $x_1, \ldots, x_n$. Здесь константа c не зависит от $x_1, \ldots, x_n$.

Равенство $f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n) + O(1)$ означает, что выполнены неравенства $f(x_1, \ldots, x_n) \leqslant g(x_1, \ldots, x_n) + O(1)$ и $g(x_1, \ldots, x_n) \leqslant \leqslant f(x_1, \ldots, x_n) + O(1)$.

В частности, сложность натурального числа n ограничена его логарифмом:

$$\mathrm{K}(n) \leqslant \log n + O(1).$$

Некоторые числа имеют значительно меньшую сложность. Например, слово $0^n = 0 \ldots 0$ длины n имеет сложность

$$\mathrm{K}(0^n) \leqslant \log l(0^n) + O(1) = \log n + O(1).$$

Из определения $\mathrm{K}(x|y) \leqslant \mathrm{K}(x) + O(1)$. Разность между правой и левой частями этого неравенства может быть максимально большой. Например, $\mathrm{K}(x|x) = O(1)$. Легко видеть, что существуют конечные последовательности x, для которых имеет место $\mathrm{K}(x) = l(x) + O(1)$.

Нетрудно доказать следующее утверждение.

Предложение 3.3. *Безусловная сложность* $\mathrm{K}(x)$ *слова* x *связана с условной сложностью* $\mathrm{K}(x|y)$ *относительно другого слова* y *и сложностью* $\mathrm{K}(y)$ *этого слова следующим образом:*

$$\mathrm{K}(x|y) - O(1) \leqslant \mathrm{K}(x) \leqslant$$
$$\mathrm{K}(x|y) + \mathrm{K}(y) + 2\min\{\log \mathrm{K}(x), \log \mathrm{K}(y)\} + O(1).$$

Доказательство. Пусть по самому короткому коду p и условию y можно восстановить x. Кроме того, пусть q – самый короткий код для восстановления y. Можно добавить код условия y к коду p и по сложному коду $\overline{\mathrm{str}(l(q))}01qp$ восстановить x. Длина такого кода равна $K(x|y) + \mathrm{K}(y) + 2\log \mathrm{K}(y) + O(1)$. Также можно по сложному коду $\overline{\mathrm{str}(l(p))}01qp$ восстановить x Длина такого кода равна $K(x|y) + \mathrm{K}(y) + 2\log \mathrm{K}(x) + O(1)$. □

Приведем еще некоторые свойства колмогоровской сложности. Пусть $\psi(x)$ – вычислимая функция типа $\{0,1\}^* \to \{0,1\}^*$. Тогда

$$\mathrm{K}(\psi(x)|y) \leqslant \mathrm{K}(x|y) + O(1), \tag{3.6}$$

где константа $O(1)$ зависит от ψ.

Для доказательства этого неравенства рассмотрим способ декодирования $B(p,y) = \psi(A(p,y))$, где $A(p,y)$ – оптимальный способ декодирования. Для этого способа декодирования выполнено $\mathrm{K}_B(\psi(x)|y) = l(p)$, где $A(p,y) = x$. Отсюда получаем (3.6).

Верно также соотношение

$$\mathrm{K}(x|y) \leqslant \mathrm{K}(x|\psi(y)) + O(1). \tag{3.7}$$

Рассмотрим способ декодирования $B(p,y) = A(p,\psi(y))$, где $A(p,y)$ – оптимальный способ декодирования. Для этого способа декодирования $\mathrm{K}_B(x|y) = l(p)$, где $A(p,\psi(y)) = x$. Отсюда получаем (3.7).

Легко видеть, что колмогоровская сложность $\mathrm{K}(x)$ неограничена. В случае ее ограниченности просто не хватило бы кодов для всех слов x (их бесконечно много).

Кроме того, она не является вычислимой. Это свойство следует из более сильного свойства – функция $\mathrm{K}(x)$ не только не вычислимая, но и не имеет неограниченной вычислимой нижней оценки.

Предложение 3.4. *Не существует вычислимой функции $\psi(x)$, которая принимает как угодно большие значения и такой, что $\psi(x) \leqslant \mathrm{K}(x)$ для всех x, для которых $\psi(x)$ определена.*

Доказательство. Допустим, что функция $\psi(x)$ является всюду определенной и $\psi(x) \leqslant \mathrm{K}(x)$ для всех x. Определим другую функцию

$$\mu(x) = \min\{y : \psi(y) > x\}. \tag{3.8}$$

По определению $\psi(\mu(x)) > x$ для всех x.

Так как $\psi(x)$ неограничена, функция $\mu(x)$ является всюду определенной и вычислимой. По свойству (3.6)

$$\mathrm{K}(\mu(x)) \leqslant \mathrm{K}(x) + O(1).$$

С другой стороны, по определению функции ψ

$$x < \psi(\mu(x)) \leqslant \mathrm{K}(\mu(x)) \leqslant \mathrm{K}(x) + O(1) \leqslant l(x) + O(1).$$

Получаем противоречие, так как длина $l(x) = O(\log x)$.

В случае, когда функция $\psi(x)$ не является всюду определенной, для вычисления функции $\mu(x)$ по формуле (3.8) может не существовать алгоритма. Однако совсем не обязательно искать минимум в (3.8), достаточно найти хотя бы какое-нибудь y такое, что $\psi(y) > x$. Определим процесс одновременного вычисления всех значений $\psi(y)$ до тех пор, пока не найдется y такое, что $\psi(y) > x$. Определим значение $\mu(x)$ равным первому такому y. Такое y найдется, так как функция $\psi(y)$ принимает бесконечно много значений. □

Функция $f : \{0,1\}^* \to \{0,1\}^*$ называется перечислимой сверху, если множество $\{(x, y) : y > f(x)\}$ (надграфик функции f) перечислимо.

Легко видеть, что колмогоровская сложность $\mathrm{K}(x|y)$ перечислима сверху, так как

$$\{(x, y, n) : n > \mathrm{K}(x|y)\} = \{(n, x, y) : \exists p(A(p, y) = x \& l(p) < n)\},$$

где $A(p, y)$ – оптимальный способ декодирования.

Утверждение 3.4 связано с одной интерпретацией теоремы Геделя о неполноте, которая была предложена Чейтиным [37].

Теория характеризуется бесконечным набором утверждений, представленных в виде формул. Каждая формула есть слово в некотором алфавите. Можно считать, что формулы закодированы двоичными строками.

Задано множество «истинных» формул TRUTH.

Имеется также некоторый алгоритм, перечисляющий множество PROOF «доказуемых» формул.

Предполагаем также, что теория непротиворечива, т.е. все доказуемые формулы являются истинными: PROOF $\subseteq$ TRUTH.

Допустим, что все истинные в обычном математическом смысле утверждения вида $\mathrm{K}(x) > n$, где $x \in \{0,1\}^*$ и $n \in \mathcal{N}$, могут быть записаны на языке нашей теории и принадлежит множеству TRUTH.

Предложение 3.5. *Не более чем конечное число утверждений*

типа $\mathrm{K}(x) > n$ *с различными* n *могут принадлежать множеству* PROOF*, т.е. могут быть доказуемыми.*

Доказательство. Допустим противное. Тогда существует алгоритм, который перечисляет (доказывает) бесконечную последовательность утверждений $\mathrm{K}(x_i) > n_i$, $i = 1,\ 2,\ \ldots$, среди которых имеется бесконечно много различных n_i.

Допускаем также, что все x_i различные (для этого не перечисляем утверждения с повторяющимися x_i).

Определим частичную вычислимую функцию $\psi(x_i) = n_i$ для всех i. Эта функция принимает как угодно большие значения.

Тогда $\mathrm{K}(x) > \psi(x)$ для всех x, для которых $\psi(x)$ определена. Получаем противоречие с утверждением 3.4.

Так как сложность неограничена, имеется бесконечно много истинных утверждений вида $\mathrm{K}(x) > n$ с различными n. Не более чем конечное число таких утверждений может быть доказано, поэтому PROOF $\neq$ TRUTH. $\square$

3.3. Несжимаемые последовательности

Имеется всего 2^n двоичных строк длины n. Для любого $k < n$ число всех двоичных строк x длины n, для которых выполнено неравенство $\mathrm{K}(x) < n - k$, не превосходит числа всех двоичных кодов p, для которых $l(p) < n - k$. Число таких p равно

$$2^0 + 2^1 + \cdots + 2^{n-k-1} < 2^{n-k}.$$

Таким образом, доля x таких, что $\mathrm{K}(x) < n - k$, оценивается сверху:

$$\frac{|\{x : l(x) = n \& \mathrm{K}(x) < n - k\}|}{2^n} < 2^{-k}.$$

Величину

$$d(x) = n - \mathrm{K}(x)$$

называем дефектом случайности последовательности x. Она обладает свойством

$$|\{x : l(x) = n \& d(x) \leqslant k\}| \geqslant 2^{n-k}.$$

Заметим, что все эти рассуждения останутся верными, если мы заменим $\mathrm{K}(x)$ на $\mathrm{K}(x|l(x))$. Сформулируем это свойство в виде утверждения. Пусть $|D|$ обозначает число элементов конечного множества D.

Предложение 3.6. *Пусть*

$$B_{n,k} = \{x : l(x) = n \& \mathrm{K}(x|l(x)) \geqslant n - k\}$$

– множество всех сжимаемых не более чем на k-битов последовательностей длины n. Тогда

$$2^n(1 - 2^{-k}) \leqslant |B_{n,k}| \leqslant 2^n.$$

Можно также рассматривать в качестве дефекта случайности величину $d(x|n) = n - \mathrm{K}(x|n)$. Для нее верны все те же свойства, что и для $d(x)$.

Таким образом, для всех последовательностей длины n, кроме малой их доли, дефект случайности ограничен. Иными словами, большинство последовательностей несжимаемые. Мы покажем, что вследствие этой несжимаемости для большинства последовательностей имеет место свойство устойчивости частот единиц и нулей.

Перейдем теперь к более точным оценкам колмогоровской сложности. Допустим, что некоторое подмножество

$$D = \{x_1, x_2, \ldots x_m\}$$

множества строк длины n задано в виде списка $(x_1, x_2, \ldots x_m)$ своих элементов. Ранее было указано, как кодировать этот список в виде одной строки. Для простоты обозначаем эту последовательность так же, как само множество D. Тогда для задания элемента $x \in D$ при известном списке D достаточно задать номер этого элемента в списке. Отсюда получаем оценку

$$\mathrm{K}(x|D) \leqslant \log |D| + O(1). \tag{3.9}$$

Кроме этого, имеет место оценка

$$(1-2^{-k})|D| \leqslant |\{x \in D : \mathrm{K}(x|D) \geqslant \log|D| - k\}| \leqslant |D|. \quad (3.10)$$

Можно определить дефект случайности элемента $x \in D$ в виде

$$d(x|D) = \log|D| - \mathrm{K}(x|D).$$

Для него верно неравенство

$$\frac{|\{x \in D : d(x|D) \geqslant k\}|}{|D|} \leqslant 2^{-k}.$$

Более точная верхняя оценка колмогоровской сложности строки длины n может быть получена, если предварительно представить множество всех двоичных строк длины n в виде объединения попарно непересекающихся подмножеств строк с заданным числом единиц:

$$\{0,1\}^n = \cup_{k=0}^{n} \Xi_n^k,$$

где

$$\Xi_n^k = \left\{ x : l(x) = n \& \sum_{i=1}^{n} x_i = k \right\},$$

а затем применить оценку (3.10) и формулу

$$\mathrm{K}(x) \leqslant \mathrm{K}(x|D) + \mathrm{K}(D) + 2\log\mathrm{K}(D) + O(1).$$

Если $k = 0$ или $k = n$, то $\mathrm{K}(x) = O(1)$. Пусть далее $0 < k < n$.

Напомним, что $|\Xi_n^k| = \binom{n}{k} = \frac{n!}{k!(n-k)!}$. Применяя эти оценки, получим для строки $x \in \Xi_n^k$

$$\mathrm{K}(x) \leqslant \mathrm{K}(x|\Xi_n^k) + \mathrm{K}(\Xi_n^k) + 2\log\mathrm{K}(\Xi_n^k) + O(1). \quad (3.11)$$

Далее, применяя формулу Стирлинга

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1 + O(1/n))$$

для каждого факториала из биномиального коэффициента, получим

$$\mathrm{K}(x|\Xi_n^k) = \mathrm{K}(x|n,k) + O(1) \leqslant \log \frac{n!}{k!(n-k)!} + O(1) = \\ = n\left(-\frac{k}{n}\log\frac{k}{n} - \left(1-\frac{k}{n}\right)\log\left(1-\frac{k}{n}\right)\right) - \\ \frac{1}{2}\log\left(\frac{k}{n}\big(1-\frac{k}{n}\big)\right) - \frac{1}{2}\log n + O(1). \quad (3.12)$$

Для задания строки, представляющей множество Ξ_n^k, достаточно знать числа n и k. Поэтому

$$\mathrm{K}(\Xi_n^k) \leqslant \log n + \log k + 2\log\log k + O(1). \quad (3.13)$$

Напомним определение энтропии Шеннона:

$$H(p) = -p\log p - (1-p)\log(1-p),$$

где $0 \leqslant p \leqslant 1$ (полагаем $0\log 0 = 0$).

Учитывая это представление и оценки (3.12) и (3.13), перепишем оценку (3.11) в упрощенном виде:

$$\mathrm{K}(x) \leqslant nH\left(\frac{k}{n}\right) + \frac{1}{2}\log n + \log k + 2\log\log k + O(1) \leqslant \\ nH\left(\frac{k}{n}\right) + O(\log n). \quad (3.14)$$

Верхняя оценка (3.14) позволяет вывести некоторые статистические закономерности для несжимаемых последовательностей.

Допустим, что длина x равна n. Рассмотрим разложение энтропии $H(p)$ по формуле Тэйлора в окрестности точки $p = \frac{1}{2}$. Легко проверить, что $H(1/2) = 1$, $H'(\frac{1}{2}) = 0$ и $H''(1/2) < 0$, где

$$H''(p) = -\frac{1}{p(1-p)\ln 2}.$$

Имеем

$$H(p) = 1 - 2\log e\left(p - \frac{1}{2}\right)^2 + o\left(\left(p - \frac{1}{2}\right)^2\right). \qquad (3.15)$$

Напомним, что величина

$$d(x) = n - \mathrm{K}(x)$$

называется дефектом случайности последовательности x.

Из неравенств (3.14), (3.15) и равенства $\mathrm{K}(x) = n - d(x)$ получим

$$\left(\frac{k}{n} - \frac{1}{2}\right)^2 = O\left(\frac{\log n + d(x)}{n}\right).$$

Таким образом, мы доказали следующую теорему – закон больших чисел для несжимаемых последовательностей.

Теорема 3.2. *Существует константа C такая, что для любой конечной двоичной последовательности x длины n, содержащей k единиц, выполнено*

$$\left|\frac{k}{n} - \frac{1}{2}\right| \leqslant C\sqrt{\frac{\log n + d(x)}{n}}.$$

В статье [11] А. Н. Колмогоров предложил считать признаком случайности конечной последовательности x отсутствие в ней закономерностей, что выражается в невозможности более конечного описания x, чем ее длина. Мы называем такие последовательности несжимаемыми. Для несжимаемой последовательности x величина $d(x) = n - \mathrm{K}(x)$ мала. Теорема 3.2 показывает, что для несжимаемых последовательностей частота единиц близка к $\frac{1}{2}$. Оценка отклонения частоты от $\frac{1}{2}$ зависит от степени несжимаемости последовательности x.

А. Н. Колмогоров придавал большое значение изучению понятия алгоритмической случайности *конечного объекта*. При этом понятие меры не должно входить в это определение. Приведем интерпретацию идеи Колмогорова в следующем виде. Вместо меры граничные условия на случайность задаются в виде

разбиения множества всех конечных последовательностей длины n на конечные попарно непересекающиеся подмножества. Конечная последовательность x является случайной, если выполнено $K(x|D) \approx \log |D|$, где D – тот элемент разбиения, которому принадлежит x, $x \in D$, $|D|$ – число его элементов. Из теории кодирования следует, что $K(x|D) \leqslant \log |D| + c$ для любого $x \in D$, где c – константа. Для характеризации «степени случайности» конечной последовательности Колмогоров вводит дефект случайности конечной последовательности x относительно конечного множества D:

$$d(x|D) = \log |D| - K(x|D), \tag{3.16}$$

где $K(x|D)$ – условная сложность конечной последовательности x относительно множества D, заданного списком его элементов.

Основная идея Колмогорова заключалась в том, чтобы находить подходящие разбиения множества всех конечных последовательностей и выводить стохастические свойства конечной последовательности из предположения о том, что ее сложность относительно соответствующего элемента разбиения D (для которого $x \in D$) близка к своему максимальному значению, а именно, $\mathrm{K}(x|D) \approx \log |D|$.

Мы продолжим обсуждение стохастических свойств несжимаемых последовательностей в разделе 9

3.4. Сложность пары

В этом разделе докажем теорему Колмогорова–Левина о декомпозиции сложности пары. Впервые это доказательство было опубликовано в обзоре [10].

Теорема 3.3.

$$\mathrm{K}(x, y) = \mathrm{K}(x) + \mathrm{K}(y|x) + O(\log \mathrm{K}(x, y)). \tag{3.17}$$

Доказательство. Пусть по самому короткому коду p и условию x можно восстановить y. Кроме того, пусть q – самый короткий код для восстановления x. Можно добавить код условия

x к коду p и по сложному коду $\overline{\mathrm{str}(l(q))}01qp$ восстановить пару (x,y). Длина такого кода равна правой части неравенства (3.17). Неравенство $\leqslant$ доказано.

Доказательство обратного неравенства $\geqslant$ намного сложнее. Фиксируем пару (x,y). Обозначим $a=\mathrm{K}(x,y)$. Множество

$$A=\{(x',y'):\mathrm{K}(x',y')\leqslant a\}$$

является перечислимым. Кроме этого, $|A|<2^{a+1}$. Для каждой строки x' рассмотрим сечение

$$A_{x'}=\{y':(x',y')\in A\}.$$

Обозначим $m=\lfloor\log|A_x|\rfloor$. Тогда $2^m\leqslant|A_x|<2^{m+1}$.

1) Оценим сверху величину $\mathrm{K}(y|x)$. Для этого по числу a перечисляем все пары $(x',y')\in A$ и откладываем те пары, для которых $x'=x$. Таким образом, мы перечисляем множество A_x. Для задания $y\in A_x$ при известном x можно использовать число a и порядковый номер перечисления $y\in A_x$, т.е. число, не превосходящее $|A_x|<2^{m+1}$. Таким образом, учитывая неравенство $m\leqslant a$, получаем

$$\begin{gathered}\mathrm{K}(y|x)\leqslant\log|A_x|+2l(\mathrm{str}(a))+O(1)\leqslant\\ \leqslant m+2\log\mathrm{K}(x,y)+O(1)\leqslant m+3\log\mathrm{K}(x,y)+O(1).\end{gathered}\quad(3.18)$$

2) Оценим сверху величину $\mathrm{K}(x)$. Пусть

$$B=\{x':|A_{x'}|\geqslant 2^m\}.$$

Из неравенства

$$2^m|B|\leqslant\sum_{x'\in B}|A_{x'}|\leqslant|A|<2^{a+1}$$

следует $|B|<2^{a-m+1}$.

Элементы множества B можно перечислять зная a и m.

По определению $x\in B$. Для задания x по a перечисляем пары из множества A. При этом, как только наберется не менее

2^m пар (x', y') с одинаковой первой координатой x', перечисляем x' в B.

Так как $|A_x| \geqslant 2^m$, среди всех x', перечисленных в B, будет и x. Для задания x надо, кроме выше перечисленной информации, знать порядковый номер его перечисления, т.е. число, не превосходящее $|B| < 2^{a-m+1}$. Двоичный код этого числа имеет длину не более $a - m + 1$. Мы использовали также $a = \mathrm{K}(x, y)$ и $m \leqslant a + 1$. Их двоичные коды имеют длину не более $3 \log \mathrm{K}(x, y)$. Отсюда

$$\mathrm{K}(x) \leqslant a - m + O(\log \mathrm{K}(x, y)). \tag{3.19}$$

Сложим неравенства (3.18) и (3.19) и получим необходимое неравенство

$$\mathrm{K}(y|x) + \mathrm{K}(x) \leqslant \mathrm{K}(x, y) + O(\log \mathrm{K}(x, y)).$$

Теорема доказана. □

Задача 10 из раздела (3.6) утверждает, что оценка (3.17) не улучшаемая.

3.5. Количество информации

На основе понятия сложности А. Н. Колмогоров предложил в [11] определение количества информации в слове y о слове x:

$$\mathrm{I}(y : x) = \mathrm{K}(x) - \mathrm{K}(x|y).$$

Величину $\mathrm{K}(x)$ можно интерпретировать как минимальное количество информации, необходимое для воспроизведения слова x; $\mathrm{K}(x|y)$ интерпретируется как минимальное количество информации, которое необходимо добавить к информации, содержащейся в y, чтобы восстановить x. Разность между ними естественно интерпретировать как количество информации, содержащейся в слове y о слове x.

Это определение аналогично вероятностному определению информации, содержащейся в случайной величине ψ о случайной величине ξ:

$$IH(\psi : \xi) = H(\xi) - H(\xi|\psi),$$

где H – энтропия Шеннона.

В отличие от вероятностного определения количества информации $IH(\psi : \xi)$, величина $\mathrm{I}(y : x)$ не коммутативна. Покажем это на примере.

Для любого m найдется x длины m такое, что $\mathrm{K}(x|m) \geqslant m$. Действительно, если бы такого x не существовало, для любого y длины m существует p длины $\leqslant m - 1$ такое, что $A(p, m) = y$. Здесь A – оптимальный метод декодирования. Все такие p различные. Число таких p, а значит, число и таких y, не превосходит числа всех двоичных строк длины $< m$. Это число равно $2^m - 1$. Так как число всех двоичных строк длины m равно 2^m, найдется x такое, что $\mathrm{K}(x|m) \geqslant m$.

Аналогичным образом найдутся сколь угодно большие m, для которых $\mathrm{K}(m) \geqslant l(m)$.

Очевидно, $\mathrm{K}(l(m)|m) = O(1)$. Для каждого такого m и для соответствующего x длины m получаем

$$\begin{aligned} \mathrm{I}(x : m) = \mathrm{K}(m) - \mathrm{K}(m|x) &\geqslant l(m) - O(1), \\ \mathrm{I}(m : x) = \mathrm{K}(x) - \mathrm{K}(x|m) &\leqslant \\ \leqslant l(x) - m + O(1) &= O(1). \end{aligned}$$

Значит, $\mathrm{I}(x : m) - \mathrm{I}(m : x) \geqslant l(m) - O(1)$. Длина двоичной записи числа m удовлетворяет неравенству $l(m) \geqslant \log m - 1$.

Дальнейшее расхождение между величинами $\mathrm{I}(x : y)$ и $\mathrm{I}(y : x)$ в общем случае нельзя увеличить. Величина $\mathrm{I}(x : y)$ коммутативна с точностью до $O(\log \mathrm{K}(x, y))$. По теореме 3.3

$$\begin{aligned} \mathrm{K}(x, y) &= \mathrm{K}(x) + \mathrm{K}(y|x) + O(\log \mathrm{K}(x, y)), \\ \mathrm{K}(x, y) &= \mathrm{K}(y) + \mathrm{K}(x|y) + O(\log \mathrm{K}(x, y)). \end{aligned}$$

Отсюда получаем

$$|\mathrm{I}(y:x) - \mathrm{I}(x:y)| = O(\log \mathrm{K}(x,y)),$$
$$|\mathrm{I}(y:x) - (\mathrm{K}(x) + \mathrm{K}(y) - \mathrm{K}(x,y))| = O(\log \mathrm{K}(x,y)).$$

3.6. Задачи и упражнения

1. Доказать, что универсальная функция $U(x,y)$ не является всюду определенной.

2. Объясните, почему функция сложности $\mathrm{K}(x)$ определена для любого x, а также, почему она неограничена.

3. Доказать что

(a) $\mathrm{K}(0^n|n) = 0(1)$, $\mathrm{K}(0^n) \leqslant \log n + 0(1)$, $\mathrm{K}(0^n) = \mathrm{K}(n) + 0(1)$, где 0^n – слово, состоящее из n нулей; $\mathrm{K}(2^n) \leqslant \log n + 0(1)$ и $\mathrm{K}(2^{2^n}) \leqslant \log n + 0(1)$, где 2^n – натуральное число – степень двойки, понимаемое в обычном смысле.

(b) Существует константа $c \geqslant 0$ такая, что $\mathrm{K}(0^n) \geqslant \log n - c$ для бесконечно многих n.

(c) $\mathrm{K}(x|l(x)) \leqslant \mathrm{K}(x) + 0(1) \leqslant \mathrm{K}(x|l(x)) + \log l(x) + \log\log l(x) + 2\log\log\log l(x) + 0(1)$.

(d) $\mathrm{K}(x,x) = \mathrm{K}(x) + 0(1)$; $\mathrm{K}(x,\mathrm{K}(x)) = \mathrm{K}(x) + 0(1)$.

(e) $\mathrm{K}(x0) = \mathrm{K}(x1) + 0(1) = \mathrm{K}(0x) + 0(1) = \mathrm{K}(1x) + 0(1) = \mathrm{K}(x) + 0(1)$.

(f) $\mathrm{K}(x|y0) = \mathrm{K}(x|y1) + 0(1) = \mathrm{K}(x|0y) + 0(1) = \mathrm{K}(x|1y) + 0(1) = \mathrm{K}(x|y) + 0(1)$.

4. Доказать, что оптимальный способ описания $A(p)$ не является всюду определенной функцией и для него не существует соответствующего алгоритма кодирования, который по произвольной конечной последовательности x выдавал бы какой-нибудь самый короткий код p, для которого $A(p) = x$.

5. Доказать, что функция сложности $\mathrm{K}(x)$ не является перечислимой снизу, но является перечислимой сверху.

6. Доказать, что существует константа c такая, что для любого N найдется пара последовательностей (x, y), для которой выполнено $l(x) + l(y) = N$ и $\mathrm{K}(x, y) \geqslant N + \log N - c$.

7. Доказать, что для любого n существует последовательность x длины $\leqslant n$ такая, что замена некоторого бита в ней на противоположный приводит к последовательности x', где

$$\mathrm{K}(x') \geqslant \mathrm{K}(x) + \log n - O(1).$$

Как надо исправить это неравенство, если потребовать существование такой последовательности x длины равной n.

При любой такой замене

$$\mathrm{K}(x') \leqslant \mathrm{K}(x) + \log n + O(\log \log n).$$

8. Пусть $K(x) \geqslant n - c$, $c > 0$, и $x = yz$, где $l(y) = l(z) = n/2$. Тогда $\mathrm{K}(y) \geqslant n/2 - O(\log n)$ и $\mathrm{K}(z) \geqslant n/2 - O(\log n)$.

9. Провести доказательство неравенства (3.10).

10. Доказать, что неравенства $\mathrm{K}(x, y) \leqslant \mathrm{K}(x) + \mathrm{K}(y|x) + O(1)$ и $\mathrm{K}(x, y) \leqslant \mathrm{K}(x) + \mathrm{K}(y|x) + O(\log \log \mathrm{K}(x, y))$, а также неравенство $\mathrm{K}(x, y) \leqslant \mathrm{K}(x) + \mathrm{K}(y|x) + \log \mathrm{K}(x, y) + O(1)$ в общем случае неверны.

Привести нетривиальные примеры последовательностей, для которых первое из неравенств выполнено.

11. Пусть A – перечислимое множество и $\omega = \omega_1\omega_2 \ldots$ – его характеристическая последовательность, где

$$\omega_i = \begin{cases} 1, \text{ если } i \in A, \\ 0 \text{ в противном случае.} \end{cases}$$

Обозначаем $\omega^n = \omega_1 \ldots \omega_n$ – последовательность первых n битов бесконечной последовательности ω.

Доказать, что $\mathrm{K}(\omega^n|n) \leqslant \log n + O(1)$, где $\omega^n = \omega_1\omega_2 \ldots \omega_n$.

Оценить сверху $\mathrm{K}(\omega^n)$. Как изменятся эти оценки, если множество A разрешимо?

12. Даны два слова x и y – два слова одной длины n. Оценить $\mathrm{I}(x : y)$ сверху и по возможности привести оценки снизу в наихудшем случае:

(a) $x = 0101\ldots01$ и $y = 1010\ldots10$ длины $2n$;
(b) $x = 0^n$ и $y = 1^{n/2}0^{n/2}$;
(c) $x = 0^{n/2}1^{n/2}$ и $y = 1^{n/2}0^{n/2}$;
(d) $x = x_1\ldots x_n$ и $x' = x_1x_1\ldots x_nx_n$;
(e) $x = uvw$ и $y = usw$, слова u, v, w, s – длины n;
(f) $x = uvw$ и $y = svt$, слова u, v, w, s, t – длины n;

13. Доказать неравенства:
(a) $\mathrm{K}(x) \leqslant \mathrm{K}(xy) + 2\log l(x) + O(1)$;
(b) $\mathrm{K}(x) \leqslant \mathrm{K}(xy) + 2\log l(y) + O(1)$;
(c) привести примеры конечных последовательностей x и y, для которых неравенство $\mathrm{K}(x) \leqslant \mathrm{K}(xy) + O(1)$ и даже $\mathrm{K}(x) \leqslant \mathrm{K}(xy) + 2\log \mathrm{K}(x) + O(1)$ неверно. Привести примеры последовательностей x длины n, у которых существуют подпоследовательности на порядок более сложные, чем вся последовательность.

14. Доказать, что для почти любой бесконечной последовательности ω существует такое число m, что $\mathrm{K}(\omega^n) \geqslant n - m$ для бесконечно многих n.

15. Доказать, что среди натуральных чисел от 1 до n найдется число сложности $\geqslant \log n - 1$. Оценить долю чисел от 1 до n сложность которых $\geqslant \log n - c$, где $c \geqslant 1$.

16. Доказать, что для любого y число всех x длины n таких, что $\mathrm{K}(x|y) \leqslant \mathrm{K}(x) - m$ не превосходит 2^{n-m+c}, где константа c не зависит от m и y.

17. Доказать, что для дюбых строк y, z, u длины n найдется строка x длины n такая, что $\mathrm{K}(x|y) \geqslant n - 2$, $\mathrm{K}(x|z) \geqslant n - 2$ и $\mathrm{K}(x|u) \geqslant n - 2$.

18. Доказать, что для любой бесконечной последовательности ω будет $\sup_n \mathrm{K}(\omega^n|n) < \infty$ тогда и только тогда, когда ω является вычислимой.

19. Существует такая константа c, что для любой бесконечной последовательности ω выполнено $\mathrm{K}(\omega^n) \leqslant n - \log n + c$ для бесконечно многих n.

20. Существуют бесконечная ω и константа c такие, что

$$\mathrm{K}(\omega^n) \geqslant n - 2\log n - c$$

для всех n.

21. Доказать, что существует такая константа c, что для любых x, n и k, если имеется $\geqslant 2^k$ таких p, что $A(p) = x$ и $l(p) \leqslant n$, то $\mathrm{K}(x|k) \leqslant n-k+c$ (здесь $A(p)$ – оптимальный способ описания).

22. Доказать, что для любого безусловного способа описания $A(p)$ существует такая константа c, что для любого x число его кратчайших описаний не превосходит c. Указание: использовать предыдущую задачу.

Глава 4

Случайность по Мартин-Лефу

В этой главе мы рассмотрим конструктивные варианты классических понятий из топологии и теории меры. В частности, будет определено понятие бесконечной случайной по Мартин-Лефу последовательности. Мы покажем, что известные асимптотические законы теории вероятностей выполнены для каждой такой случайной последовательности.

Будет сформулирована новая логика теории вероятностей. Согласно этой логике законы теории вероятностей выполнены не только для почти всюду, – как это имеет место в классической теории вероятностей, – но и для каждой индивидуальной последовательности, которая выдерживает универсальный тест Мартин-Лефа.

4.1. Тесты Мартин-Лефа

Пусть Ω – множество всех бесконечных (двоичных или бинарных) последовательностей, состоящих из 0 и 1. Топология на этом множестве задается интервалами вида

$$\Gamma_x = \{\omega \in \Omega : x \subset \omega\},$$

где x – конечная двоичная последовательность.

Множество Ω можно изображать в виде бесконечного двоичного дерева, вершиной которого является пустая последовательность λ. Остальные его вершины представлены всеми конечными двоичными последовательностями. Порядок между этими вершинами задается отношением продолжения последовательностей. Каждая бесконечная последовательность, состоящая из 0 и 1, изображается бесконечным путем на дереве, стартующем из корня.

Интервал Γ_x состоит из всех бесконечных продолжений конечной последовательности x. Любые два интервала Γ_x и Γ_y либо не пересекаются: $\Gamma_x \cap \Gamma_y = \emptyset$, если последовательности x и y не продолжают друг друга, либо один из них является подмножеством другого: $\Gamma_x \subseteq \Gamma_y$ или $\Gamma_y \subseteq \Gamma_x$ в противоположном случае.

Открытые множества представляют собой объединения таких интервалов. Каждое открытое множество можно представить в виде объединения попарно непересекающихся интервалов. Замкнутые множества – это дополнения открытых множеств.

Сначала мы будем рассматривать равномерную меру L на множестве Ω. Она задается своими значениями на интервалах

$$L(\Gamma_x) = 2^{-l(x)}$$

для всех $x \in \{0,1\}^*$. Далее эта мера может быть продолжена естественным образом на все открытые и замкнутые множества, а затем и на все борелевские подмножества Ω.

Мы рассмотрим конструктивные аналоги этих понятий. Конструктивизация означает, что все функции и операции должны быть в каком-нибудь смысле вычислимыми.

Интервал Γ_x однозначно задается конечной последовательностью x и поэтому является конструктивным объектом. Равномерная мера интервалов по определению – вычислимая функция, переводящая конечные последовательности x в рациональные числа $2^{-l(x)}$.

Назовем открытое множество U *эффективно открытым*, если его можно представить в виде объединения вычислимой по-

следовательности интервалов

$$U = \cup_{i=1}^{\infty} \Gamma_{x_i},$$

где функция $f(i) = x_i$ является вычислимой. Дополнение эффективно открытого множества называется *эффективно замкнутым* множеством.

В теории вероятностей особое значение имеют множества меры 0. В частности, асимптотические законы теории вероятностей, такие как усиленнй закон больших чисел или закон повторного логарифма, имеют место для всех последовательностей, кроме множества меры 0. Говорят, что они имеют место почти всюду.

Измеримое подмножество $A \subset \Omega$ имеет меру 0, если для любого $\epsilon > 0$ существует такое открытое множество $U = \cup_i \Gamma_{x_i}$, что $A \subseteq U$ и $L(U) < \epsilon$.

Определим конструктивный аналог множества меры 0. Измеримое множество $A \subset \Omega$ является эффективно нулевым, если такая последовательность интервалов задается по рациональному числу ϵ некоторой вычислимой функцией.

Более точно, множество $A \subset \Omega$ является *эффективно нулевым*, если существует такая вычислимая функция $x(i, \epsilon)$, где i – натуральное, а ϵ – положительное рациональное число, что

- $L(\cup_i \Gamma_{x(i,\epsilon)}) < \epsilon$ и
- $A \subseteq \cup_i \Gamma_{x(i,\epsilon)}$ для всех рациональных $\epsilon > 0$.

Основываясь на этих соображениях, перейдем к определению теста Мартин-Лефа. Рассмотрим убывающую последовательность рациональных чисел $\epsilon_m = 2^{-m}$, $m = 1, 2, \ldots$ Множество

$$T = \{(m, x(i, 2^{-m})) : i, m = 1, 2, \ldots\}$$

является перечислимым. Этому множеству соответствует последовательность эффективно открытых множеств $\{U_m\}$ такая, что

- $U_m = \cup\{\Gamma_x : (m,x) \in T\}$,[1]
- $L(U_m) \leqslant 2^{-m}$ для всех m,
- $A \subseteq \cap_m U_m$.

Можно взять эти свойства в качестве определения эффективно нулевого множества. Множество пар T называется вычислимой основой для системы эффективно открытых множеств $\{U_m\}$.

Множество T определяет равномерный способ перечисления интервалов, составляющих семейство $\{U_m\}$. Назовем такое семейство эффективно открытых множеств *равномерно перечислимым*.

Можно потребовать, чтобы выполнялось еще одно свойство системы $\{U_m\}$:

- $U_{m+1} \subseteq U_m$ для всех m.

Семейство множеств $\{U_m\}$, удовлетворяющее первым трем свойствам, легко перестроить в другую последовательность $\{U'_m\}$, удовлетворяющую четвертому свойству, т.е. такую, что

$$U'_m = \cup_{n>m} U_n.$$

Тогда $L(U'_m) \leqslant \sum_{n>m} 2^{-n} \leqslant 2^{-m}$.

Система эффективно открытых множеств $\{U_m\}$, удовлетворяющая первым трем условиям, называется *тестом проверки на случайность* по Мартин-Лефу. Каждый тест Мартин-Лефа определяет эффективно нулевое множество $\cap_m U_m$, которое также будет называться *нулевым множеством теста*.

Бесконечная двоичная последовательность *отвергается* таким тестом, если она лежит в его нулевом множестве. Мы говорим также, что такая последовательность не случайная по Мартин-Лефу. Последовательность *выдерживает* тест Мартин-Лефа, если она не принадлежит его нулевому множеству.

[1] Говорим также, что множества U_m *равномерно эффективно открытые*.

Мы будем называть бесконечную двоичную последовательность *случайной по Мартин-Лефу*, если она не принадлежит никакому эффективно нулевому множеству. Другими словами, случайная последовательность выдерживает любой тест Мартин-Лефа.

По существу, эффективно нулевые множества – это все подмножества нулевых множеств тестов Мартин-Лефа $\{U_m\}$. Поскольку множество всех тестов Мартин-Лефа счетно, мера объединения их нулевых множеств равна нулю. Таким образом, мера множества всех случайных последовательностей равна единице.

Приведем некоторые примеры тестов Мартин-Лефа и соответствующих эффективно нулевых множеств.

Множество, состоящее из одной бесконечной последовательности $0^\infty = 00\ldots$, является эффективно нулевым, так как $0^\infty \in \cap_n \Gamma_{0^n}$ и $L(\Gamma_{0^n}) = 2^{-n}$ для всех n. Кроме того, последовательность интервалов Γ_{0^n} является перечислимой.

Множество U, состоящее из всех двоичных последовательностей вида $\omega = \omega_1 0 \omega_2 0 \ldots$, также является эффективно нулевым, так как содержится в пересечении перечислимой системы эффективно открытых множеств:

$$U_m = \cup\{\Gamma_{x_1 0 x_2 0 \ldots x_m 0} : x_i \in \{0,1\}, i = 1, \ldots m\},$$

где $L(U_m) = 2^{-m}$ для всех m.

Приведем более сложный пример – эффективно нулевое множество, связанное с усиленным законом больших чисел, который сформулируем следующим образом. Обозначим

$$\mathcal{L} = \left\{ \omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \frac{1}{2} \right\},$$

где $S_n(\omega) = \sum_{i=1}^{n} \omega_i$. Усиленный закон больших чисел утверждает, что

$$L(\mathcal{L}) = 1.$$

Определим тест Мартин-Лефа, который отвергает любую бесконечную двоичную последовательность ω, для которой усиленный закон больших чисел нарушается, т.е. $\omega \notin \mathcal{L}$.

Построим равномерно перечислимое семейство эффективно открытых множеств U_m, $m = 1, 2 \ldots$, такое, что $\Omega \setminus \mathcal{L} \subseteq \cap_m U_m$.

Для построения этой системы множеств мы будем использовать неравенство Хефдинга

$$L\left\{\omega : \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| > \delta\right\} < 2e^{-2n\delta^2}$$

для всех n, δ.

Далее считаем, что δ – положительное рациональное число. Определим семейство эффективно открытых множеств

$$U_n^\delta = \left\{\omega : \sup_{k \geqslant n} \left|\frac{S_k(\omega)}{k} - \frac{1}{2}\right| > \delta\right\}.$$

Нетрудно проверить, что

$$U_n^\delta = \bigcup_{k \geqslant n} \bigcup_x \left\{\Gamma_x : l(x) = k \& \left|\frac{S_k(x)}{k} - \frac{1}{2}\right| > \delta\right\}$$

– эффективно открытое множество, а его мера также убывает экспоненциально по n:

$$L(U_n^\delta) < \frac{1}{\delta^2} e^{-2n\delta^2} = e^{-2n\delta^2 - 2\ln\delta}.$$

Пусть $\mathcal{U} = \Omega \setminus \mathcal{L}$ – множество всех бесконечных последовательностей ω, для которых усиленный закон больших чисел нарушается.

По определению $\omega \in \mathcal{U}$ тогда и только тогда, когда существует δ такое, что $\omega \in \cap_n U_n^\delta$. Иными словами,

$$\mathcal{U} \subseteq \cup_\delta \cap_n U_n^\delta.$$

Заметим, что $L(\cap_n U_n^\delta) = 0$ для любого $\delta > 0$.

Нам необходимо построить равномерно перечислимую последовательность эффективно открытых множеств, пересечение которых включает $\mathcal{U}$.

Рассмотрим счетную вычислимую последовательность убывающих рациональных чисел $\delta_i = 2^{-i}$. Для каждых i и m эффективно находим $n_{i,m}$ такое, что

$$L(U^{\delta_i}_{n_{i,m}}) < e^{-2n_{i,m}\delta_i^2 - 2\ln\delta_i} < 2^{-m-i}.$$

Для каждого m выберем из каждого семейства $\{U^{\delta_i}_n\}$ множество $U^{\delta_i}_{n_{i,m}}$ и возьмем объединение всех таких множеств:

$$U_m = \cup_i U^{\delta_i}_{n_{i,m}}.$$

Тогда

$$L(U_m) \leqslant \sum_{i=1}^{\infty} 2^{-m-i} = 2^{-m}$$

для всех m. Кроме этого,

$$\mathcal{U} \subseteq \cap_m U_m.$$

Таким образом, мы определили равномерно перечислимую последовательность эффективно открытых множеств $\{U_m\}$, пересечение которых содержит все бесконечные последовательности, на которых нарушается усиленный закон больших чисел. В частности, истинна импликация:

$$\omega \in \mathcal{L} \Rightarrow \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \frac{1}{2}.$$

4.2. Универсальный тест Мартин-Лефа

В этом разделе мы докажем основной результат конструктивного подхода к теории вероятностей – теорему о существовании максимального по включению эффективно нулевого множества.

Теорема 4.1. *Существует максимальное по включению эффективно нулевое множество.*

Доказательство. Имеется счетное число равномерно перечислимых семейств эффективно открытых множеств

$$\{U_n^i\},\ n = 1, 2, \ldots,$$

нулевые множества $\cap_n U_n^i$ которых задают все эффективно нулевые множества. Нам необходимо доказать, что объединение всех эффективно нулевых множеств само содержится в эффективно нулевом множестве из этого же семейства. Идея такого построения аналогична способу, который был использован при анализе усиленного закона больших чисел. Для каждого m мы отберем из каждого равномерно перечислимого семейства эффективно открытых множеств $\{U_n^i\}$ одно множество $U_{n_i}^i$ такое, что $L(U_{n_i}^i) < 2^{-m-i}$, и определим семейство эффективно открытых множеств $U_m = \cup_i U_{n_i}^i$. Тогда $L(U_m) \leqslant 2^{-m}$ для всех m и пересечение этих множеств $\cap_m U_m$ будет содержать все эффективно нулевые множества.

Проблема заключается в том, что все указанные операции должны быть эффективными. Для этого, аналогично тому как это делалось в частном примере – при анализе усиленного закона больших чисел, мы должны определить равномерно перечислимую последовательность $\{U_m^i, m = 1, 2, \ldots\}$, $i = 1, 2, \ldots$, семейств эффективно открытых множеств, среди которых содержатся все перечислимые семейства.

Здесь мы также рассмотрим вычислимую универсальную основу – перечислимое множество троек $\mathcal{T} = \{(i, m, x)\}$, где i, m – натуральные числа, x – конечная двоичная последовательность.

Лемма 4.1. *Существует универсальная вычислимая основа $\mathcal{T}$ такая, что*

- *для любого i система эффективно открытых множеств*

$$U_m = \cup_x \{\Gamma_x : (i, m, x) \in \mathcal{T}\} \tag{4.1}$$

является тестом Мартин-Лефа;

- *для любого теста Мартин-Лефа* $\{U_m, m = 1, 2, \dots\}$ *найдется* i *такое, что выполнено равенство* (4.1).

Доказательство. Для построения вычислимой основы $\mathcal{T}$, обладающей необходимыми свойствами, мы используем теорему о существовании универсальной функции. Пусть $U(i, m, x)$ – функция универсальная для всех вычислимых функций $\phi(m, x)$ от двух аргументов. Здесь удобно считать, что i и m – натуральные числа, а x – двоичная строка.

Как было замечено в разделе 3.1.2, универсальная функция $U(i, m, x)$ определяет универсальную систему перечислимых множеств:

$$W_i = \{(m, x) : U(i, m, x) \text{ определена}\}.$$

Эта система удовлетворяет условиям:

- W_i – перечислимое множество пар для любого i;
- для любого перечислимого множества пар W найдется такое i, что $W = W_i$;
- множество $\{(i, m, x) : (m, x) \in W_i\}$ перечислимо.

Мы перестроим семейство множеств W_i в перечислимое семейство T_i так, что

1) $T_i \subseteq W_i$ для всех i;
2) T_i – вычислимая основа некоторого теста Мартин-Лефа для любого i;
3) для любой вычислимой основы некоторого теста Мартин-Лефа T найдется такое i, что $T = T_i = W_i$;
4) множество $\mathcal{T} = \{(i, m, x) : (m, x) \in T_i\}$ перечислимо.

Произведем перестройку системы W_i следующим образом. Развернем процесс перечисления множеств W_i: на каждом шаге s этого процесса делается один шаг вычисления значения

$U(i,m,x)$ для одного из наборов (i,m,x). Для этого мы каким-либо образом просматриваем каждый набор (i,m,x) на бесконечном числе шагов процесса.

FOR $s = 1, 2, \ldots$

Пусть T_i^{s-1} – все пары (m,x), перечисленные в множество T_i за шаги $< s$. Полагаем $T_i^0 = \emptyset$.

Пусть на шаге s мы просматриваем набор (i,m,x).

Если значение $U(i,m,x)$ ранее не было определено, то выполняем очередной шаг вычисления значения $U(i,m,x)$ для просматриваемого набора (i,m,x).

Если значение $U(i,m,x)$ впервые определено, то проверяем условие

$$L\left(\cup_z\{\Gamma_z : (m,z) \in T_i^{s-1}\} \cup \Gamma_x\right) \leqslant 2^{-m}. \tag{4.2}$$

Если это условие выполнено, то определяем

$$T_i^s = T_i^{s-1} \cup \{(m,x)\},$$

в противном случае определим $T_i^s = T_i^{s-1}$.

Полагаем $T_j^s = T_j^{s-1}$ для всех $j \neq i$.

ENDFOR

Определим $T_i = \cup_s T_i^s$ для каждого i и

$$\mathcal{T} = \{(i,m,x) : (m,x) \in T_i\}.$$

Легко видеть, условия 1) – 2) выполнены, так как они проверялись в процессе конструкции. Условие 4) выполнено по природе самой конструкции. Покажем, что условие 3) также выполнено. Пусть T – вычислимая основа для теста Мартин-Лефа. Как перечислимое множество, $T = W_i$ для некоторого i. Для такого i условие (4.2) всегда будет выполнено и все пары из W_i будут перечислены в T_i. Таким образом, $T = W_i = T_i$.

Лемма доказана. □

Завершим доказательство теоремы 4.1. Определим серию тестов Мартин-Лефа:

$$U_m^i = \cup_x\{\Gamma_x : (i,m,x) \in \mathcal{T}\}.$$

По лемме 4.1 для каждого теста Мартин-Лефа $\{U_m\}$ найдется i такое, что $U_m = U_m^i$ для всех m.

Определим максимальный тест $\{U_m\}$ следующим образом:

$$U_m = \cup_i U_{i+m}^i$$

при $m = 1, 2, \ldots$ Тогда для любого m

$$L(U_m) \leqslant \sum_{i=1}^{\infty} L(U_{i+m}^i) \leqslant \sum_{i=1}^{\infty} 2^{-i-m} = 2^{-m}.$$

По свойству 4) множество

$$T = \cup_i \{(m+i, x) : (i, m, x) \in \mathcal{T}\}$$

перечислимое. Легко видеть, что оно является вычислимой основой теста $\{U_m\}$.

Для любого теста Мартин-Лефа $\{U_m^i\}$ будет выполнено

$$U_{m+i}^i \subseteq U_m$$

для любого m. Поэтому нулевое множество теста $\{U_m^i\}$ является подмножеством нулевого множества теста $\{U_m\}$:

$$\cap_n U_n^i \subseteq \cap_m U_m.$$

Теорема 4.1 доказана. □

Построенный в теореме 4.1 тест называется *универсальным тестом* Мартин-Лефа.

На самом деле мы доказали даже более сильное утверждение про универсальный тест.

Следствие 4.1. *Универсальный тест Мартин-Лефа $\{U_m\}$ обладает следующим свойством: для произвольного теста $\{V_m\}$ найдется число i такое, что выполнено*

$$V_{m+i} \subseteq U_m$$

для всех m.

Из определения следует, что бесконечная последовательность ω является случайной по Мартин-Лефу тогда и только тогда, когда она не содержится в нулевом множестве универсального теста.[2]

Заметим, что в определении теста Мартин-Лефа можно потребовать $L(U_m) \leqslant \rho(m)$ для всех m, где $\rho(m)$ – произвольная вычислимая функция такая, что $\rho(m) \to 0$ при $m \to \infty$. Это определение теста приводит к тому же классу случайных последовательностей (см. задачу 10 из раздела 4.3).

Алгоритмический подход к теории вероятностей предлагает новую логику для интерпретации вероятностных законов.

Пусть $\Phi(\omega)$ – некоторое утверждение о бесконечной последовательности ω, которое может быть истинным или ложным для каждой конкретной последовательности ω.

Под законом теории вероятностей мы понимаем утверждение $\Phi(\omega)$, которое истинно для почти всех ω. [3] Пример такого закона – усиленный закон больших чисел. В последующем мы также рассмотрим закон повторного логарифма и эргодическую теорему Биркгофа.

Напомним, что $\mathcal{L}$ обозначает множество всех бесконечных двоичных последовательностей, случайных по Мартин-Лефу. Эквивалентно, это множество всех последовательностей, которые выдерживают универсальный тест Мартин-Лефа.

Алгоритмический подход к теории вероятностей предлагает более точную формулировку вероятностных законов. Закон может быть сформулирован в форме, свободной от понятия вероятностного распределения:

$$\omega \in \mathcal{L} \Longrightarrow \Phi(\omega)$$

для всех ω.

[2] Напомним, что понятие случайной по Мартин-Лефу последовательности было определено без использования универсального теста.

[3] В этом разделе мы для простоты рассматриваем равномерную меру на множестве Ω всех бесконечных двоичных последовательностей. Выражение «почти всюду» означает «за исключением множества меры нуль».

4.3. Задачи и упражнения

1. Доказать, что следующие множества бесконечных последовательностей являются эффективно нулевыми:

(a) $\{0\omega_2 0\omega_3 0 \cdots : \omega_i \in \{0,1\}\}$;

(b) $\{0^\infty, 1^\infty\}$;

(c) множество, состоящее из одной вычислимой последовательности;

(d) множество всех вычислимых последовательностей;

(e) множество, состоящее из всех бесконечных последовательностей ω таких, что $K(\omega^n) \leqslant \log n + O(1)$ для всех n;

(f) множество, состоящее из всех бесконечных последовательностей ω таких, что $K(\omega^n) \leqslant f(n)$ для всех n. Для каких функций f это верно?

2. Докажите, что объединение и пересечение конечного числа эффективно нулевых множеств также является эффективно нулевым множеством.

3. Докажите, что если некоторая бесконечная последовательность $\omega = \omega_1\omega_2\ldots$ является случайной, то

(a) 0ω, 1ω, $x\omega$ – также случайные последовательности, где $x = x_1 \ldots x_k$ – конечная последовательность;

(b) $\omega' = \omega_1 \ldots \omega_{n-1} x_1 \ldots x_k \omega_n \ldots$ – также случайная последовательность;

(c) также является случайной последовательностью ω', у которой каждый бит противоположен соответствующему биту последовательности ω;

(d) последовательность $\omega_n\omega_{n+1}\ldots$ является случайной для любого n; верно и обратное: для любого n, если последовательность $\omega_n\omega_{n+1}\ldots$ случайная, то $\omega_1\omega_2\ldots$ – также случайная последовательность.

(e) является ли случайной последовательность вида $\omega_1\omega_1\omega_2\omega_2\ldots$?

4. Пусть A – разрешимое множество и $\alpha = \alpha_1\alpha_2\ldots$ – его характеристическая последовательность. Доказать, что α не является случайной последовательностью.

5. Пусть последовательность $\omega = \omega_1\omega_2\ldots$ является случайной и $n_1 < n_2 < \ldots$ – вычислимая последовательность номеров. Тогда последовательность $\omega_{n_1}\omega_{n_2}\ldots$ случайная.

6. Пусть $\omega = \omega_1\omega_2\ldots$ случайная последовательность, а последовательность $\alpha = \alpha_1\alpha_2\ldots$ вычислимая. Тогда последовательность $\omega \oplus \alpha$ случайная. Здесь $\omega \oplus \alpha = \omega_1 \oplus \alpha_1\omega_2 \oplus \alpha_2\ldots$ и $\oplus$ – сложение по модулю 2 ($0 \oplus 0 = 0$, $0 \oplus 1 = 1$, $1 \oplus 1 = 0$).

7. Пусть A – перечислимое множество и $\omega = \omega_1\omega_2\ldots$ – его характеристическая последовательность, где

$$\omega_i = \begin{cases} 1, \text{ если } i \in A, \\ 0 \text{ в противном случае.} \end{cases}$$

Доказать, что последовательность ω не является случайной. Построить тест Мартин-Лефа, который отвергает такие последовательности.

8. Тест Соловея – это вычислимая последовательность строк $x_1, x_2, \ldots$ такая, что $\sum\limits_{n=1}^{\infty} 2^{-l(x_n)} \leqslant 1$.

Бесконечная последовательность ω выдерживает тест Соловея, если $x_n \subset \omega$ для не более чем конечного числа различных n. В противном случае тест Соловея отвергает последовательность ω.

Доказать, что бесконечная последовательность ω является случайной по Мартин-Лефу тогда и только тогда, когда выдерживает любой тест Соловея.

Доказать, что для любого теста Соловея $x_1, x_2, \ldots$ семейство множеств

$$U_m = \{\omega : x_n \subset \omega \text{ для } \geqslant 2^m \text{ различных } n\}$$

является тестом Мартин-Лефа, который отвергает то же множество последовательностей.

Покажите, как из теста Мартин-Лефа построить тест Соловея, который отвергает то же множество последовательностей.

9. Сформулировать свойство универсальности для тестов Соловея и привести независимую конструкцию такого теста.

10. Доказать, что в определении теста Мартин-Лефа можно потребовать $L(U_m) \leqslant \rho(m)$ для всех m, где $\rho(m)$ – произвольная вычислимая функция такая, что $\rho(m) \to 0$ при $m \to \infty$. Это определение теста приводит к тому же классу случайных последовательностей.

11. Для любого теста Мартин-Лефа $\{U_m\}$ можно определить его численное представление – функцию

$$d(\omega) = \sup\{m : \omega \in U_m\}. \tag{4.3}$$

Эту функцию также называем тестом. Доказать, что

(a) $U_m = \{\omega : d(\omega) \geqslant m\}$, где тест d определен по (4.3).

(b) $L\{\omega : d(\omega) = \infty\} = 0$ для любого m.

(c) Бесконечная последовательность ω отвергается тестом $\{U_m\}$ тогда и только тогда, когда $d(\omega) = \infty$.

12. Обозначим посредством $\hat{d}(\omega)$ численное представление универсального теста Мартин-Лефа. Тогда для численного представления $d(\omega)$ любого теста Мартин-Лефа существует такая константа c что $c\hat{d}(\omega) \geqslant d(\omega)$ для всех ω.

Глава 5

Специальные виды алгоритмической сложности

В этой главе мы дадим эквивалентное описание случайных по Мартин-Лефу последовательностей с помощью алгоритмической сложности. Тем самым первоначальная программа Колмогорова по сложностному определению случайности будет выполнена.

Естественный путь для сложностного определения случайности заключается в перенесении понятия несжимаемой конечной последовательности на бесконечные последовательности. В этом случае следовало бы считать случайной бесконечную последовательность ω, для которой было бы выполнено $\mathrm{K}(\omega^n) = n + O(1)$, где $\omega^n = \omega_1 \ldots \omega_n$ – последовательность первых n битов бесконечной последовательности ω.

Простая колмогоровская сложность $\mathrm{K}(x)$ не подходит для такого определения, так как для любой бесконечной последовательности ω выполнено $K(\omega^n) \leqslant n - \log n + O(1)$ для бесконечно многих n (см. задачу 16 из раздела 3.6). В связи с этим мы рассмотрим два модифицированных варианта колмогоровской сложности – префиксную и монотонную сложности. С по-

мощью каждой из этих сложностей можно будет дать определение случайной последовательности, эквивалентное определению Мартин-Лефа.

5.1. Префиксное декодирование

Классическая теория информации решает задачу построения кодовых слов для символов конечного алфавита, на котором формулируются передаваемые сообщения. При этом используются коды, которые допускают возможность декодирования. Простейшее достаточное условие для существования алгоритма декодирования заключается в том, что кодовые слова не должны продолжать друг друга – не должны быть префиксами друг друга. В этом случае мы можем закодировать сообщение простой последовательностью слов, кодирующих передаваемые символы. Благодаря безпрефиксности кодовых слов простейший алгоритм может декодировать закодированное сообщение просто читая его подряд и проверяя на совпадение с возможными кодовыми словами.

В этом разделе мы рассмотрим безпрефиксные (или префиксно-корректные) методы декодирования конечных объектов и соответствующую префиксную сложность, которая является модификацией колмогоровской сложности.

5.1.1. Префиксная сложность

Множество строк C называется безпрефиксным, если для любых $p, q \in C$ будет $p \not\sqsubseteq q$.

Мы будем рассматривать способы декодирования, для которых множества кодов являются безпрефиксными.

Функция $B(p, y)$ называется *префиксно-корректной* (по первому аргументу), если для любого y множество всех p, для которых $B(p, y)$ определена, является безпрефиксным. Здесь аргумент y рассматривается в качестве параметра.

Это условие эквивалентно тому, что для любых p и p', если $B(p,y)$ и $B(p',y)$ определены, то $p \not\subseteq p'$ и $p' \not\subseteq p$.

Пусть $B(p,y)$ – произвольная вычислимая префиксно-корректная функция. Определим меру (условной) сложности относительно функции B:

$$\mathrm{KP}_B(x|y) = \min\{l(p) : B(p,y) = x\}.$$

Здесь считаем, что $\min \emptyset = \infty$.

Для мер сложности относительно префиксно-корректных функций также имеет место теорема инвариантности.

Теорема 5.1. *Существует вычислимая префиксно-корректная функция $A(p,y)$ такая, что для любой вычислимой префиксно-корректной функции $B(p,y)$ выполнено*

$$\mathrm{KP}_A(x|y) \leqslant \mathrm{KP}_B(x|y) + c$$

для всех x и y, где c – константа, зависящая от B (но не зависящая от x и y).

Доказательство. Схема доказательства та же, что и у доказательства теоремы инвариантности для простой колмогоровской сложности. Предварительно нужно показать, что существует универсальная функция для класса всех вычислимых префиксно-корректных функций.

Лемма 5.1. *Существует вычислимая префиксно-корректная по первому аргументу p функция $\tilde{U}(q,p,y)$ такая, что для любой вычислимой префиксно-корректной по p функции $B(p,y)$ существует q такое, что $B(p,y) = \tilde{U}(q,p,y)$ для всех p и y.*

Доказательство. Пусть $U(q,p,y)$ – вычислимая функция, универсальная для всех вычислимых функций от двух аргументов p и y.

Заметим, что функция вычислима тогда и только тогда, когда ее график является перечислимым множеством.

Развернем процесс перечисления графика универсальной функции $U(q,p,y)$. Одновременно перечисляем график функции $\tilde{U}(q,p,y)$.

На каждом шаге s просматриваем только один набор (q,p,y) и делаем дополнительный шаг вычисления значения $U(q,p,y)$, если это значение еще не определилось на предыдущих шагах. Просмотр наборов (q,p,y) осуществляем каким-либо вычислительно эффективным образом. При этом, каждый набор (q,p,y) просматриваем на бесконечном числе шагов процесса.

FOR $s = 1, 2, \ldots$

Пусть на шаге s просматривается набор (q,p,y).

Если значение $\tilde{U}(q,p,y)$ было определено на предыдущих шагах, то на шаге s ничего не делаем.

Если значение универсальной функции $U(q,p,y)$ впервые определилось на шаге s и множество, состоящее из p и всех p' таких, что значения $U(q,p',y)$ определены на предыдущих шагах, является безпрефиксным, то определим $\tilde{U}(q,p,y) = U(q,p,y)$.[1] В противном случае, значение $\tilde{U}(q,p,y)$ навсегда остается неопределенным.

ENDFOR

По построению функция $\tilde{U}(q,p,y)$ обладает свойствами:

- для любых q и y функция $\tilde{U}(q,p,y)$ является префиксно-корректной по p;
- для любых q и y, если функция $U(q,p,y)$ является префиксно-корректной по p, то $\tilde{U}(q,p,y) = U(q,p,y)$.
- если функция $B(p,y)$ является префиксно-корректной по p, то найдется такое q, что $B(p,y) = \tilde{U}(q,p,y)$ выполнено для всех p и y;

Лемма доказана. □

Переходим к доказательству теоремы 5.1.

[1] На каждом шаге s процесса значения $U(q,p,y)$ определены только для конечного числа троек (q,p,y).

Пусть $\tilde{U}(q,p,y)$ – функция универсальная для всех префиксно-корректных функций. Определим способ декодирования

$$A(\overline{q}01p,y) = \tilde{U}(q,p,y)$$

для всех $p,q,y \in \{0,1\}^*$.[2]

Для всех остальных входов, не имеющих вида $\overline{q}01p,y$, значения $A(\overline{q}01p,y)$ не определены.

Легко видеть, что эта функция $A(p,y)$ является префиксно-корректной по p.

Пусть $B(p,y)$ – произвольная вычислимая префиксно-корректная по p функция. Тогда по лемме 5.1 для некоторого q

$$B(p,y) = \tilde{U}(q,p,y)$$

для всех p и y. Отсюда вытекает, что если для некоторой строки x выполнено $B(p,y) = x$, то $A(\overline{q}01p,y) = x$. Отсюда

$$\mathrm{KP}_A(x|y) \leqslant \mathrm{KP}_B(x|y) + 2l(q) + 2.$$

Соответствующая константа c имеет вид: $c = 2l(q) + 2$. □

Префиксно-корректный метод декодирования, удовлетворяющий заключению теоремы 5.1, называется оптимальным (или универсальным). Фиксируем один из оптимальных префиксно-корректных методов декодирования A и назовем соответствующую меру сложности $\mathrm{KP}_A(x|y)$ условной префиксной сложностью. В дальнейшем нижний индекс опускаем.

Определим безусловную префиксную сложность

$$\mathrm{KP}(x) = \mathrm{KP}(x|\lambda).$$

В разделе 3.2 было получено простейшее неравенство $\mathrm{K}(x) \leqslant l(x)+O(1)$ для простой колмогоровской сложности. В случае префиксной сложности, доказательство из раздела 3.2 не годится,

[2]Здесь имеется ввиду, что обе части равенства определены или неопределены одновременно.

так как способ декодирования $B(p) = p$ не является префиксно-корректным. Можно его модифицировать: $B(\overline{\mathrm{str}(l(p))}01p) = p$. Легко проверить, что коды вида $\overline{\mathrm{str}(l(p))}01p$ образуют безпрефиксное множество. Тогда $\mathrm{KP}(x) \leqslant l(x)+2\log l(x)+O(1)$. Можно также получить оценку $\mathrm{KP}(x) \leqslant l(x)+\log l(x)+2\log\log l(x)+O(1)$ и так далее. Из дальнейшего будет следовать, что слагаемое $\log l(x)$ невозможно устранить (см. раздел 5.1.3).

Приведем соотношения, связывающие префиксную и простую условные колмогоровские сложности:

Предложение 5.1.

$$\mathrm{K}(x|y) - O(1) \leqslant \mathrm{KP}(x|y) \leqslant \mathrm{K}(x|y) + 2\log \mathrm{K}(x|y) + O(1).$$

Доказательство. Первое неравенство выполнено, так как для задания простой колмогоровской сложности используется более широкий класс методов декодирования.

Для доказательства второго неравенства нам необходимо по универсальному способу декодирования $A(p, y)$ для простой колмогоровской сложности построить префиксно-корректный способ декодирования. Определим префиксно-корректный метод декодирования следующим образом:

$$B(\overline{\mathrm{str}(l(p))}01p, y) = A(p, y).$$

Утверждение доказано. □

Отметим также следующее неравенство для условной префиксной сложности.

Предложение 5.2.

$$\mathrm{KP}(x|l(x)) \leqslant l(x) + O(1).$$

Доказательство. Определим префиксно-корректную по первому аргументу функцию

$$B(p, n) = \begin{cases} p, \text{ если } l(p) = n, \\ \text{неопределено, в противном случае,} \end{cases}$$

Тогда $\mathrm{KP}(x|l(x) \leqslant \mathrm{KP}_B(x|l(x)) + O(1) = l(x) + O(1)$. $\square$

Более нетривиальные соотношения для префиксной сложности будут доказаны в следующих разделах.

5.1.2. Модель вычисления

Существует интерпретация префиксно-корректных способов декодирования в терминах машин Тьюринга (МТ). Для простоты рассмотрим безусловные методы декодирования.

Значения функции $B(p)$, задающей некоторый метод декодирования, можно вычислять на МТ с одной входной и одной рабочей лентой. Входное слово p помещается на входной ленте, вычисления производятся на рабочей ленте, результат вычисления помещается там же.

При обычной интерпретации границы входного слова каким-либо образом обозначены на ленте: в начале работы головка МТ обозревает первую букву входного слова p; входное слово ограничено справа маркером конца входа. По этой причине нам приходится использовать разделитель для пары входных слов – МТ не сможет разделить два входных слова, если они записаны подряд без разделения.

Оказывается, что функции $B(p)$, представляющие префиксно-корректные методы декодирования, можно вычислять на МТ без маркера конца входа.

Уточним, что это значит. В процессе вычисления на такой МТ головка читает входное слово и производит вычисления на рабочей ленте. При этом в том случае, когда МТ останавливается и печатает результат, выполнены условия:

- в процессе вычисления головка входной ленты никогда не выходит за пределы входного слова,
- в момент остановки головка обозревает последнюю букву входного слова.

Данная интерпретация была предложена Чейтиным [38].

Мы сформулируем соответствующее утверждение.

Предложение 5.3. *Любая префиксно-корректная вычислимая функция вычислима на МТ без маркера конца входа.*

Доказательство. Пусть префиксно-корректная функция $B(p)$ вычислима на МТ $\mathcal{M}_1$, которая использует маркер конца входа.

Мы неформально опишем работу МТ $\mathcal{M}_2$, которая моделирует работу МТ $\mathcal{M}_1$ и при этом не использует маркер конца входа. Моделирование происходит в виде следующего цикла.

Пусть $\mathcal{M}_2$ уже прочитала часть входа $p' \subseteq p$. Запускаем $\mathcal{M}_1$ на всех возможных входах x таких, что $p' \subseteq x$, параллельно.

Если на некотором входе x машина $\mathcal{M}_1$ остановилась и выдала слово y, делаем следующее:

(a) Если $p' \subset x$, то машина $\mathcal{M}_2$ читает следующий бит b входа, переходит на начало цикла и в дальнейшем обрабатывает слово $p'' = p'b$. Нарушения условия не происходит, так как из префиксной корректности функции B следует, что машина $\mathcal{M}_1$ не может быть определена на слове p' и мы можем читать следующий бит входа без нарушения требований к машине $\mathcal{M}_2$.

(b) Если $p' = x$, то машина $\mathcal{M}_2$ выдает y и останавливается. Если при этом $p' = p$, то вычисление проведено корректно, так как в этом случае головка машины $\mathcal{M}_2$ стоит на последней букве входного слова и машина выдала результат. Если $p' \subset p$, то из-за префиксной корректности функции B машина $\mathcal{M}_1$, а тем самым и машина $\mathcal{M}_2$, не может быть определена на p и условие на $\mathcal{M}_2$ по-прежнему выполнено.

По конструкции машина $\mathcal{M}_2$ заканчивает работу и выдает результат не выходя за пределы входного слова. Утверждение доказано. □

Приведем пример применения рассмотренной модели вычисления для доказательства одного из неравенств для префиксной сложности пары.

Для префиксной сложности имеет место неравенство

$$\mathrm{KP}(x, y) \leqslant \mathrm{KP}(x) + \mathrm{KP}(y) + O(1). \tag{5.1}$$

Для доказательства рассмотрим префиксно-корректный метод декодирования пар. Пусть по программе p можно восстановить слово x, а по программе q можно восстановить слово y.

По предложению 5.3 можно считать, что слово x восстанавливается по слову p с помощью машины $\mathcal{M}_1$ а слово y восстанавливается по слову q с помощью машины $\mathcal{M}_2$. Обе машины не используют маркер конца входа. Тогда в качестве программы для восстановления пары можно рассмотреть слово pq. Сначала машина $\mathcal{M}_1$ применяется к слову pq и выдает x. При этом ее головка обозревает последнюю букву слова p, и тем самым мы знаем начало слова q и можем применить машину $\mathcal{M}_2$ к слову q и получить y. Отсюда следует неравенство (5.1).

5.1.3. Априорная полумера на дискретном множестве

Префиксной сложности соответствует двойственное понятие – понятие перечислимого распределения вероятностей на множестве всех натуральных чисел.

Напомним, что $\mathcal{R}$ обозначает множество всех действительных чисел, $\mathcal{N}$ – множество всех натуральных чисел. Обозначим $\mathcal{R}_+$ множество всех неотрицательных действительных чисел, $\mathcal{N}_+$ – множество всех натуральных чисел вместе с 0, которое отождествлено с множеством всех двоичных слов Ξ.

Распределение вероятностей на множестве натуральных чисел – это функция $f : \mathcal{N}_+ \to \mathcal{R}_+$, удовлетворяющая условию

$$\sum_{n=0}^{\infty} f(n) = 1.$$

Из технических соображений нам придется ослабить как понятие распределения на множестве натуральных чисел, так и понятие его вычислимости.

Эффективная вычислимость функций типа $f : \mathcal{N}_+ \to \mathcal{R}$. Мы введем структуру эффективной вычислимости для функций с вещественными значениями. Область определения функ-

ции типа $f : \mathcal{N}_+ \to \mathcal{R}$ состоит из конструктивных объектов. [3] Значения такой функции не являются конструктивными объектами. Конструктивными являются естественные приближения к вещественным числам – рациональные числа. Существуют различные уровни эффективной вычислимости для функций типа $f : \mathcal{N}_+ \to \mathcal{R}$.

Функция $f : \mathcal{N}_+ \to \mathcal{R}$ называется перечислимой снизу, если множество $\{(r, x) : r < f(x)\}$, где r обозначает рациональное число, является перечислимым.

Другое эквивалентное определение перечислимой снизу функции следующее.

Последовательностью функций $f_n(x)$ называется вычислимой, если функция $f(n, x) = f_n(x)$ вычислимая.

Предложение 5.4. *Функция f перечислима снизу, тогда и только тогда, когда существует неубывающая вычислимая последовательность функций $f_n(x)$ с рациональными значениями, т.е. $f_{n+1}(x) \geqslant f_n(x)$ для всех n и x такая, что $f(x) = \lim\limits_{n\to\infty} f_n(x)$ для всех x.*

Доказательство. Пусть $A = \{(r, x) : r < f(x)\}$ и A^n – конечное подмножество A, перечисленное за n шагов перечисления. Определим $f_n(x) = \max(\{r : (r, x) \in A^n\} \cup \{0\})$. Легко проверить, что последовательность $f_n(x)$ удовлетворяет условию предложения.

Для доказательства обратного утверждения, пусть $f(x) = \lim_{n\to\infty} f_n(x)$ и $f_n(x) \leqslant f_{n+1}(x)$ для всех x и n. Тогда из вычислимости $f_n(x)$ по n и x следует, что

$$\{(r, x) : r < f(x)\} = \{(r, x) : \exists n(r < f_n(x))\}$$

– перечислимое множество. □

Аналогично функция $f : \mathcal{N}_+ \to \mathcal{R}$ называется перечислимой сверху, если множество $\{(r, x) : r > f(x)\}$, где r обозначает рациональное число, является перечислимым. Другое эквивалентное

[3] Напомним установленное в разделе 3.1 соответствие между натуральными числами и двоичными строками.

определение перечислимой сверху функции следующее. Функция f перечислима сверху, если существует невозрастающая вычислимая последовательность функций $f_n(x)$ с рациональными значениями (т.е. $f_{n+1}(x) \leqslant f_n(x)$ для всех n и x) такая, что $f(x) = \lim\limits_{n\to\infty} f_n(x)$ для всех x.

Функция $f : \mathcal{N}_+ \to \mathcal{R}$ называется вычислимой, если она одновременно перечислима снизу и сверху. Для такой функции существует алгоритм, который по входу x и произвольному положительному рациональному числу ϵ вычисляет рациональное приближение к $f(x)$ с точностью до ϵ. Действительно, достаточно перечислять рациональные числа $r_1 < f(x)$ и $r_2 > f(x)$ до тех пор, пока не найдутся такие два числа, для которых $r_2 - r_1 < \epsilon$. Любое из них можно взять в качестве необходимого рационального приближения к значению $f(x)$.

Перечислимые полумеры. Функция $Q : \mathcal{N}_+ \to \mathcal{R}_+$ называется перечислимой снизу полумерой, если она перечислима снизу и

$$\sum_{n=0}^{\infty} Q(n) \leqslant 1.$$

В множестве всех перечислимых снизу полумер существует универсальный объект – максимальная с точностью до мультипликативной константы перечислимая снизу полумера.

Теорема 5.2. *Существует такая перечислимая снизу полумера P, что для любой перечислимой снизу полумеры Q найдется константа c такая, что*

$$cP(x) \geqslant Q(x)$$

для всех x. Здесь константа c зависит от полумеры Q.

Из свойств эффективности следует, что множество всех перечислимых снизу полумер счетно. Мы покажем, что все перечислимые снизу полумеры можно перечислять снизу равномерно одним алгоритмом.

Лемма 5.2. *Существует такая последовательность полумер P_i, что*

- *множество $\{(i, r, x) : r < P_i(x)\}$ перечислимо (r – рациональное);*[4]
- *для любой перечислимой снизу полумеры Q найдется такое i, что $Q = P_i$.*

Доказательство. Как и прежде, для построения такой последовательности полумер используем универсальную функцию. Рассмотрим перечислимые множества, состоящие из пар (r, x), где r – рациональное число, x – натуральное число (точнее, отождествленная с ним строка). Мы также используем какое-нибудь отождествление рациональных чисел и двоичных строк. Пары двоичных строк также отождествлены со строками.

Пусть $U(i, r, x)$ – универсальная функция. Каждое перечислимое множество пар (r, x) имеет вид W_i, которое есть область определения функции $U(i, r, x)$ при фиксированном i (см. раздел 3.1).

Запустим процесс вычисления всех значений универсальной функции $U(i, r, x)$ в виде цикла, на каждой итерации которого просматривается только одна тройка (i, r, x) и моделируется один шаг машины Тьюринга, вычисляющей значение $U(i, r, x)$. При этом каждая тройка (i, r, x) просматривается на бесконечном числе итераций цикла. Если на очередной итерации s цикла значение $U(i, r, x)$ впервые определено, то определим множество $W_i^s = W_i^{s-1} \cup \{(r, x)\}$; полагаем $W_i^s = W_i^{s-1}$, в противном случае. Пусть $W_i^0 = \emptyset$. Таким образом, W_i^s обозначает конечное подмножество пар, перечисленных в W_i за s шагов процесса вычисления (графика) универсальной функции. Определим

$$f_i^s(x) = \max(\{r : (r, x) \in W_i^s\} \cup \{0\}).$$

Так как W_i^s – конечное множество для любых i и s, $f_i^s(x) > 0$ только для не более конечного числа различных x. По определению $f_i^s(x) \leqslant f_i^{s+1}(x)$ для всех i, s, x.

[4] Говорим, что полумеры P_i *равномерно перечислимые снизу.*

Для каждого i определим $P_i^0(x) = 0$ для всех x, при $s > 0$ определим $P_i^s(x) = f_i^s(x)$ для всех x, если

$$\sum_y f_i^s(y) \leqslant 1, \tag{5.2}$$

определим $P_i^s(x) = P^{s-1}(x)$ в противном случае. Заметим, что если условие (5.2) нарушается для какого-нибудь s, то оно не выполнено и для больших s, поэтому $P_i^s(x) = P^{s-1}(x)$ для всех таких s.

Из определения следует, что для любого x выполнено $P_i^s(x) \leqslant P_i^{s+1}(x)$ для всех i, s. Для произвольного i определим

$$P_i(x) = \sup_s P_i^s(x) = \lim_{s\to\infty} P_i^s(x)$$

для всех x. Тогда $\sum\limits_y P_i(y) \leqslant 1$, так как $\sum\limits_y P_i^s(y) \leqslant 1$ для любого s. Кроме этого, $r < P_i(x)$ тогда и только тогда, когда $r < P_i^s(x)$ для некоторого s. По определению, последовательность рациональных чисел $P_i^s(x)$ вычислима как функция от i, s и x. Поэтому множество $\{(i, r, x) : r < P_i(x)\}$ перечислимо.

Для любой перечислимой снизу полумеры Q имеем

$$W_i = \{(r, x) : r < Q(x)\}$$

для некоторого i. Тогда условие (5.2) всегда выполнено и поэтому $Q(x) = P_i(x)$ для всех x. □

Доказательство теоремы. Определим

$$P(x) = \sum_{i=1}^{\infty} \frac{1}{i(i+1)} P_i(x).$$

Функция P перечислима снизу как двойной предел неубывающей вычислимой последовательности функций с рациональными значениями. Здесь мы используем некоторое обобщение предложения 5.4, согласно которому $P_i(x) = \lim_{n\to\infty} f_n(i, x)$, где $f_n(i, x)$

– вычислимая по x, i, n и неубывающая по n последовательность функций с рациональными значениями.

Она является полумерой, так как

$$\sum_x P(x) = \sum_x \sum_{i=1}^{\infty} \frac{1}{i(i+1)} P_i(x) =$$
$$= \sum_{i=1}^{\infty} \frac{1}{i(i+1)} \sum_x P(x) \leqslant \sum_{i=1}^{\infty} \frac{1}{i(i+1)} = 1.$$

Для любой перечислимой снизу полумеры Q имеет место равенство $Q = P_i$ для некоторого i. Поэтому

$$i(i+1)P(x) \geqslant Q(x)$$

для всех x. Теорема доказана. □

В дальнейшем мы зарезервируем обозначение P для максимальной перечислимой снизу полумеры. Мы будем называть P *априорной полумерой* на множестве всех натуральных чисел и нуле (множестве отождествленных с ними двоичных строк), поскольку она приписывает самую большую полувычислимую вероятность натуральному числу.

Члены любого перечислимого снизу сходящегося ряда определяют оценки снизу для априорной полумеры. Например, можно утверждать, что для некоторой константы $c > 0$

$$P(n) \geqslant \frac{1}{cn \log n (\log \log n)^2}$$

для всех $n \geqslant 3$, так как

$$\sum_{n=3}^{\infty} \frac{1}{n \log n (\log \log n)^2} < \infty.$$

Из этой оценки получаем следствие.

Следствие 5.1. *$P(x) > 0$ для всех x.*

5.1.4. Двойственность

Следующая теорема о двойственном представлении префиксной сложности указывает на связь между префиксной сложностью и априорной полумерой.

Теорема 5.3. $\mathrm{KP}(x) = -\log P(x) + O(1)$.

Доказательство. Если $x \neq y$, то по определению $\mathrm{KP}(x)$ и $\mathrm{KP}(y)$ – длины двух попарно несравнимых последовательностей p_x и p_y. Поэтому

$$\sum_x 2^{-\mathrm{KP}(x)} = \sum_{x \in \Xi}^{\infty} L(\Gamma_{p_x}) \leqslant 1,$$

где все двоичные строки p_x, $x \in \Xi$, попарно несравнимы, поэтому интервалы Γ_{p_x} попарно не пересекаются.

Функция $Q(x) = 2^{-\mathrm{KP}(x)}$ перечислима снизу. Следовательно, функция $Q(x)$ есть перечислимая снизу полумера на натуральных числах. Отсюда следует, что

$$cP(x) \geqslant 2^{-\mathrm{KP}(x)}$$

для всех x, где c – константа. Таким образом, мы доказали, что

$$-\log P(x) \leqslant \mathrm{KP}(x) + O(1).$$

Для доказательства противоположного неравенства нам потребуется вспомогательное утверждение. Известное в теории информации неравенство Крафта заключается в том, что для любой конечной последовательности натуральных чисел $k_1, \ldots, k_n$ такой, что

$$\sum_{i=1}^{n} 2^{-k_i} \leqslant 1,$$

существуют попарно несравнимые двоичные строки $p_1, \ldots, p_n$ такие, что $l(p_i) = k_i$ при $i = 1, \ldots, n$. Данное неравенство используется в теории информации для эффективного построения

префиксных кодов [5] . Поскольку мы строим коды для бесконечного множества всех двоичных слов, нам потребуется обобщенное неравенство Крафта.

Лемма 5.3. *Для любой вычислимой последовательности натуральных чисел $k_1, k_2 \ldots$ такой, что*

$$\sum_{i=1}^{\infty} 2^{-k_i} \leqslant 1,$$

можно построить вычислимую последовательность попарно несравнимых двоичных строк $p_1, p_2, \ldots$ такую, что $l(p_i) = k_i$ при $i = 1, 2, \ldots$

Доказательство. Построим нужную последовательность по индукции. Пусть $p_1, \ldots, p_n$ уже определены так, что выполнено $l(p_i) = k_i$ при $i = 1, \ldots, n$. Найдем p_{n+1}. Используем еще одно предположение индукции: существуют конечные последовательности $t_1, \ldots, t_m$ такие, что $l(t_i) \neq l(t_j)$ при $i \neq j$ и

$$\Omega \setminus \cup_{i=1}^{n} \Gamma_{p_i} = \cup_{i=1}^{m} \Gamma_{t_i}.$$

Среди слов $t_1, \ldots, t_m$ обязательно найдется слово t_i, для которого $l(t_i) \leqslant k_{n+1}$, так как иначе

$$L(\cup_{i=1}^{m} \Gamma_{t_i}) < \sum_{s > k_{n+1}} 2^{-s} = 2^{-k_{n+1}}$$

и тогда для их дополнения

$$L(\cup_{i=1}^{n} \Gamma_{p_i}) > 1 - 2^{-k_{n+1}},$$

откуда $\sum\limits_{j=1}^{n+1} 2^{-k_j} > 1$. Получаем противоречие с условием леммы.

Пусть t_1 – самое длинное слово с $l(t_1) \leqslant k_{n+1}$.

[5] Это неравенство на длины кодовых слов является необходимым и достаточным условием существования однозначно декодируемого кода.

Если $l(t_1) = k_{n+1}$, то определим $p_{n+1} = t_1$. В этом случае,

$$\Omega \setminus \cup_{i=1}^{n+1} \Gamma_{p_i} = \cup_{i=2}^{m} \Gamma_{t_i},$$

т.е. предположение индукции выполнено.

Если $l(t_1) < k_{n+1}$, то представим интервал Γ_{t_1} в виде объединения попарно несравнимых интервалов $\Gamma_{a_1}, \ldots, \Gamma_{a_s}$ так, что $l(a_s) = k_{n+1}$. Нетрудно проверить, что это всегда можно сделать. Определим $p_{n+1} = a_s$. Имеем

$$\Omega \setminus \cup_{i=1}^{n+1} \Gamma_{p_i} = \cup_{i=2}^{m} \Gamma_{t_i} \cup_{i=1}^{s-1} \Gamma_{a_i}.$$

Длины всех a_i различные и больше длины t_1. Длины всех t_i при $i > 1$ меньше чем длина t_1. Поэтому все a_i и t_i при $i > 1$ a_i имеют различные длины. Предположение индукции выполнено Лемма доказана. $\square$

Переходим к доказательству теоремы. Построим префиксно-корректную функцию $B(p)$ такую, что

$$\mathrm{KP}_B(x) \leqslant -\log P(x) + O(1).$$

Для этого перечисляем без повторения все пары (m, x) такие, что

$$2^{-m} < \frac{1}{2} P(x).$$

Так как $P(x) > 0$ для всех x, таких пар бесконечно много. Пусть (m_k, x_k) – k-я пара при таком перечислении. Тогда

$$\sum_{k=1}^{\infty} 2^{-m_k} = \sum_{x} \sum_{x_k = x} 2^{-m_k} \leqslant$$
$$\leqslant \sum_{x} 2^{-s(x)+1} \leqslant \sum_{x} P(x) \leqslant 1,$$

где $s(x) = \min\{m_k : x_k = x\}$. Для этой величины выполнено $2^{-s(x)} < \frac{1}{2} P(x)$ и $2^{-s(x)+1} \geqslant \frac{1}{2} P(x)$.

Так как $P(x) > 0$ для всех x, для каждого x существует k такое, что $x_k = x$.

По лемме 5.3 существует вычислимая последовательность

$$p_1, p_2, \ \ldots$$

попарно несравнимых слов, для которых $l(p_k) = m_k$ для всех k. Определим функцию $B(p)$:

$B(p_k) = x_k$ для всех k. Для остальных входов значение функции не определено. Тогда

$$\mathrm{KP}_B(x) = \min\{m_k : x_k = x\} = s(x).$$

Имеем $2^{-s(x)} \geqslant \frac{1}{4}P(x)$. Эквивалентно, $2^{-\mathrm{KP}_B(x)} \geqslant \frac{1}{4}P(x)$ или $\mathrm{KP}_B(x) \leqslant -\log P(x)+2$. Отсюда получаем утверждение теоремы $\mathrm{KP}(x) \leqslant -\log P(x) + O(1)$. $\square$

Можно рассмотреть условные перечислимые снизу полумеры $Q(x|y)$, для которых выполнены условия:

- множество $\{(r, x, y) : r < Q(x|y)\}$ перечислимо;
- $\sum\limits_x Q(x|y) \leqslant 1$ для каждого y.

Аналогичным образом доказывается, что существует максимальная с точностью до мультипликативной константы условная полумера $P(x|y)$ такая, что для любой перечислимой снизу условной полумеры $Q(x|y)$ существует константа c такая, что $cP(x|y) \geqslant Q(x|y)$ для всех x и y.

Непосредственным образом проверяется, что теорема двойственности имеет место для условных префиксной сложности и априорной полумеры.

Теорема 5.4. $\mathrm{KP}(x|y) = -\log P(x|y) + O(1)$.

Удобно использовать результат теоремы 5.3 для доказательства неравенств для префиксной сложности. Сформулируем необходимое следствие из теоремы 5.3.

Следствие 5.2. *Для любой перечислимой снизу последовательности вещественных чисел a_n, $n = 1, 2, \ldots$, такой что*

$$\sum_{n=1}^{\infty} a_n < \infty,$$

имеет место неравенство

$$\mathrm{KP}(n) \leqslant -\log a_n + O(1).$$

В качестве первого применения этого следствия докажем неравенство

$$\mathrm{KP}(n) \leqslant \log n + 2\log\log n + O(1). \tag{5.3}$$

Действитьно, ряд $\sum_{n=2}^{\infty} \frac{1}{n\log^2 n} < \infty$ – сходится. Неравенство (5.3) получается по следствию 5.2.

Можно усилить неравенство (5.3):

$$\mathrm{KP}(n) \leqslant \log n + \log\log n + 2\log\log\log n + O(1). \tag{5.4}$$

Оно следует из сходимости ряда $\sum_{n=4}^{\infty} \frac{1}{n\log n\log\log^2 n} < \infty$.

Можно доказать более точное неравенство

$$\mathrm{KP}(x) \leqslant l(x) + \mathrm{KP}(l(x)) + O(1).$$

Действительно, сходится ряд

$$\sum_x 2^{-l(x)-\mathrm{KP}(l(x))} = \sum_m \sum_{l(x)=m} 2^{-l(x)-\mathrm{KP}(l(x))} =$$
$$= \sum_m 2^m 2^{-m-\mathrm{KP}(m)} = \sum_m 2^{-\mathrm{KP}(m)} < \infty. \tag{5.5}$$

Для доказательства противоположных неравенств используем другое следствие из теоремы 5.3.

Следствие 5.3. *Для любой перечислимой снизу последовательности вещественных чисел a_n, $n = 1, 2, \ldots$, такой, что*

$$\sum_n a_n = \infty,$$

имеет место неравенство

$$\mathrm{KP}(n) \geqslant -\log a_n$$

для бесконечно многих n.

Доказательство. Действительно, по теореме 5.3, если $\mathrm{KP}(n) \leqslant -\log a_n + O(1)$, то имеет место $\infty > c\sum_n P(n) \geqslant \sum_n a_n = \infty$. Полученное противоречие доказывает следствие. □

Так как ряд

$$\sum_{n\geqslant 2} \frac{1}{n \log n} = \infty$$

расходится, для бесконечно многих n выполнено

$$\mathrm{KP}(n) \geqslant \log n + \log\log n.$$

Таким образом, неравенство (5.4) является почти не улучшаемым.

5.1.5. Префиксная сложность пары

Для префиксной сложности имеет место точное соотношение для декомпозиции сложности пары.

Теорема 5.5. $\mathrm{KP}(x, y) = \mathrm{KP}(x) + \mathrm{KP}(y|x, \mathrm{KP}(x)) + O(1)$.

Предварительно докажем две простые леммы.

Лемма 5.4. $\mathrm{KP}(x, y) \leqslant \mathrm{KP}(x) + \mathrm{KP}(y|x) + O(1)$.

Доказательство этой леммы предоставляется читателю в качестве задачи 9 из раздела 5.4.

Лемма 5.5. $\mathrm{KP}(x, \mathrm{KP}(x)) = \mathrm{KP}(x) + O(1)$.

Доказательство. Неравенство

$$\mathrm{KP}(x) \leqslant \mathrm{KP}(x, \mathrm{KP}(x)) + O(1)$$

очевидно.

Пусть p – самый короткий код для x. Тогда некоторый алгоритм может по p вычислить x и $\mathrm{KP}(x) = l(p)$. Отсюда

$$\mathrm{KP}(x, \mathrm{KP}(x)) \leqslant \mathrm{KP}(x) + O(1).$$

Лемма доказана. □

Пользуясь леммами 5.4 и 5.5, получим неравенство $\leqslant$:

$$\begin{aligned} \mathrm{KP}(x, y) &\leqslant \mathrm{KP}(y, x, \mathrm{KP}(x)) + O(1) \leqslant \\ &\leqslant \mathrm{KP}(x, \mathrm{KP}(x)) + \mathrm{KP}(y|x, \mathrm{KP}(x)) + O(1) = \\ &= \mathrm{KP}(x) + \mathrm{KP}(y|x, \mathrm{KP}(x)) + O(1). \end{aligned}$$

Для доказательства обратного неравенство мы воспользуемся двойственным представлением префиксной сложности через априорную полумеру.

Учитывая теорему 5.3, нам необходимо доказать неравенство

$$c_1 P(y|x, \mathrm{KP}(x)) \geqslant 2^{\mathrm{KP}(x)} P(x, y),$$

где c_1 – положительная константа.

Функция $Q(x) = \sum\limits_y P(x, y)$ является перечислимой снизу полумерой, так как

$$\sum_x Q(x) = \sum_{x,y} P(x, y) \leqslant 1.$$

Поэтому $c_2 P(x) \geqslant \sum\limits_y P(x, y)$ для некоторой положительной константы c_2. В частности, $c_3 2^{-\mathrm{KP}(x)} \geqslant \sum\limits_y P(x, y)$, для некоторой константы c_3 или

$$\sum_y c_3^{-1} 2^{\mathrm{KP}(x)} P(x, y) \leqslant 1 \tag{5.6}$$

для любого x.

Если бы функция $Q(y|x,m) = c_3^{-1} 2^m P(x,y)$ была перечислимой снизу полумерой, мы сразу получили бы необходимое неравенство

$$cP(y|x, \mathrm{KP}(x)) \geqslant 2^{\mathrm{KP}(x)} P(x,y)$$

для некоторой константы c. Однако данная функция удовлетворяет условию (5.6) только при $m = \mathrm{KP}(x)$. Мы преодолеем этот недостаток с помощью дополнительных построений.

Множество $W = \{(r,x,z) : r < P(x,z)\}$ перечислимо. Пусть W^t обозначает конечное подмножество его элементов, перечисленных за t шагов. Пусть

$$P^t(x,z) = \max(\{r : (r,x,y) \in W^t\} \cup \{0\}).$$

Определим полумеру $Q(y|x,m)$ следующим образом: по x, m и s найдем максимальное $t \leqslant s$ такое, что

$$c_3^{-1} 2^m \sum_{z \leqslant t} P^t(x,z) \leqslant 1,$$

и полагаем

$$Q^s(y|x,m) = \begin{cases} c_3^{-1} 2^m P^t(x,y), \text{ если } y \leqslant t, \\ 0, \text{ если } y > t. \end{cases}$$

По определению $Q^s(y|x,m) \leqslant Q^{s+1}(y|x,m)$ для всех x, m и s. Пусть

$$Q(y|x,m) = \sup_s Q^s(y|x,m).$$

Тогда нетрудно видеть, что функция Q перечислима снизу и

$$\sum_y Q(y|x,m) \leqslant 1$$

для любых x и m. Имеем для априорной полумеры

$$cP(y|x,m) \geqslant Q(y|x,m)$$

для некоторой константы c.

Полагаем в этом неравенстве $m = \mathrm{KP}(x)$. Так как согласно (5.6),

$$c_1^{-3} 2^{\mathrm{KP}(x)} \sum_{z \leqslant t} P^t(x,z) \leqslant$$
$$\leqslant c_3^{-1} 2^{\mathrm{KP}(x)} \sum_z P(x,z) \leqslant 1$$

для каждого t, выполнено

$$Q(y|x.\mathrm{KP}(x)) = c_3^{-1} 2^{\mathrm{KP}(x)} P(x,z).$$

Отсюда получаем

$$cP(y|x, \mathrm{KP}(x)) \geqslant 2^{\mathrm{KP}(x)} P(x,y)$$

для некоторой положительной константы c. Теорема доказана. □

5.2. Монотонные способы декодирования

В современной практике и теории информации широко используются алгоритмы, кодирующие информационные массивы, получаемые потоком в режиме онлайн. Процессы кодирования и декодирования также должны происходить в режиме онлайн, а кодовая последовательность также должна строиться потоком.

В связи с этим мы будем рассматривать так называемые монотонные методы кодирования и декодирования. Монотонным методам декодирования будет соответствовать монотонная сложность, которая является модификацией простой колмогоровской сложности.

5.2.1. Монотонная сложность

Функция $f : \{0,1\}^* \to \{0,1\}^*$ называется монотонной, если для всех x и y таких, что $x \subseteq y$, выполнено $f(x) \subseteq f(y)$. Класс таких

функций может быть использован для определения соответствующей монотонной сложности $km(x)$ (см. задачи раздела 5.4).

Для определение монотонных способов декодирования мы будем использовать определение более общего вида.

Заданием вычислимой операции называется перечислимое множество пар конечных последовательностей $\hat{F}$, которое обладает свойствами:

1) для любых пар $(p,x),(p'x') \in \hat{F}$, если $p \subseteq p'$, то $x \subseteq x'$ или $x' \subseteq x$;

2) если $(p,x) \in \hat{F}$, то $(p,x') \in \hat{F}$, то для любых x' таких, что $x' \subseteq x$; также выполнено $(p,\lambda) \in \hat{F}$ для всех $p \in \Xi$.[6]

При таком определении для любой конечной или бесконечной последовательности α, если $(p,x),(p',x') \in \hat{F}$ и $p \subseteq \alpha$, $p' \subseteq \alpha$, то одна из последовательностей x или x' продолжает другую. Это свойство обеспечивает корректность следующего определения вычислимой операции.

Для произвольной конечной или бесконечной последовательности α определим значение вычислимой операции:

$$F(\alpha) = \sup\{x : \exists p(p \subseteq \alpha \& (p,x) \in \hat{F})\}. \tag{5.7}$$

Здесь под sup понимается объединение множества попарно согласованных последовательностей в одну последовательность. Эквивалентная запись для (5.7) в случае бесконечной α:

$$F(\alpha) = \sup_n\{x : (\alpha^n, x) \in \hat{F}\}.$$

Рассмотрим примеры вычислимых операций.

Любая монотонная функция $f : \Xi \to \Xi$ задает вычислимую операцию: множество $\hat{F} = \{(p,x) : \exists p \in D_f(x \subseteq f(p))\}$, где D_F – область определения функции f, удовлетворяет условиям 1) и 2).

[6] Свойство 2) добавлено для технического удобства.

Множество $\hat{F} = \{(p, p') : p' \subseteq p\}$ задает тождественную операцию: $F(p) = p$ для всех p. В частности, $F(\alpha) = \alpha$ для любой бесконечной последовательности α.

Значением вычислимой операции на конечной последовательности может быть бесконечная последовательность. Например, для любой бесконечной вычислимой последовательности α можно рассмотреть задание вычислимой операции $\hat{F} = \{(\lambda, \alpha^n) : n \in \mathcal{N}\}$, где λ – пустая последовательность. Тогда $F(\lambda) = \alpha$. Заметим, что верно и обратное: для любой вычислимой операции F, если $F(p) = \alpha$, где $p \in \Xi$ и $\alpha \in \Omega$, то последовательность α – вычислимая (см. задачу из раздела 5.4).

Для любой вычислимой операции F определим меру монотонной сложности

$$\mathrm{KM}_F(x) = \min\{l(p) : x \subseteq F(p)\}. \tag{5.8}$$

По определению величина $\mathrm{KM}_F(x)$ равна длине самой короткой строки p такой, что для некоторой пары $(p, x') \in \hat{F}$ будет $x \subseteq x'$. По свойству 2) задания вычислимой операции в этом случае также и $(p, x) \in \hat{F}$. Поэтому величина $\mathrm{KM}_F(x)$ равна длине самой короткой строки p такой, что пара $(p, x) \in \hat{F}$, т.е., определение (5.8) эквивалентно

$$\mathrm{KM}_F(x) = \min\{l(p) : (p, x) \in \hat{F}\}.$$

Для монотонных способов декодирования также имеет место теорема инвариантности.

Теорема 5.6. *Существует вычислимая монотонная операция A такая, что для любой вычислимой монотонной операции F существует константа c такая, что*

$$\mathrm{KM}_A(x) \leqslant \mathrm{KM}_F(x) + c$$

для всех x.

Предварительно мы построим равномерно перечислимую последовательность всех эффективных способов задания вычислимых операций.

Лемма 5.6. *Существует перечислимое множество троек $\mathcal{F}$, которое обладает свойствами:*

- *для любого i множество*

$$\hat{F}_i = \{(p, y) : (i, p, y) \in \mathcal{F}\} \tag{5.9}$$

  обладает свойствами 1) *и* 2), *т.е. является заданием некоторой вычислимой операции;*

- *для любого задания вычислимой операции $\hat{F}$ найдется i такое, что $\hat{F} = \hat{F}_i$ (для которого выполнено свойство* (5.9)).

Доказательство. Пусть $U(i, p, y)$ – универсальная функция и множество W_i состоит из всех пар (p, y) таких, что значение $U(i, p, y)$ определено.

Пусть W_i^s состоит из всех пар (p, y) таких, что значение $U(i, p, y)$ определено за s шагов.

Пусть $\tilde{F}_i^0 = \emptyset$.

Полагаем $\tilde{F}_i^s = W_i^s$, если множество W_i^s удовлетворяет условию 1) определения задания эффективной операции. В противном случае, определим $\tilde{F}_i^s = \tilde{F}_i^{s-1}$.

Определим $\tilde{F}_i = \cup_s \tilde{F}_i^s$. После этого для каждого i расширим множество $\tilde{F}_i$ следующим образом: для каждой пары $(p, y) \in \tilde{F}_i$ добавим к нему все пары (p, y'), где $y' \subset y$.

Обозначим полученные множества $\hat{F}_i$, $i = 1, 2, \ldots$ По построению каждое такое множество является заданием вычислимой операции. Пусть

$$\mathcal{F} = \{(i, p, y) : (p, y) \in \hat{F}_i\}.$$

По определению $\mathcal{F}$ – перечислимое множество. Кроме того, для любого задания вычислимой операции $\hat{F}$ найдется i такое, что $\hat{F} = W_i$. По способу определения $\tilde{F}_i = W_i$, причем расширение множества $\tilde{F}_i$ до множества $\hat{F}_i$ не нарушит это равенство, так как множество W_i обладает свойством 2). Значит, $\hat{F} = \hat{F}_i$. Лемма доказана. □

Доказательство теоремы. Пусть перечислимое множество троек $\mathcal{F}$ удовлетворяет условию леммы 5.6.

Определим оптимальный монотонный способ задания $\tilde{A}$ следующим образом:

$$\tilde{A} = \{(\overline{\mathrm{str}(i)}01p, y) : (i, p, y) \in \mathcal{F}\}.$$

Условие 1) монотонного способа задания очевидным образом выполнено. Для того чтобы было выполнено условие 2), расширим множество $\tilde{A}$: для каждой пары $(x, y) \in \tilde{A}$ добавим к множеству $\tilde{A}$ все пары (x, y') такие, что $y' \subseteq y$. Пусть A – соответствующая монотонная операция. Докажем, что A определяет оптимальный монотонный способ декодирования.

Пусть $\hat{F}$ – задание некоторой вычислимой операции, также $(p, x) \in \hat{F}$ и $l(p) = \mathrm{KM}_F(x)$.

Тогда по лемме 5.6 $\hat{F} = \hat{F}_i$, $(i, p, x) \in \mathcal{F}$ и поэтому $(\overline{\mathrm{str}(i)}01p, x) \in A$ для некоторого i. Таким образом,

$$\mathrm{KM}_A(x) \leqslant l(p) + 2l(i) + O(1) = \mathrm{KM}_F(x) + 2l(i) + O(1).$$

Теорема доказана. □

Фиксируем одну из оптимальных вычислимых операций A, удовлетворяющих теореме 5.6, и назовем меру сложности $\mathrm{KM}_A(x)$ *монотонной сложностью* слова x. Нижний индекс в дальнейшем опускаем.

Имеет место неравенство $\mathrm{KM}(x) \leqslant l(x) + O(1)$. Действительно, множество пар $\hat{F} = \{(p, p') : p' \subseteq p, p \in \Xi\}$ удовлетворяет условиям 1) и 2) определения задания вычислимой операции.

Монотонная сложность связана со сложностями других видов следующим образом:

$$\mathrm{KM}(x) \leqslant \mathrm{KP}(x) + O(1) \leqslant \mathrm{K}(x) + 2\log \mathrm{K}(x) + O(1).$$

Действительно, для любой вычислимой префиксно-корректной функции $F(p)$ множество $\{(p, F(p)) : p \in D_F\}$, где D_F – область определения функции F, удовлетворяет условию 1) определения задания вычислимой операции.

Монотонная сложность является монотонной функцией, а именно, $\mathrm{KM}(x) \leqslant \mathrm{KM}(y)$ при $x \subseteq y$.

5.2.2. Теорема Левина–Шнорра

Теорема Левина–Шнорра о характеризации случайной по Мартин-Лефу последовательности завершает программу Колмогорова по сложностному описанию случайных последовательностей.

Напомним, что $\omega^n = \omega_1 \dots \omega_n$ обозначает последовательность первых n битов бесконечной последовательности ω.

Теорема 5.7. *Бесконечная двоичная последовательность ω случайна по Мартин-Лефу тогда и только тогда, когда*

$$\mathrm{KM}(\omega^n) = n + O(1).$$

Доказательство. Докажем, что множество

$$\begin{aligned} &\{\omega \in \Omega : \mathrm{KM}(\omega^n) \not\geqslant n + O(1)\} = \\ &\{\omega \in \Omega : \forall m \exists n (\mathrm{KM}(\omega^n) < n - m)\} \end{aligned} \tag{5.10}$$

является эффективно нулевым.

Предварительно докажем лемму.

Лемма 5.7. *Для любой последовательности попарно несогласованных слов $x_1, x_2, \dots$ выполнено $\sum\limits_{n=1}^{\infty} 2^{-\mathrm{KM}(x_n)} \leqslant 1$.*

Доказательство. Пусть $x_i \subseteq A(p_i)$ и $l(p_i) = \mathrm{KM}(x_i)$ для всех i. Тогда из монотонности сложности слова $p_1, p_2, \dots$ также попарно не согласованы и поэтому $\sum_i 2^{-l(p_i)} \leqslant 1$. □

Переходим к доказательству теоремы. Определим

$$U_{m,n} = \cup\{\Gamma_x : l(x) \leqslant n \& l(x) - \mathrm{KM}(x) > m\}$$

и $U_m = \cup_n U_{m,n}$. По определению U_m – эффективно открытое множество, более того, последовательность $\{U_m\}$ равномерно эффективно открытая, т.е., существует вычислимая основа, задающая множества из этой последовательности.

Докажем второе свойство теста Мартин-Лефа. Представим множество $U_{m,n}$ в виде объединения конечного числа попарно

непересекающихся интервалов: $U_{m,n} = \cup_{j=1}^{k} \Gamma_{x_j}$, где x_j – попарно несогласованы и $l(x_j) - \mathrm{KM}(x_j) > m$ и $l(x_j) \leqslant n$ для всех j.[7] Тогда по лемме 5.7

$$L(U_{m,n}) = \sum_{j=1}^{k} 2^{-l(x_j)} < 2^{-m} \sum_{j=1}^{k} 2^{-\mathrm{KM}(x_j)} < 2^{-m}$$

для всех m и n. По определению

$$U_m = \cup_n U_{m,n} = \cup\{\Gamma_x : l(x) - \mathrm{KM}(x) > m\}.$$

$U_{m,n} \subseteq U_{m,n+1}$ для всех m и n. Поэтому $L(U_m) \leqslant 2^{-m}$ для всех m.

Если ω принадлежит множеству (5.10), то для каждого m существует n такое, что $\mathrm{KM}(\omega^n) > n - m$, т.е. $\omega \in U_{m,n} \subseteq U_m$. Значит, ω является элементом эффективно нулевого множества $\cap_m U_m$.

Докажем обратное утверждение. Пусть $\{U_m\}$ – произвольный тест Мартин-Лефа. Мы докажем более сильное утверждение: если ω отвергается тестом $\{U_m\}$, то величина $n - \mathrm{KP}(\omega^n)$ неограничена. [8]

Пусть $U = \cap_m U_m$, где $L(U_m) \leqslant 2^{-m}$ для всех m и

$$U_m = \cup\{\Gamma_x : (m, x) \in T\},$$

где T – вычислимая основа теста. Пусть также выполнено условие: если $(m, x), (m, x') \in T$, то слова x и x' несравнимы. Доказательство существования основы с таким свойством предоставляется читателю в качестве задачи из раздела 5.4.

Определим семейство равномерно перечислимых снизу полумер на множестве всех натуральных чисел:

$$P_m(x) = \begin{cases} 2^m 2^{-l(x)}, \text{ если } (m, x) \in T, \\ 0 \text{ в противном случае.} \end{cases}$$

[7] Это представление может быть неэффективным.

[8] Напомним, что $n - \mathrm{KM}(\omega^n) \geqslant n - \mathrm{KP}(\omega^n) - O(1)$.

Для каждой из этих полумер выполнено неравенство

$$\sum_x P_m(x) = 2^m \sum_{(m,x)\in T} 2^{-l(x)} = 2^m L(U_m) \leqslant 1.$$

Рассмотрим смесь этих полумер – перечислимую снизу полумеру:

$$R(x) = \sum_{m=1}^{\infty} \frac{1}{m(m+1)} P_m(x).$$

Из определения $\sum_x R(x) \leqslant 1$. Так как $cP(x) \geqslant R(x)$ для некоторой константы c, выполнены неравенства

$$\mathrm{KP}(x) \leqslant -\log R(x) + O(1) \leqslant -\log P_m(x) + 2\log m + O(1).$$

Кроме этого, при $(m, x) \in T$

$$-\log P_m(x) = l(x) - m.$$

Если $\omega \in \cap U_m$, то для каждого m существует n такое, что выполнено $(m, \omega^n) \in T$ и, значит,

$$n - \mathrm{KP}(\omega^n) \geqslant m - 2\log m - O(1).$$

Следовательно,

$$\sup_n (n - \mathrm{KP}(\omega^n)) = \infty.$$

Теорема доказана. □

На самом деле, теорема 5.7 дает характеризацию случайных последовательностей не только в терминах монотонной сложности, но и в терминах префиксной сложности. Первая часть доказательства совпадает с доказательством необходимости в теореме 5.7. Вторая часть этой теоремы и была доказана для префиксной сложности.

Ввиду важности этого результата, сформулируем его в виде теоремы.

Теорема 5.8. *Бесконечная двоичная последовательность ω случайна по Мартин-Лефу тогда и только тогда, когда*

$$\mathrm{KP}(\omega^n) \geqslant n - O(1).$$

Обратное неравенство не имеет место для префиксной сложности (см. задачу из раздела 5.4).

5.3. Вычислимые меры

До этого места мы для простоты рассматривали только равномерную меру. На самом деле все основные результаты алгоритмической теории вероятности выполнены и для произвольных вычислимых мер.

Пусть P – произвольная мера на множестве Ω всех бесконечных двоичных последовательностей. Рассмотрим функцию, принимающую вещественные значения:

$$P(x) = P(\Gamma_x),$$

где x – произвольная конечная последовательность. Функция $P(x)$ удовлетворяет условиям:

$$\begin{aligned} P(\lambda) &= 1; \\ P(x) &= P(x0) + P(x1). \end{aligned} \tag{5.11}$$

Верно и обратное. Для задания меры P на всех борелевских подмножествах Ω достаточно задать для всех конечных последовательностей x значения функции $P(x)$, удовлетворяющие условиям (5.11). После этого определяется мера интервалов в виде $P(\Gamma_x) = P(x)$ для всех x, которая может быть стандартным образом распространена на любое объединение счетной последовательности попарно непересекающихся интервалов – любое открытое множество, а тем самым и на любое замкнутое множество. Далее можно распространить меру P на любое борелевское подмножество Ω.

Простейший пример неравномерной меры – произвольная бернуллиевская мера с вероятностью единицы, равной p:

$$B_p(x) = p^k(1-p)^{n-k},$$

где n – длина последовательности x, а k – число единиц в ней.

Если $p = 0$, то мера P сосредоточена только на одной последовательности 0^∞: $P(0^n) = 1$ для всех n и $P(x) = 0$ для всех остальных x.

Мера P называется вычислимой, если функция $P(x)$ является вычислимой. Определение вычислимой функции с вещественными значениями было дано в разделе 5.1.3.

Известно, что функция вычислима тогда и только тогда, когда она перечислима снизу и сверху. Если функция является мерой, то достаточно требовать только перечислимость снизу или перечислимость сверху.

Например, если функция $P(x)$ перечислима снизу, то она перечислима и сверху (а тем самым и вычислима), так как

$$r > P(x) \Leftrightarrow 1 - r < \sum_{z \neq x, l(z)=l(x)} P(z).$$

Бернуллиевская $B_p(x)$ мера вычислима тогда и только тогда, когда вещественное число p вычислимо.

Аналогичным образом определяется понятие последовательности случайной по Мартин-Лефу относительно вычислимой меры P. Определяется понятие P-теста: это перечислимая последовательность эффективно открытых множеств $\{U_m\}, m = 1, 2, \ldots$ такая, что $P(U_m) \leqslant 2^{-m}$.

Множество A – эффективно P-нулевое, если $A \subseteq \cap_m U_m$ для некоторого P-теста $\{U_m\}, m = 1, 2, \ldots$.

Точно так же, как теорема 4.1, доказывается следующая теорема. Надо только в доказательстве заменить меру L на P, а значение $2^{-l(x)}$ на $P(x)$.

Теорема 5.9. *Для любой вычислимой меры P существует максимальное по включению эффективное P-нулевое множество.*

Теорема Левина–Шнорра также имеет место для произвольной вычислимой меры.

Теорема 5.10. *Бесконечная двоичная последовательность ω случайна по Мартин-Лефу относительно вычислимой меры P тогда и только тогда, когда*

$$\mathrm{KM}(\omega^n) = -\log P(\omega^n) + O(1).$$

Утверждение о том, что последовательность ω случайна по Мартин-Лефу относительно вычислимой меры P тогда и только тогда, когда $\mathrm{KM}(\omega^n) \geqslant -\log P(\omega^n) + O(1)$, доказывается точно так же, как и теорема 5.7.

Тривиальная часть доказательства теоремы 5.7 о том, что $\mathrm{KM}(x) \leqslant l(x) + O(1)$, становится нетривиальной в случае произвольной вычислимой меры P. Мы сформулируем и докажем аналогичное утверждение в виде леммы.

Лемма 5.8. *Для любой вычислимой меры P найдется константа c такая, что $\mathrm{KM}(x) \leqslant -\log P(x) + c$ для всех x.* [9]

Доказательство. Построим иерархическую систему отрезков с вычислимыми концами $\pi_x \subseteq [0,1]$ такую, что

- длина π_x равна $P(x)$;
- $\pi_\lambda = [0,1]$;
- $\pi_x = \pi_{x0} \cup \pi_{x1}$ для всех x.

Такая система отрезков существует, и концы этих отрезков можно вычислять с любой степенью точности.

Рассмотрим также отрезки с двоично-рациональными концами. Произвольной двоичной последовательности $x = x_1 \ldots x_n$

[9] Впервые формулировка этой леммы опубликована в [15]. Наиболее точное изложение доказательства имеется в [27]. В конструкции этой леммы содержится идея метода арифметического кодирования, см. [10], [34].

сопоставим отрезок с двоично-рациональными концами длиной 2^{-n}:

$$I_x = \left[\sum_{i=1}^{n} x_i 2^{-i}, \sum_{i=1}^{n} x_i 2^{-i} + 2^{-n} \right].$$

Рассмотрим множество

$$\hat{F} = \{(x, y) : I_x \subset \pi_y\}. \tag{5.12}$$

Множество $\hat{F}$ является перечислимым. Так как концы отрезков I_x являются конечными объектами – рациональными числами, а концы отрезков π_y можно вычислять с любой степенью точности, можно определить алгоритм, который определит строгое включение $I_x \subset \pi_y$, если оно имеет место, и никогда не остановится в противном случае.

Кроме этого, из определения выполнены следующие свойства:

1) $(x, \lambda) \in \hat{F}$ для всех x;

2) если $(x, y) \in \hat{F}$ и $x \subseteq x'$, то $(x', y) \in \hat{F}$;

3) если $(x, y), (x, y') \in \hat{F}$, то $\pi_y \cap \pi_{y'} \neq \emptyset$ и, следовательно, $y \subseteq y'$ или $y' \subseteq y$.

Отсюда можно вывести второе условие из определения задания эффективной операции: если $(x, y), (x', y') \in \hat{F}$ и $x \subseteq x'$, то по свойству 1) имеет место $(x', y) \in \hat{F}$ и по свойству 2) последовательности y и y' сравнимы.

Третье условие условие определения задания эффективной операции – если $(x, y) \in \hat{F}$, то $(x, y') \in \hat{F}$ для всех $y' \subseteq y$ – прямо следует из определения системы отрезков π_x.

Таким образом, задание $\hat{F}$ определяет вычислимую операцию F и соответствующую меру сложности:

$$\mathrm{KM}_F(x) = \min\{l(p) : (p, x) \in \hat{F}\} = \min\{l(p) : I_p \subset \pi_x\}.$$

Отсюда следует, что $\mathrm{KM}_F(x)$ равно минус логарифму от длины самого большого двоичного отрезка I_p, строго содержащегося в π_x.

В любом отрезке π_x строго содержится отрезок с двоичными концами длины не менее $\frac{1}{4}|\pi_x|$. Отсюда для самого большого двоичного отрезка I_p, строго содержащегося в π_x, выполнено $\frac{1}{4}P(x) \leqslant 2^{-l(p)}$ или

$$\mathrm{KM}_F(x) = l(p) \leqslant -\log P(x) + 2.$$

Отсюда $\mathrm{KM}(x) \leqslant -\log P(x) + O(1)$. $\square$

Определим дефект случайности конечной последовательности x относительно вычислимой меры P

$$d_P(x) = -\log P(x) - \mathrm{KM}(x). \tag{5.13}$$

Для любого натурального числа m выполнено неравенство $P(\cup\{\Gamma_x : d_P(x) > m\}) = P(\cup\{\Gamma_x : P(x) < 2^{-\mathrm{KM}(x)-m}\}) \leqslant 2^{-m}$. Доказательство этого неравенства аналогично доказательству необходимости из теоремы 5.7.

Теорема 5.10 влечет

Следствие 5.4. *Бесконечная последовательность ω является случайной в смысле Мартин-Лефа относительно вычислимой меры P тогда и только тогда, когда $\sup_n d_P(\omega^n) < \infty$.*

Можно определить понятие алгоритмической случайности непосредственно в терминах тестов типа (5.13). Пусть Q – вычислимая мера. Функция $F : \Xi \to \mathcal{R}_+$ называется тестом случайности относительно меры Q, если выполнены свойства:

- (i) Функция F перечислимая снизу;
- (ii) $Q(\cup\{\Gamma_x : F(x) > m\}) \leqslant 2^{-m}$ для всех m.

Для бесконечной последовательности ω определим $F(\omega) = \sup_n F(\omega^n)$. Из определения видно, что семейство эффективно открытых множеств $U_m = \cup\{\Gamma_x : F(x) > m\}$ образует тест Мартин-Лефа, корректный относительно меры Q.[10] С другой

[10] который ранее был определен в разделе 4.

стороны, по любому тесту Мартин-Лефа $\{U_m\}$ можно определить функцию $F(x) = \max_m(\Gamma_x \subseteq U_m)$. Легко видеть, что эта функция удовлетворяет условиям (i) и (ii).

В заключение заметим, что сложностные определения случайности и определение случайности по Мартин-Лефу относительно вычислимой меры P приводят к одному и тому же классу бесконечных последовательностей. Последовательности из этого класса будем называть просто *случайными последовательностями* относительно вычислимой меры P.

5.4. Задачи и упражнения

1. Доказать, что $\mathrm{KP}(x) \leqslant \mathrm{K}(x)+\log \mathrm{K}(x)+2\log\log \mathrm{K}(x)+O(1)$. Доказать, что слагаемое $\log \mathrm{K}(x)$ нельзя устранить из правой части неравенства.

2. Доказать, что $\mathrm{KP}(x,y) \leqslant \mathrm{KP}(x) + \mathrm{K}(y) + O(1)$, где $\mathrm{K}(y)$ – простая колмогоровская сложность.

3. Доказать, что $\mathrm{KP}(\phi(x,y)) \leqslant \mathrm{KP}(x) + \mathrm{KP}(y) + O(1)$, где $\phi(x,y)$ – произвольная вычислимая функция.

4. Доказать, что $\mathrm{KP}(x,y) \leqslant \mathrm{KP}(x) + \mathrm{KP}(y|x) + O(1)$.

5. Доказать, что $\mathrm{KP}(x,y,z) \leqslant \mathrm{KP}(x|y) + \mathrm{KP}(y|z) + \mathrm{KP}(z) + O(1)$.

6. Доказать, что $\mathrm{KP}(x)+\mathrm{KP}(y|x)-\log \mathrm{KP}(x)-2\log\log \mathrm{KP}(x)-O(1) \leqslant \mathrm{KP}(x,y) \leqslant\leqslant \mathrm{KP}(x) + \mathrm{KP}(y|x) + O(1)$.

7. Доказать, что $\mathrm{KM}(x) \leqslant \mathrm{KP}(x) + O(1) \leqslant \mathrm{K}(x) + 2\log \mathrm{K}(x) + O(1)$.

8. Доказать неравенство (5.1), используя теорему 5.3 о двойственности.

9. Доказать неравенство (5.4).

10. Доказать, что универсальная перечислимая снизу полумера на натуральных числах $P(n)$ не является вычислимой функцией и $\sum\limits_n P(n) < 1$.

11. Доказать, что двоичное разложение действительного числа $\sum\limits_n P(n)$ является случайной по Мартин-Лефу последователь-

ностью.

12. Доказать, что такое же утверждение верно для вещественного числа $\sum_n 2^{-\mathrm{KP}(n)}$.

13. Пусть F – оптимальная префиксно-корректная функция. Для произвольной строки x можно рассмотреть вероятность того, что машина F на случайном входе выдаст x

$$\mathcal{P}(x) = L\{\exists p(F(p) = x)\} = \sum_{F(p)=x} 2^{-l(p)}.$$

a) Доказать, что

$$2^{-\mathrm{KP}(x)} \leqslant \mathcal{P}(x) \leqslant 2^{-\mathrm{KP}(x)+1},$$

таким образом, величина $\mathcal{P}(x)$ и априорная полумера $P(x)$ совпадают с точностью до мультипликативной константы.

b) Доказать, что число

$$\sum_x \mathcal{P}(x) = \sum_{F(p) \text{ определено}} 2^{-l(p)}$$

является невычислимым и, более того, случайным относительно равномерной меры. Это число называется числом Чейтина.

14. Доказать, что для почти любой бесконечной последовательности ω найдется число m такое, что $\mathrm{KP}(\omega^n) \geqslant n + \mathrm{KP}(n) - m$ для бесконечно многих n.

15. Существует другое определение префиксно-корректного способа декодирования. Вычислимая функция $B(p, y)$ называется префиксно-корректной, если для любых пар (p, y) и (p', y) из ее области определения таких, что $p \subseteq p'$, выполнено $B(p, y) = B(p', y)$. Таким образом, коды могут продолжать друг друга, но тогда они являются кодами одного и того же конечного объекта. На основе каждого такого способа декодирования $B(p, y)$ определяется мера сложности

$$\mathrm{KP}'_B(x|y) = \min\{l(p) : B(p, y) = x\}.$$

Доказать, что существует соответствующая префиксная сложность $\mathrm{KP}'(x|y)$. Доказать, что она совпадает с ранее определенной префиксной сложностью с точностью до константы:

$$\mathrm{KP}'(x|y) = \mathrm{KP}(x|y) + O(1).$$

16. Доказать, что префиксная и монотонная сложности не являются вычислимыми функциями. Доказать, что для оптимальных способов декодирования для этих сложностей не существует соответствующих вычислимых способов кодирования.

17. Доказать, что $\mathrm{KM}(x) \leqslant \mathrm{KM}(y)$ при $x \subseteq y$.

18. Доказать, что $\mathrm{K}(x) \not\leqslant \mathrm{KM}(x) + O(1)$. Привести пример, где нарушается соответствующее неравенство.

19. Доказать, что для любого теста случайности существует основа, для которой выполнено условие: если $(m, x), (m, x') \in T$, то слова x и x' несравнимы.

20. Даны вычислимая мера Q и бесконечная последовательность ω. Доказать, что если $Q(\omega^n) = 0$ для всех n, то последовательность ω не случайная по мере Q. Построить тест Мартин-Лефа, отвергающий ω.

21. Привести пример, когда для некоторой вычислимой операции $F(x) = \omega$, где x – конечная последовательность, а ω – бесконечная. Что можно сказать о последовательности ω.

22. Доказать, что вычислимая последовательность случайна по некоторой вычислимой мере.

23. Бернуллиевская мера B_p является вычислимой, если вычислимо число p. Доказать, что верно и обратное утверждение.

24. Пусть R_p – множество всех бесконечных последовательностей, случайных относительно меры B_p. Доказать, что $R_p \cap R_q = \emptyset$, если $p \neq q$.

25. Пусть P – произвольная вычислимая мера. Доказать, что существует неслучайная по этой мере бесконечная последовательность. Доказать, что множество всех неслучайных последовательностей бесконечно (и имеет мощность континуума).

26. Доказать, что $\sup_n \mathrm{KM}(\omega^n) < \infty$ тогда и только тогда, когда бесконечная последовательность ω является вычислимой.

27. Функция $f : \{0,1\}^* \to \{0,1\}^*$ называется монотонной, если $f(x) \subseteq f(y)$ при $x \subseteq y$ для всех x и y из области определения функции f. Для любой вычислимой монотонной функции f можно также определить монотонную меру сложности:

$$km_f(x) = \min\{l(p) : x = f(p)\}.$$

a) Доказать, что для таких мер сложности также имеет место теорема инвариантности: существует такая вычислимая монотонная функция f, что для любой вычислимой монотонной функции g существует константа c такая, что неравенство

$$km_f(x) \leqslant km_g(x) + c$$

выполнено для всех x. Фиксируем одну из таких функций f и рассмотрим монотонную меру сложности $km(x) = km_f(x)$.

b) Доказать, что $km(x) \leqslant l(x) + O(1)$.

c) Доказать, что $\mathrm{K}(x) - O(1) \leqslant km(x) \leqslant \mathrm{KP}(x) + O(1)$.

d) Доказать, что $\mathrm{KM}(x) \leqslant km(x) + O(1)$. Доказать, что обратное неравенство неверно.

e) Доказать, что $\sup_n km(\omega^n) = \infty$ для любой бесконечной последовательности ω.

f) Доказать, что теорема Левина–Шнорра для равномерной меры верна и для сложности $km(x)$. Будет ли эта теорема верна в случае произвольной вычислимой меры? Для каких мер она может быть верна (и для каких мер она может быть неверна)?

28. Доказать, что $\mathrm{KP}(\omega^n) \leqslant n + \mathrm{KP}(n) + O(1)$.

29. Доказать, что $\mathrm{KP}(\omega^n) = \mathrm{KP}(n) + O(1)$ для любой вычислимой последовательности ω.

30. Проверить утверждение из доказательства леммы 5.8: в любом отрезке π_x строго содержится отрезок с двоичными концами длины не менее $\frac{1}{4}|\pi_x|$.

31. Докажите, что среди всех тестов $F(x)$, удовлетворяющих условиям (i) и (ii), корректных относительно вычислимой меры Q, существует максимальный с точностью до мультипликативной константы, т.е. существует тест $U(x)$, удовлетворяю-

щий условиям (i) и (ii), такой, что для любого теста F, корректного относительно меры Q, найдется константа c такая, что $cU(x) \geqslant F(x)$ для всех x. Придумайте доказательство, независимое от использования универсального теста из раздела 4.

Глава 6

Универсальное прогнозирование

6.1. Универсальная полумера на дереве всех двоичных последовательностей

В этом разделе мы определим априорную перечислимую снизу полумеру M на множестве всех двоичных последовательностей. Эта полумера определяется аналогично априорной полумере P на множестве всех натуральных чисел. Поскольку мы отождествили натуральные числа и конечные двоичные последовательности, формально носитель у полумер P и M один и тот же. Различие определений заключается в том, что при определении полумеры P мы не учитывали структуру носителя (и поэтому называли его множеством натуральных чисел). Определение полумеры M будет существенно использовать структуру множества всех конечных двоичных последовательностей, а именно эта полумера будет согласована с отношением продолжения: $x \subseteq y$.

Напомним,что $\Xi = \{0,1\}^*$ – множество всех конечных двоичных последовательностей (слов или строк), $\Omega = \{0,1\}^\infty$ – множество всех бесконечных двоичных последовательностей. Конечные двоичные последовательности с отношением продолжения

$x \subseteq y$ образуют бесконечное дерево, вершиной которого является пустая последовательность λ.

Полумерой (на дереве двоичных последовательностей) называется всюду определенная функция $Q : \Xi \to \mathcal{R}_+$, удовлетворяющая свойствам

1) $Q(\lambda) \leqslant 1$;

2) $Q(x) \geqslant Q(x0) + Q(x1)$ для всех x.

Полумера Q называется перечислимой снизу, если множество $\{(r, x) : r < Q(x)\}$ перечислимо, где r обозначает рациональное число.

Среди всех перечислимых снизу полумер также существует универсальный объект. Мы докажем, что существует максимальная с точностью до мультипликативной константы перечислимая снизу полумера.

Теорема 6.1. *Существует перечислимая снизу полумера M такая, что для любой перечислимой снизу полумеры Q существует константа c, для которой*

$$cM(x) \geqslant Q(x)$$

для всех x.

Доказательство. Доказательство аналогично доказательству теоремы 5.2. Предварительно покажем, что все перечислимые снизу полумеры можно перечислять снизу равномерно одним алгоритмом.

Лемма 6.1. *Существует такая последовательность полумер Q_i, что*

- *множество $\{(i, r, x) : r < Q_i(x)\}$ перечислимо (r – рациональное);*
- *для любой перечислимой снизу полумеры Q найдется такое i, что $Q = Q_i$.*

Доказательство. Для построения такой последовательности полумер используем универсальную функцию. Рассмотрим перечислимые множества, состоящие из пар (r, x), где r – рациональное число, x – двоичная последовательность.

Пусть $U(i, r, x)$ – универсальная функция. Каждое перечислимое множество пар (r, x), есть область определения функции $U(i, r, x)$ при фиксированном i. Такое множество обозначается W_i (см. раздел 3.1).

Запустим процесс вычисления всех значений универсальной функции $U(i, r, x)$ в виде цикла, на каждой итерации которого просматривается только одна тройка (i, r, x) и моделируется один шаг машины Тьюринга, вычисляющей значение $U(i, r, x)$. При этом каждая тройка (i, r, x) просматривается на бесконечном числе итераций цикла. Если на очередной итерации s цикла значение $U(i, r, x)$ впервые определено, то определим множество $W_i^s = W_i^{s-1} \cup \{(r, x)\}$; полагаем $W_i^s = W_i^{s-1}$, в противном случае. Пусть $W_i^0 = \emptyset$.

Таким образом, W_i^s обозначает конечное подмножество пар, перечисленных в W_i за s шагов цикла. Определим

$$R_i^s(x) = \max(\{r : (r, x) \in W_i^s\} \cup \{0\}).$$

Функция $R_i^s(x) > 0$ для не более чем конечного числа x.

Пусть для любых i и s число $t(i, s)$ равно максимальному t такому, что $t \leqslant s$ и функция $R_i^t(x)$ удовлетворяет свойствам 1) и 2) из определения полумеры. Определим

$$Q_i(x) = \sup_s R_i^{t(i,s)}(x).$$

Легко видеть, что Q_i является равномерно перечислимой снизу последовательностью полумер: множество $\{(i, r, x) : r < Q_i(x)\}$ перечислимо.

Для любой перечислимой снизу полумеры Q имеем

$$W_i = \{(r, x) : r < Q(x)\}$$

для некоторого i. Легко видеть, что $Q(x) = Q_i(x)$ для всех x. □

Доказательство теоремы. Определим

$$M(x) = \sum_{i=1}^{\infty} \frac{1}{i(i+1)} Q_i(x).$$

Функция M перечислима снизу. Нетрудно проверить, что она, как смесь полумер, также является полумерой.

Кроме этого, для любой перечислимой снизу полумеры Q будет $Q = Q_i$ для некоторого i. Поэтому $i(i+1)M(x) \geqslant Q(x)$ для всех x. $\square$

Фиксируем одну из функций M, удовлетворяющих заключению теоремы 6.1, назовем ее *универсальной* (*априорной*) *полумерой* на дереве всех двоичных последовательностей.

В частности, $cM(x) \geqslant Q(x)$ для любой вычислимой меры Q, где c – константа. Например, $cM(x) \geqslant 2^{-l(x)}$, откуда

$$\mathrm{KA}(x) \leqslant l(x) + O(1),$$

где $\mathrm{KA}(x) = -\log M(x)$.

Функция $\mathrm{KA}(x) = -\log M(x)$ играет роль алгоритмической сложности. Доказано, что она не совпадает с монотонной сложностью $\mathrm{KM}(x)$, хотя и близка к ней.

Предложение 6.1. *Имеют место следующие соотношения:*

$$\mathrm{KA}(x) \leqslant \mathrm{KM}(x) + O(1);$$
$$\mathrm{KM}(x) \leqslant \mathrm{KA}(x) + 2\log l(x) + O(1).$$

Доказательство. Пусть $\mathrm{KM}(x) = \mathrm{KM}_A(x)$, где A – вычислимая операция. Определим перечислимую снизу функцию

$$Q(x) = L(\cup_p \{\Gamma_p : x \subseteq A(p)\}). \tag{6.1}$$

Нетрудно доказать, что функция Q является полумерой. Действительно, $Q(\lambda) \leqslant 1$. Кроме этого, если имеет место $x0 \subseteq A(p)$ или $x1 \subseteq A(p)$, то $x \subseteq A(p)$. Поэтому для такого p интервал Γ_p попадет в объединение всех Γ_z таких, что $x \subseteq A(z)$. Поэтому $Q(x) \geqslant Q(x0) + Q(x1)$.

Среди интервалов y из объединения (6.1) имеется и тот, для которого $l(y) = \mathrm{KM}(x)$. Поэтому $Q(x) \geqslant 2^{-\mathrm{KM}(x)}$, а значит, и $\mathrm{KA}(x) \leqslant \mathrm{KM}(x) + O(1)$.

Доказательство второго утверждения оставляем читателю в качестве упражнения. $\square$

В терминах универсальной полумеры на дереве двоичных последовательностей можно сформулировать критерий случайности по Мартин-Лефу.

Как уже было отмечено, для любой вычислимой меры Q найдется константа $c > 0$ такая, что $cM(x) \geqslant Q(x)$ для всех x. Иными словами, $M(x)/Q(x) \geqslant 1/c > 0$ для всех x. Оказывается, обратное неравенство выполнено для всех начальных фрагментов случайных последовательностей.

Теорема 6.2. *Бесконечная последовательность ω случайна по вычислимой мере Q тогда и только тогда, когда*

$$\sup_n \frac{M(\omega^n)}{Q(\omega^n)} < \infty.$$

Доказательство. Пусть $\{U_m : m = 1, 2, \ldots\}$ – тест случайности по Мартин-Лефу, корректный относительно вычислимой меры Q:

$Q(U_m) \leqslant 2^{-m}$ и $U_{m+1} \subseteq U_m$ для всех m.

Определим последовательность функций:

$$Q'_m(x) = 2^m Q(\Gamma_x \cap U_m) = \begin{cases} 2^m Q(x), \text{ если } \Gamma_x \subseteq U_m, \\ 0 \text{ в противном случае.} \end{cases}$$

Каждая функция Q'_m удовлетворяет свойству 2) определения полумеры на всех продолжениях последовательностей из U_m. Мы перенесем это свойство на все конечные последовательности, если определим для всех x

$$Q_m(x) = \sup_n \sum_{x \subseteq z, l(z)=n} Q'_m(z).$$

Имеем $Q_m(x) \geqslant Q'_m(x)$ для всех x, причем на всех продолжениях последовательностей из U_m имеет место равенство.

Имеем $Q_m(\lambda) \leqslant 2^m Q(U_m) \leqslant 1$ для всех m. Каждая функция Q_m является полумерой.

Из определения следует, что функции Q_m равномерно по m перечислимы снизу. Определим смесь полумер:

$$R(x) = \sum_{m=1}^{\infty} \frac{1}{m(m+1)} Q_m(x).$$

Полумера $R(x)$ перечислима снизу, поэтому найдется константа c такая, что $cM(x) \geqslant R(x)$ для всех x.

Допустим, что последовательность $\omega \in \cap_m U_m$. Тогда для каждого m существует n такое, что

$$\Gamma_{\omega^n} \subseteq U_m,$$

поэтому $Q'_m(\omega^n) = 2^m Q(\omega^n)$. Отсюда получаем

$$\frac{M(\omega^n)}{Q(\omega^n)} \geqslant \frac{R(\omega^n)}{cQ(\omega^n)} \geqslant \frac{Q'_m(\omega^n)}{cm(m+1)Q(\omega^n)} = \\ = \frac{2^m Q(\omega^n)}{cm(m+1)Q(\omega^n)} = \frac{2^m}{cm(m+1)}.$$

Следовательно,

$$\sup_n \frac{M(\omega^n)}{Q(\omega^n)} = \infty.$$

Докажем обратное утверждение. Допустим, что

$$\sup_n \frac{M(\omega^n)}{Q(\omega^n)} = \infty. \tag{6.2}$$

Определим тест Мартин-Лефа:

$$U_m = \cup \left\{ \Gamma_x : \frac{Q(x)}{M(x)} < 2^{-m} \right\}.$$

Представим

$$U_m = \cup_i \{\Gamma_{x_i} : Q(x_i) < 2^{-m} M(x_i)\},$$

где все последовательности x_i попарно несравнимы. Тогда

$$Q(U_m) = \sum_i Q(x_i) < 2^{-m} \sum_i M(x_i) < 2^{-m} M(\lambda) < 2^{-m}$$

для всех m.

Если последовательность ω удовлетворяет условию (6.2), то для каждого m существует n такое, что $Q(\omega^n) < 2^{-m} M(\omega^n)$. Тогда $\Gamma_{\omega^n} \subseteq U_m$ и, значит, $\omega \in U_m$. Следовательно, $\omega \in \cap_m U_m$, т.е. ω не является случайной по мере Q.

Теорема доказана. □

Назовем функцию

$$d_Q(x) = -\log Q(x) - \mathrm{KA}(x) = \log \frac{M(x)}{Q(x)} \tag{6.3}$$

дефектом случайности конечной последовательности x относительно вычислимой меры Q.[1]

По теореме 6.2 Бесконечная последовательность ω случайна по вычислимой мере Q тогда и только тогда, когда

$$\sup_n d_Q(\omega^n) < \infty.$$

Вероятностные машины. Под вероятностной машиной понимаем пару (Q, F), где Q – вычислимая мера, а F – вычислимая операция. В дальнейшем мы считаем, что $Q = L$ – равномерная мера. С каждой вероятностной машиной связываем функцию

$$P(x) = L\{\omega : x \subseteq F(\omega)\}. \tag{6.4}$$

Легко видеть, что функция $P(x)$ является перечислимой снизу полумерой.

[1] См. также аналогичное определение (5.13).

Верно и обратное: для любой перечислимой снизу полумеры P можно построить такую вычислимую операцию F, что для нее выполнено (6.4).

Покажем, что любая перечислимая снизу мера является прообразом равномерной меры относительно некоторого вычислимого оператора.

Предложение 6.2. *Для любой перечислимой снизу полумеры P существует вычислимая операция F такая, что*

$$P(x) = L\{\omega \in \Omega : x \subseteq F(\omega)\}$$

для всех x.

Доказательство. Доказательство аналогично доказательству предложения 12.2 с несколько модернизированной системой отрезков π_x. Построим систему отрезков π_x, такую что

- длина π_x равна $P(x)$;
- $\pi_\lambda = [0, 1]$;
- $\pi_{x0} \cup \pi_{x1} \subseteq \pi_x$ для всех x.

В отличие от системы отрезков из доказательства леммы 5.8, концы этих отрезков будут полувычислимыми вещественными числами: левые концы перечислимы сверху, а правые – перечислимы снизу. В этом случае отношение $I_z \subset \pi_x$ по прежнему перечислимо.

Используем оператор F, построенный в доказательстве леммы 5.8:

$$\hat{F} = \{(z, y) : I_z \subset \pi_y\}.$$

По определению

$$F(\alpha) = \sup\{y : (z, y) \in \hat{F} \& z \subseteq \alpha\}$$

для любой конечной или бесконечной последовательности α. По определению

$$L\{\alpha \in \Omega : x \subseteq F(\alpha)\} = L\left(\cup\{\Gamma_z : (z,x) \in \hat{F}\}\right) = L\left(\cup\{\Gamma_z : I_z \subset \pi_x\}\right) = |\pi_x| = P(x).$$

Утверждение доказано.[2] □

Вычислительной моделью вероятностной машины является машина Тьюринга с дополнительной лентой, на которой некоторое аналоговое устройство последовательно печатает двоичные биты, составляющие потенциально бесконечную последовательность ω. Например, биты ω могут получаться в результате записи результатов подбрасывания симметричной монеты.

Машина Тьюринга в процессе вычисления последовательно читает ω и также последовательно печатает бит за битом некоторую выходную последовательность. Поскольку на множестве всех ω имеется вероятностное распределение, можно рассматривать вероятность (6.4) события, состоящего в том, что машина напечатает на выходе некоторую последовательность, началом которой является конечная последовательность x.

Можно рассматривать вероятность более сложных событий. Например, для произвольного борелевского подмножества $A \subseteq \Omega$ можно рассмотреть вероятность выдать последовательность из A:

$$P(A) = L\{\omega : F(\omega) \in A\}.$$

Соотношение (6.4) устанавливает взаимно однозначное соответствие между вероятностными машинами и перечислимыми снизу полумерами.

В частности, универсальная полумера M также допускает представление (6.4):

$$M(x) = L\{\omega : x \subseteq F_M(\omega)\}$$

[2]Утверждение впервые сформулировано и доказано в [10]. Приведенное доказательство заимствовано из [27].

для некоторой вычислимой операции F_M. По свойству универсальной полумеры вероятностная машина (L, F_M) выдает конечные последовательности x с самой большой вероятностью: для произвольной вероятностной машины (L, F) найдется такая константа c, что

$$cL\{\omega : x \subseteq F_M(\omega)\} \geqslant L\{\omega : x \subseteq F(\omega)\}$$

для любой вычислимой операции F.

Напомним, что последовательность $\alpha = \alpha_1\alpha_2\ldots$ является вычислимой, если вычислимой является функция $f(i) = \alpha_i$.

В качестве примера использования вероятностной машины ответим на вопрос: может ли вероятностная машина с положительной вероятностью выдать бесконечную невычислимую последовательность?

Теорема Де Леу, Мура, Шеннона, Шапиро [20] утверждает, что это невозможно. Для любой вероятностной машины (L, F) и бесконечной последовательности α, если $L\{\omega : F(\omega) = \alpha\} > 0$, то последовательность α является вычислимой.

Мы сформулируем это утверждение на языке полумер. Ясно, что достаточно формулировать это утверждение только для универсальной полумеры.

Теорема 6.3. *Пусть для некоторой положительной константы $c > 0$ неравенство $M(\alpha^n) > c$ выполнено для всех n. Тогда последовательность α вычислима.*

Доказательство. Пусть $M(\alpha^n) > c$ выполнено для всех n, где $c > 0$ – рациональное. Может существовать не более чем $1/c$ попарно несравнимых конечных последовательностей $x_1, \ldots, x_m$ таких, что $M(x_i) > c$ для всех i. Это следует из неравенства

$$cm < \sum_{i=1}^{m} M(x_i) \leqslant 1.$$

Выберем такой набор $x_1, \ldots, x_m$, число элементов m в котором максимальное. Тогда для любого i не может существовать двух

несравнимых продолжений $x_i \subset y$ и $x_i \subset y'$ таких, что выполнено $M(y) > c$ и $M(y') > c$. Кроме этого, $x_i \subset \alpha$ для одного из этих i. Отсюда следует, что можно вычислять начальные фрагменты α, перечисляя все y такие, что $M(y) > c$ и $x_i \subseteq y$. Все такие y будут попарно сравнимы и будут являться начальными фрагментами α. $\square$

Супермартингалы. Понятие полумеры связано с понятием супермартингала. Мы сформулируем понятие супермартингала, адаптированное для двоичных последовательностей.

Пусть Q – вычислимая мера на пространстве всех бесконечных двоичных последовательностей. Функция $\mathcal{P} : \Xi \to \mathcal{R}_+$ называется Q-супермартингалом, если она удовлетворяет условиям

- $\mathcal{P}(\lambda) \leqslant 1$;
- $\mathcal{P}(x)Q(x) \geqslant \mathcal{P}(x0)Q(x0) + \mathcal{P}(x1)Q(x1)$ для всех x.

Легко видеть, что в том случае, когда $Q(x) > 0$ для всех x, второе условие эквивалентно условию

$$\mathcal{P}(x) \geqslant \mathcal{P}(x0)Q(0|x) + \mathcal{P}(x1)Q(1|x)$$

для всех x, где $Q(i|x) = Q(xi)/Q(x)$ – условная вероятность того, что $i \in \{0, 1\}$ при известном x. Здесь справа написано математическое ожидание супермартингала по условному распределению $Q(\cdot|x)$.

Супермартингал $\mathcal{P}$ является перечислимым снизу, если множество $\{(r, x) : r < \mathcal{P}(x)\}$ перечислимо.

Если заменить неравенства на равенства, получим определение Q-мартингала.

Определения мартингала и супермартингала допускают игровую интерпретацию. Можно рассмотреть игру, которая идет по раундам (шагам) между игроком и казино. Начальный капитал игрока $\mathcal{P}(\lambda)$ не превосходит 1 (или равен 1 в случае мартингала).

Пусть на раунде n казино уже выдало последовательность случайных битов $x = x_1 \ldots x_n$. Текущий капитал игрока равен

$\mathcal{P}(x)$. На шаге $n+1$ игрок ставит весь свой текущий капитал $\mathcal{P}(x)$ на 0 и 1. Он договаривается с казино, что его выигрыш будет равен $\mathcal{P}(x0)$, если источник случайных битов выдаст 0, или равен $\mathcal{P}(x1)$, если источник случайных битов выдаст 1. Источник случайных исходов описывается мерой Q.

Казино ограничивает будущие выигрыши игрока $\mathcal{P}(x0)$ и $\mathcal{P}(x1)$ игрока так, чтобы его средний выигрыш не превосходил вложенный капитал:

$$\mathcal{P}(x0)Q(0|x) + \mathcal{P}(x1)Q(1|x) \leqslant \mathcal{P}(x).$$

Мы говорим, что такая игра является справедливой. Если $\mathcal{P}$ – супермартингал, то наличие неравенства $\leqslant$ можно интерпретировать как то, что часть выигрыша казино может забирать себе в качестве комиссионных.

Например, если источник случайных битов выдает 0 и 1 с равными вероятностями, то приемлемым будет соглашение, что в случае выпадения 1 игрок получит $2\mathcal{P}(x)$, а в случае выпадения 0 он не получит ничего, т.е. потеряет весь свой накопленный капитал. Другое возможное соглашение: в случае выпадения 1 игрок получает $1.5\mathcal{P}(x)$, а в случае выпадения 0 он получает $0.5\mathcal{P}(x)$.

В некоторых случаях игрок может так удачно делать ставки, что его текущий капитал $\mathcal{P}(\omega^n)$ становится неограниченным, где $\omega = \omega_1\omega_2\ldots$ – последовательность исходов в случае неограниченного продолжения игры. Теорема Дуба об ограниченных снизу мартингалах (супермартингалах) утверждает, что вероятность этого события равна нулю (см. [25]). Эта теорема позволяет казино с вероятностью единица не разориться.

Пусть Q – мера и $\mathcal{P}$ – Q-супермартингал. Легко видеть, что функция $P(x) = \mathcal{P}(x)Q(x)$ является полумерой. Если Q – вычислимая мера и $\mathcal{P}$ – перечислимый снизу Q-супермартингал, то P – перечислимая снизу полумера. Верно и обратное: для любых перечислимой снизу полумеры полумеры P и вычислимой меры Q функция $\mathcal{P}(x) = \frac{P(x)}{O(x)}$ является перечислимым снизу Q-супермартингалом.

Для любой вычислимой меры Q и априорной полумеры M Q-супермартингал $\mathcal{M}_Q(x) = \frac{M(x)}{Q(x)}$ является максимальным с точностью до мультипликативной константы перечислимым снизу Q-супермартингалом.

Универсальная полумера M в отличие от супермартингалом $\mathcal{M}_Q$ не зависит от какой-либо меры.

Имеет место аналог теоремы 6.2.

Теорема 6.4. *Бесконечная последовательность ω случайна по вычислимой мере Q тогда и только тогда, когда*

$$\sup_n \mathcal{M}_Q(\omega^n) < \infty.$$

6.2. Универсальный предиктор

Рассмотрим задачу прогнозирования последующего бита некоторой последовательности $\omega_1\omega_2\ldots$, поступающей в режиме онлайн. При этом, при построении предиктора мы хотим использовать как можно меньше предположений об источнике, генерирующем последовательность.

Предполагаем, что источник определелеяется некоторой неизвестной нам вычислимой мерой. Допустим, что уже известны первые $n-1$ битов

$$\omega_1,\ \ldots,\ \omega_{n-1}$$

последовательности. Необходимо предсказать вероятность того, что следующий бит ω_n равен 0 или 1.

Пусть источник генерирует бит $\omega_n = 0$ с неизвестной нам вероятностью [3]

$$P(0|\omega^{n-1}) = \frac{P(\omega^{n-1}0)}{P(\omega^{n-1})}.$$

Соответственно, вероятность события $\omega_n = 1$ равна

$$P(1|\omega^{n-1}) = \frac{P(\omega^{n-1}1)}{P(\omega^{n-1})}.$$

[3] Для простоты мы предполагаем, что $P(x) > 0$ для всех x.

Предсказателю эти вероятности не известны. Эадача предсказателя состоит в том, чтобы построить некоторый универсальный метод предсказания, который вычислял бы эти вероятности или некоторые приближения к ним на основе только последовательности $\omega_1 \dots \omega_{n-1}$ без использования распределения P источника.

Будут представлены два универсальных метода предсказания. В разделе 6.2.1 будет рассмотрено правило Лапласа. Этот метод предсказания использует дополнительное предположение о том, что биты последовательности $\omega_1 \dots \omega_{n-1}$ независимо и одинаково распределены, однако вероятность каждого бита ω_n не известна предсказателю. В разделе 6.2.2 будет представлен универсальный предиктор Соломонова, который имеет более общий характер и не использует предположения о независимости, используется только то, что распределение источника представляет собой вычислимую меру. Заметим, что оба метода прогнозирования построены на основе сходной идеи – построить метод прогнозирования, который работает не хуже чем любой метод из широкого класса.

6.2.1. Правило Лапласа

Исторически первой процедурой универсального прогнозирования является правило Лапласа. [4] Эта процедура использует гипотезу о том, что исходы ω_i порождаются некоторым источником, который генерирует их независимо друг от друга с одной и той же вероятностью единицы, равной p.

[4] Лаплас рассматривал задачу вычисления вероятности события, которое заключается в том, что солнце взойдет завтра, если известно, что оно всходило каждый день последние 5000 лет (1826251 день). Эта задача находится на границе применимости частотной интерпретации вероятности, так как не выполнено условие повторяемости данного опыта. Кроме того, известная последовательность состоит из одних восходов. Тем не менее, для того, чтобы выразить степень нашей субъективной неопределенности о значении p этой вероятности, мы считаем все значения p равновозможными. Согласно формуле, приведенной далее, вероятность рассматриваемого события будет равна $\frac{1826252}{1826253}$.

Пусть исходы ω_i принадлежат множеству $\{0,1\}$. Мы также предполагаем, что в каждый момент времени $i = 1, 2, \ldots$ исход ω_i порождается независимо от предыдущих исходов с неизвестными нам постоянными вероятностями $p = P\{\omega_i = 1\}$ и $q = P\{\omega_i = 0\} = 1 - p$.

Пусть мы наблюдаем исходы $\omega^n = \omega_1, \ldots \omega_n$, в которых имеется n_1 единиц и n_2 нулей, $n_1 + n_2 = n$. Вероятность получить такую последовательность исходов равна $p^{n_1}(1-p)^{n_2}$, если вероятность единицы равна p. Так как истинная вероятность p неизвестна, рассмотрим байесовскую смесь вероятностей последовательности длины n по всем возможным p :

$$P(\omega^n) = \int_0^1 p^{n_1}(1-p)^{n_2} dp.$$

Этот интеграл равен $\frac{1}{(n+1)\binom{n}{n_1}}$. Доказательство см. в виде задачи из раздела 6.3.

Условная вероятность события $\omega_{n+1} = 1$ при известных исходах $\omega^n = \omega_1, \ \ldots, \ \omega_n$ равна

$$P\{\omega_{n+1} = 1|\omega^n\} = \frac{P(\omega^n 1)}{P(\omega^n)} = \frac{\frac{1}{(n+2)\binom{n+1}{n_1+1}}}{\frac{1}{(n+1)\binom{n}{n_1}}} = \frac{n_1+1}{n+2}.$$

Таким образом, получаем правило Лапласа:

$$P\{\omega_{n+1} = 1|\omega^n\} = \frac{n_1+1}{n+2},$$
$$P\{\omega_{n+1} = 0|\omega^n\} = \frac{n_2+1}{n+2}.$$

Качество такой процедуры прогнозирования можно оценивать с помощью оценки избыточности соответствующего кода. Величина

$$L_p(\omega^n) = -\log(p^{n_1}(1-p)^{n_2}),$$

с точностью до 1, совпадает с количеством двоичных битов, необходимых для кодирования последовательностей ω^n (с помощбю кода Шеннона для блоков длины n), состоящих из n_1 единиц и n_2 нулей и порождаемых источником с вероятностью единицы, равной p. Нетрудно проверить, что

$$\sup_{0\leqslant p\leqslant 1} p^{n_1}(1-p)^{n_2} = \left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}.$$

Для правила Лапласа

$$L(\omega^n) = -\log P(\omega^n) = -\log \int\limits_0^1 p^{n_1}(1-p)^{n_2}dp.$$

Допустим, что последовательность ω^n порождена источником с вероятностью $\omega_n = 1$, равной p_0. Тогда для произвольной последовательности ω^n выполнено

$$L(\omega^n) - L_{p_0}(\omega^n) \leqslant L(\omega^n) - \inf_{0\leqslant p\leqslant 1} L_p(\omega^n) = \log \frac{\sup\limits_{0\leqslant p\leqslant 1} p^{n_1}(1-p)^{n_2}}{\int\limits_0^1 p^{n_1}(1-p)^{n_2}dp} = \log \frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}}{\frac{1}{(n+1)\binom{n}{n_1}}} \leqslant \log(n+1).$$

Таким образом, используя для кодирования вероятности, вычисленные по правилу Лапласа, мы истратим $\log(n+1)$ дополнительных битов по сравнению с длиной оптимального кода, т.е. кода, построенного на основе истинной вероятности p_0 источника, порождающего исходы ω_i.

Другой, более точный, метод прогнозирования был предложен Кричевским и Трофимовым (см.[13], [14], [44]). Рассматривается байесовская смесь вероятностей последовательности длины

n по всем возможным p с плотностью $1/(\pi\sqrt{p(1-p)})$:

$$P(\omega^n) = \int\limits_0^1 \frac{p^{n_1}(1-p)^{n_2}}{\pi\sqrt{p(1-p)}}dp.$$

В этом случае условная вероятность появления единицы после n наблюдений $\omega^n = \omega_1,\ \ldots,\ \omega_n$ равна

$$P(1|\omega^n) = \frac{n_1 + 1/2}{n+1}.$$

Имеет место неравенство

$$\int\limits_0^1 \frac{p^{n_1}(1-p)^{n_2}}{\pi\sqrt{p(1-p)}}dp \geqslant \frac{1}{2\sqrt{n}}\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}.$$

Это утверждение предлагается далее в разделе 6.3 в виде задачи.

Отсюда получаем оценку на дополнительное число битов при кодировании с использованием прогнозирования по методу Кричевского и Трофимова:

$$L(\omega^n) - \inf_{0\leqslant p\leqslant 1} L_p(\omega^n) = \log \frac{\sup\limits_{0\leqslant p\leqslant 1} p^{n_1}(1-p)^{n_2}}{\int_0^1 \frac{p^{n_1}(1-p)^{n_2}}{\pi\sqrt{p(1-p)}}dp} \leqslant$$

$$\leqslant \log \frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}}{\frac{1}{2\sqrt{n}}\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}} \leqslant \log(2\sqrt{n}) = \frac{1}{2}\log n + 1.$$

В этой оценке ошибка асимптотически в два раза меньше, чем в соответствующей оценке для метода Лапласа.

6.2.2. Универсальный предиктор Соломонова

В этом разделе мы рассмотрим теорию Р. Соломонова об универсальных предсказаниях (см. его работы [53], [54]). Будет доказано, что универсальная полумера в среднем прогнозирует биты

бесконечной последовательности не хуже чем произвольная вычислимая мера, которая генерирует эту последовательность.

Универсальный предсказатель Соломонова будет строиться с помощью универсальной полумеры. Пусть M – универсальная перечислимая снизу полумера на дереве двоичных последовательностей. Определим «условную полумеру» бита $b \in \{0,1\}$ при известной конечной последовательности x:

$$M(b|x) = \frac{M(xb)}{M(x)}. \tag{6.5}$$

Функция (6.5) не является мерой (см. задачу из раздела 6.3). Эта функция лишь удовлетворяет соотношению

$$M(0|x) + M(1|x) \leqslant 1$$

для всех x. Кроме того, она не перечислима сверху или снизу как отношение двух перечислимых (но не вычислимых) функций (см. упр. 8 из раздела 6.3). Поэтому эта функция сама по себе имеет чисто теоретическое значение.[5]

Пусть P и Q – две меры на двоичных последовательностях. Нам будет удобно обозначать посредством P_n и Q_n их ограничения на множестве $\{0,1\}^n$.

Рассморим расстояние Кульбака–Лейблера между мерами P_n и Q_n на $\{0,1\}^n$

$$D_n(P_n\|Q_n) = \sum_{l(x)=n} P_n(x) \log \frac{P_n(x)}{Q_n(x)}. \tag{6.6}$$

Ранее было отмечено, что $D_n(P_n\|Q_n) \geqslant 0$ и $D_n(P_n|Q_n) = 0$ тогда и только тогда, когда $P_n = Q_n$.

В теории информации мера Q называется универсальной для класса источников (мер) $\mathcal{P}$, если для любой меры $P \in \mathcal{P}$ выполнено

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l(x)=n} D_n(P_n\|Q_n) = 0.$$

[5]Можно рассматривать различные вычислимые приближения к ней эвристического типа.

Расстояние (6.6) имеет смысл и для полумеры M:

$$D_n(P_n\|M_n) = \sum_{l(x)=n} P(x)\log\frac{P(x)}{M(x)}.$$

Предложение 6.3. *Априорная полумера M является универсальной для класса всех вычислимых мер в смысле теории информации:*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l(x)=n} D_n(P_n\|M_n) = 0$$

для любой вычислимой меры P.

Доказательство. По основному свойству априорной полумеры $cM(x) \geqslant P(x)$ для всех x, где c – константа, зависящая от меры P. Отсюда для любой вычислимой меры P получаем

$$D_n(P_n\|M_n) = \sum_{l(x)=n} P(x)\log\frac{P(x)}{M(x)} \leqslant \log c \sum_{l(x)=n} P(x) \leqslant \log c$$

для всех n. Отсюда для любой вычислимой меры P

$$\lim_{n\to\infty}\frac{1}{n}D_n(P_n\|M_n) = 0.$$

□

Следующая теорема показывает силу прогностических возможностей универсального предиктора (6.5).

Теорема 6.5. *Для любой вычислимой меры P и для любого бита $b \in \{0,1\}$ выполнено*

$$\sum_{n=1}^{\infty}\sum_{l(x)=n} P(x)(M(b|x) - P(b|x))^2 < \infty. \tag{6.7}$$

Выражение (6.7) представляет собой сумму математических ожиданий по мере P (рассматриваемой на последовательностях длины n) квадратов разностей между условными вероятностями $n+1$ бита, выдаваемых неизвестным нам источником, и прогнозами нашего универсального предиктора (6.5). Из того, что эта сумма ограничена, в частности, следует, что эти математические ожидания стремятся к нулю.

Доказательство. Мы докажем теорему для $b = 0$. Для $b = 1$ доказательство аналогичное.

Пусть P и Q – две меры на двоичных последовательностях, P_n и Q_n – их ограничения на множестве $\{0,1\}^n$.

Расстояние Кульбака–Лейблера между P_n и Q_n равно

$$D_n(P_n\|Q_n) = \sum_{l(x)=n} P_n(x)\log\frac{P_n(x)}{Q_n(x)}. \tag{6.8}$$

Выражение $D_n(P_n\|Q_n)$ имеет смысл также в случае, когда P_n или Q_n являются полумерами. Для этого случая нам потребуется свойство 2) следующей технической леммы. Свойство 1) связывает неравенством два способа измерения расстояния между мерами на на множестве $\{0,1\}$.

Лемма 6.2. 1) *(Неравенство Пинскера) Для мер P_1 и Q_1 на двухэлементном множестве $\{0,1\}$ выполнено неравенство*

$$D_1(P_1\|Q_1) \geqslant 2\log e(P_1(0) - Q_1(0))^2. \tag{6.9}$$

2) *Пусть функция Q_1 удовлетворяет более слабому неравенству $Q_1(0)+Q_1(1) \leqslant 1$ и пусть $Q'_1(0) = Q_1(0)$, $Q'_1(1) = 1-Q_1(0)$ – «расширяющая» ее мера. Тогда для любой меры P_1 на $\{0,1\}$ выполнено неравенство*

$$D_1(P_1\|Q_1) \geqslant D_1(P_1\|Q'_1). \tag{6.10}$$

Доказательство этой леммы предлагается в виде задачи из раздела 6.3.

В дальнейшем мы применим эту лемму к мере P и априорной полумере M. Искусственным образом увеличим значения M так, чтобы она стала мерой M', причем выполнялось бы $M'(0|x) = M(0|x)$ для каждого x. Тогда для нее будет выполнено неравенство

$$D_1(P_1\|M_1) \geqslant D_1(P_1\|M'_1).$$

Для $x \in \{0,1\}^n$ и $b \in \{0,1\}$ обозначим

$$P_{n+1}(b|x) = \frac{P_{n+1}(xb)}{P_n(x)}.$$

Пусть $P_{n+1}(\cdot|x)$ – соответствующее вероятностное распределение на двухэлементном множестве $\{0,1\}$. Доказательство теоремы будет основываться на следующей лемме.

Лемма 6.3. *Пусть P – мера, Q – полумера. Тогда для любого n выполнено*

$$D_{n+1}(P_{n+1}\|Q_{n+1}) = D_n(P_n\|Q_n) + \sum_{l(x)=n} P_n(x) D_1(P_1(\cdot|x)\|Q_{n+1}(\cdot|x)).$$

Доказательство. Раскроем величину $D_1(P_{n+1}\|Q_{n+1})$:

$$\begin{aligned}
D_{n+1}(P_{n+1}\|Q_{n+1}) &= \sum_{l(z)=n+1} P_{n+1}(z) \log \frac{P_{n+1}(z)}{Q_{n+1}(z)} = \\
&= \sum_{l(x)=n, b\in\{0,1\}} P_{n+1}(xb) \log \frac{P_{n+1}(xb)}{Q_{n+1}(xb)} = \\
&= \sum_{l(x)=n, b\in\{0,1\}} P_n(x) P_{n+1}(b|x) \log \frac{P_n(x) P_{n+1}(b|x)}{Q_n(x) Q_{n+1}(b|x)} = \\
&= \sum_{l(x)=n} P_n(x) \log \frac{P_n(x)}{Q_n(x)} \sum_{b\in\{0,1\}} P_{n+1}(b|x) + \\
&\quad + \sum_{l(x)=n} P_n(x) \sum_{b\in\{0,1\}} P_{n+1}(b|x) \log \frac{P_{n+1}(b|x)}{Q_{n+1}(b|x)} = \\
&= D_n(P_n|Q_n) + \sum_{l(x)=n} P_n(x) D_1(P_{n+1}(\cdot|x)\|Q_{n+1}(\cdot|x)).
\end{aligned}$$

Здесь мы использовали равенство $\sum_{b\in\{0,1\}} P_{n+1}(b|x) = 1$ и определение (6.8). Лемма доказана. $\square$

Раскрывая сумму в лемме 6.2, получаем следующее следствие.

Следствие 6.1. *Для любого n выполнено*

$$D_n(P_n|Q_n) = \sum_{i=1}^{n} \sum_{l(x)=i} P_i(x) D_1(P_i(\cdot|x)\|Q_i(\cdot|x)). \qquad (6.11)$$

Переходим к доказательству теоремы. Пусть теперь P – вычислимая мера из условия теоремы, M' – дополненная до меры универсальная полумера. Применяем лемму 6.3 к каждому слагаемому суммы (6.11), где $Q = M$, получаем

$$\begin{aligned} D_n(P_n\|M_n) &= \sum_{i=1}^{n} \sum_{l(x)=i} P_i(x) D_1(P_i(\cdot|x)\|M_i(\cdot|x)) \geqslant \\ &\geqslant \sum_{i=1}^{n} \sum_{l(x)=i} P_i(x) D_1(P_i(\cdot|x)\|M'_i(\cdot|x)) \geqslant \\ &\geqslant 2\log e \sum_{i=1}^{n} \sum_{l(x)=i} P(x)(M(0|x) - P(0|x))^2. \qquad (6.12) \end{aligned}$$

По основному свойству априорной полумеры $cM(x) \geqslant P(x)$ для всех x, где c – константа, зависящая от меры P. Отсюда получаем

$$D_n(P_n\|M_n) = \sum_{l(x)=n} P(x) \log \frac{P(x)}{M(x)} \leqslant \log c \sum_{l(x)=n} P(x) \leqslant \log c$$

для всех n. Соединяем это неравенство с (6.12) и получаем необходимую нам оценку (6.7):

$$\sum_{i=1}^{\infty} \sum_{l(x)=i} P(x)(M(0|x) - P(0|x))^2 \leqslant \frac{\ln 2}{2} \sup_n D_n(P_n\|M_n) \leqslant \frac{\ln 2}{2} \log c.$$

Теорема доказана. $\square$

6.3. Задачи и упражнения

1. Доказать, что
(a) $\mathrm{KM}(x) \leqslant \mathrm{KA}(x) + 2\log l(x) + O(1)$;
(b) $\mathrm{KP}(x) \leqslant \mathrm{KA}(x) + 2\log l(x) + O(1)$.

2. Можем ли мы утверждать, что функции Q_m из доказательства теоремы 6.2 вычислимые?

3. Задана вероятностная машина (L, F). Определим

$$P(x) = L\{\omega : x \subseteq F(\omega)\}.$$

Доказать, что функция $P(x)$ является перечислимой снизу полумерой.

4. Доказать, что для любой перечислимой снизу полумеры P можно построить такую вычислимую операцию F, что для нее выполнено (6.4).

5. Доказать, что для любой вычислимой последовательности ω существует такая константа $c > 0$, что $M(\omega^n) > c$ для всех n.

6. Доказать теорему 6.4.

7. Доказать, что не существует максимальной с точностью до мультипликативной константы вычислимой меры и, таким образом, априорная полумера M не является мерой.

8. Доказать, что функция (6.5) не является перечислимой сверху или снизу.

9. Доказать, что $\int\limits_0^1 p^{n_1}(1-p)^{n_2}dp = \frac{1}{(n+1)\binom{n}{n_1}}$.

Указание. Проверим это равенство обратной индукцией по n_1. При $n_1 = n$ имеем $\int\limits_0^1 p^n dp = \frac{1}{(n+1)}$.

Предположим, что

$$\int\limits_0^1 p^{n_1+1}(1-p)^{n_2-1}dp = \frac{1}{(n+1)\binom{n}{n_1+1}}.$$

Интегрируя по частям, получим

$$\int_0^1 p^{n_1}(1-p)^{n_2}dp = \frac{n-n_1}{n_1+1}\int_0^1 p^{n_1+1}(1-p)^{n_2-1}dp = \\ = \frac{n-n_1}{n_1+1}\frac{1}{(n+1)\binom{n}{n_1+1}} = \frac{1}{(n+1)\binom{n}{n_1}}.$$

10. Доказать нераенство $\int\limits_0^1 \frac{p^{n_1}(1-p)^{n_2}}{\pi\sqrt{p(1-p)}}dp \geqslant \frac{1}{2\sqrt{n}}\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}$.

11. Доказать

1) утверждение (6.9) леммы 6.2.

2) утверждение (6.10) леммы 6.2.

Решение. Утверждение 1). Для мер P_1 и Q_1 на двухэлементном множестве $\{0,1\}$ выполнено неравенство $D_1(P_1\|Q_1) \geqslant 2(P_1(0)-Q_1(0))^2$.

Обозначим $p = P_1(0)$ и $q = Q_1(0)$. Тогда $D_1(P_1\|Q_1) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$.

Фиксируем p и изучим поведение функции

$$f(q) = D_1(P_1\|Q_1) - 2(p-q)^2.$$

При $q = p$, $f(q) = 0$. Временно заменим в определении $D(P\|Q)$ log на ln.

При $q < p$, $f(q)$ убывает, так как $f'(q) < 0$, а при $q > p$ функция $f(q)$ возрастает, так как $f'(q) > 0$. Действительно, $f(q) = p\ln p - p\ln q + (1-p)\ln(1-p) - (1-p)\ln(1-q) - 2(p-q)^2$, ее производная по q равна

$$f'(q) = -\frac{p}{q} + \frac{1-p}{1-q} + 4(p-q)^2 = \frac{q-p+4q(1-p)(p-q)}{q(1-q)} = \\ \frac{(p-q)(4q(1-q)-1)}{q(1-q)}.$$

Так как $q(1-q) \leqslant \frac{1}{4}$, $f'(q) < 0$ при $q < p$, и $f'(q) > 0$ при $q > p$. Значит $f(q) \geqslant 0$ для всех p и q.

Утверждение 2). Пусть P_1 – мера, а функция Q_1 удовлетворяет более слабому неравенству $Q_1(0) + Q_1(1) \leqslant 1$ и пусть $Q'_1(0) = Q_1(0)$, $Q'_1(1) = 1 - Q_1(0)$ – «расширяющая» ее мера. Тогда для любой меры P_1 на $\{0,1\}$ выполнено неравенство $D(P_1|Q_1) \geqslant D(P_1|Q'_1)$.

Обозначим $p = P_1(0)$, $1 - p = P_1(1)$, $q = Q_1(0) = Q'(0)$, $1 - q = Q'_1(1)$. Тогда $Q_1(1) \leqslant 1 - Q_1(0) = Q'(1) = 1 - q$. Тогда выполнено неравенство

$$D_1(P_1\|Q_1) = p\ln\frac{p}{q} + (1-p)\frac{1-p}{Q_1(1)} \geqslant$$
$$p\ln\frac{p}{q} + (1-p)\frac{1-p}{1-q} = D(P_1\|Q'_1).$$

12. Доказать, что существует такая константа c, что для любой перечислимой снизу полумеры P выполнено

$$cM(x) \geqslant 2^{-\mathrm{KP}(P)}P(x)$$

для всех x, где $\mathrm{KP}(P)$ – префиксная сложность программы, перечисляющей снизу значения полумеры P.

13. Доказать, что

a) не существует максимальной с точностью до мультипликативной константы вычислимой меры P, т.е. такой, что для любой вычислимой меры Q выполнено $cP(x) \geqslant Q(x)$ для всех x, где c – константа, зависящая от Q.

b) универсальная перечислимая снизу полумера на дереве двоичных последовательностей $M(x)$ не является вычислимой функцией и $0 < \inf_n \sum_{l(x)=n} M(x) < 1$.

14. Для любой полумеры P определим функцию

$$\bar{P}(x) = \inf_{n \geqslant l(x)} \sum_{y: x \subseteq y, l(y)=n} P(y). \qquad (6.13)$$

Функцию $\bar{P}$ можно распространить на произвольные борелевские подмножества множества Ω, полагая $\bar{P}(\Gamma_x) = \bar{P}(x)$.

Доказать, что

a) $\bar{P}$ – мера;

b) если P мера, то $\bar{P}(x) = P(x)$ для всех x, а также $\bar{P}(A) = P(A)$ для всех борелевских $A \subseteq \Omega$;

c) для любой вычислимой меры P существует константа c такая, что $c\bar{M}(x) \geqslant P(x)$ для всех x и $c\bar{M}(A) \geqslant P(A)$ для всех измеримых $A \subseteq \Omega$;

d) $\bar{M}$ совпадает с максимальной мерой Q такой, что $Q(x) \leqslant M(x)$ для всех x;

e) $0 < \bar{M}(\lambda) < 1$, где M – универсальная полумера;

f) мера $\bar{M}$ не является ни вычислимой, ни перечислимой снизу или сверху.

15. Пусть α – бесконечная последовательность. Доказать, что

a) для любой полумеры P, $P(\alpha^n) > 0$ для всех n тогда и только тогда, когда $\bar{P}(\{\alpha\}) > 0$;

b) бесконечная последовательность α вычислима тогда и только тогда, когда она является атомом меры $\bar{M}$, т.е. когда $\bar{M}(\{\alpha\}) > 0$.

16. Доказать, что для любой перечислимой полумеры P существует вычислимый оператор F такой, что

$$P(x) = L\{\omega \in \Omega : x \subseteq F(\omega)\}. \tag{6.14}$$

Доказать, что $P(x) = L(\cup_{x \subseteq F(z)} \Gamma_z)$.

17. Пусть полумера P определена согласно (6.14), а мера $\bar{P}$ определена согласно (6.13). Доказать, что для любого измеримого множества $A \subseteq \Omega$ будет

$$\bar{P}(A) = L\{\omega \in \Omega : F(\omega) \in A\} = F^{-1}(A).$$

18. Доказать, что для любой меры Q, для Q-почти всех бесконечных последовательнстей ω существует предел $\lim\limits_{n\to\infty} \frac{M(\omega^n)}{Q(\omega^n)}$.[6]

[6]Существует ли этот предел для любой бесконечной последовательности ω, случайной относительно вычислимй меры Q, является открытой проблемой.

Часть III

Алгоритмический анализ утверждений теории вероятностей

Идеи алгоритмической вычислимости могут быть использованы для алгоритмического анализа теории вероятностей. Намного ранее подобный анализ был проведен в других областях классической математики – в топологии и теории метрических пространств, в математическом анализе. Алгоритмический анализ заключается в проверке доказательств на их конструктивность. При этом обычно происходит расслоение классических утверждений по их степени конструктивности, появляются новые утверждения о невозможности построения конструктивных аналогов некоторых утверждений, даже в том случае, когда в классическом случае они верны.

Как оказалось, вероятностные утверждения также могут быть в разной степени конструктивными. Большинство утверждений теории вероятностей являются конструктивными в самом сильном смысле, поскольку для этих утверждений существуют вычислимые оценки скорости сходимости. Эти оценки позволяют доказывать, что асимптотические законы теории вероятности, такие как усиленный закон больших чисел или закон повторного логарифма, выполнены потраекторно – для алгоритмически случайных последовательностей исходов. Для усиленного закона больших чисел это было показано в главе 4. Для закона повторного логарифма даже удалось найти абсолютно новое доказательство, основанное на идее оптимального кодирования в духе колмогоровского подхода. Это доказательство приведено в разделе 7.

Как мы покажем в разделе 8, с эргодической теоремой Биркгофа дело обстоит сложнее. Эргодическая теорема Биркгофа не является алгоритмически эффективной при классическом понимании процесса конструктивизации. Мы покажем, что не существует вычислимой оценки скорости сходимости в этой теореме. Тем не менее мы покажем, что просто сходимость (без вычислимой оценки ее скорости) имеет место на каждой индивидуальной алгоритмически случайной последовательности. Таким образом, понятие *алгоритмически случайной последовательности* позволяет построить конструктивную эргодическую теорию.

Глава 7

Сложностное доказательство закона повторного логарифма

В этом разделе мы приведем доказательство первой части закона повторного логарифма для случайных последовательностей. Хотя классическое доказательство этой теоремы (см. [25]) прямо транслируется в конструктивную форму, полезно рассмотреть новое доказательство этой теоремы в духе идей колмогоровского алгоритмического подхода к теории вероятностей. Это доказательство было предложено Вовком [2]. Первая часть этого доказательства (неравенства $\leqslant$) также представлена в монографии [27]. Мы следуем этому изложению.

Приводимое ниже доказательство отличается от классических вероятностных доказательств и основано на идеях оптимального кодирования.

Для призвольной двоичной последоватедьности ω обозначим $S_n(\omega) = \sum_{i=1}^{n} \omega_i$. Мы докажем первую часть закона повторного логарифма.

Теорема 7.1. *Для любой бесконечной двоичной последователь-*

ности ω выполнена импликация:

$$\mathrm{KM}(\omega^n) \geqslant n - O(1) \Rightarrow \limsup_{n\to\infty} \frac{|S_n(\omega) - \frac{n}{2}|}{\sqrt{\frac{1}{2} n \ln\ln n}} \leqslant 1. \qquad (7.1)$$

Доказательство теоремы будет основано на достаточно точной верхней оценке для сложности $\mathrm{KM}(\omega^n)$.

Предварительно мы приведем доказательство более слабой оценки. По существу, это будет новое доказательство теоремы 3.2.

По лемме 5.8 для любой вычислимой меры Q выполнено

$$\mathrm{KM}(\omega^n) \leqslant -\log Q(\omega^n) + O(1).$$

Найдем удобную меру Q. Пусть $p_n = \frac{k}{n}$, $k = S_n(\omega^n)$ – число единиц в начальном фрагменте ω длины n. Рассмотрим для фиксированного n бернуллиевскую меру с вероятностью единицы p_n на последовательностях длины n. Получаем

$$\begin{aligned} -\log B_{p_n}(\omega^n) &= -\log(p_n^k (1-p_n)^{n-k}) = \\ &= -n \log(p_n^{p_n} (1-p_n)^{1-p_n}) = \\ = n(-p_n \log p_n - (1-p_n)\log(1-p_n)) &= nH(p_n), \end{aligned} \qquad (7.2)$$

где $H(p_n)$ – энтропия Шеннона.

Однако функция $B_{p_n}(\omega^n)$ не распространяется на все конечные последовательности, так как параметр меры зависит от аргумента. Для того чтобы избавиться от этой зависимости, введем смесь бернуллиевских мер – перечислимую снизу полумеру:

$$Q(x) = \sum_{r\in\mathcal{Q}_+} B_r(x) P(r),$$

где $\mathcal{Q}_+$ – множество всех положительных рациональных чисел (которые отождествлены со всеми натуральными числами), P – априорная полумера на множестве всех натуральных чисел.

По лемме 5.8

$$\begin{aligned}\mathrm{KM}(x) &\leqslant -\log Q(x) + O(1) \leqslant \\ &\leqslant -\log B_r(x) - \log P(r) + O(1)\end{aligned}$$

для любого r. При $x = \omega^n$ и $r = p_n$

$$\mathrm{KM}(\omega^n) \leqslant -\log B_{p_n}(\omega^n) + \mathrm{KP}(p_n) + O(1).$$

Для произвольного $\epsilon > 0$ имеет место оценка

$$\mathrm{KP}(p_n) \leqslant \mathrm{KP}(n) + \mathrm{KP}(k) + O(1) \leqslant (2+\epsilon)\log n + O(1),$$

так как $k \leqslant n$.

Сравниваем нижнюю и верхние оценки для сложности:

$$n - O(1) \leqslant \mathrm{KM}(\omega^n) \leqslant nH(p_n) + (2+\epsilon)\log n + O(1). \qquad (7.3)$$

Напишем формулу Тэйлора для функции $H(p)$ в окрестности точки $p = \frac{1}{2}$. Представим $p_n = \frac{1}{2} + \nu_n$. Имеем $H'(\frac{1}{2}) = 0$, а также $H''(\frac{1}{2}) = -\frac{4}{\ln 2}$. Отсюда

$$H(p_n) = 1 - \frac{2}{\ln 2}\nu_n^2 + o(\nu_n^2). \qquad (7.4)$$

Из неравенства (7.3) получаем

$$\nu_n^2 - o(\nu_n^2) = \frac{\ln 2}{2}(1 - H(p_n)).$$

Отсюда и из (7.4), для произвольного $\epsilon > 0$, получаем оценку

$$\left| p_n - \frac{1}{2} \right| \leqslant \sqrt{\frac{\ln 2}{2}(2+\epsilon)\frac{\ln n}{n} + O\left(\frac{1}{n}\right)}. \qquad (7.5)$$

Доказательство теоремы. Переходим теперь к доказательству точного результата (7.1).

Оценку (7.5) можно улучшить за счет уменьшения члена $\mathrm{KP}(p_n)$, если использовать более простое по сложности приближение к ν_n из представления $p_n = \frac{1}{2} + \nu_n$. Рассмотрим случай

$\nu_n > 0$. В качестве такого приближения возьмем рациональное число $\nu'_n = (1-\epsilon)^m$, где m такое, что

$$(1-\epsilon)^m \leqslant \nu_n < (1-\epsilon)^{m-1}.$$

Отсюда $m = \lfloor \log_{1-\epsilon} \nu_n \rfloor$.

Оцениваем префиксную сложность этого параметра:

$$\begin{aligned} \mathrm{KP}(m) &\leqslant (1+\epsilon)\log m + O(1) = \\ &= (1+\epsilon)\log\left(\frac{\log \nu_n}{\log(1-\epsilon)}\right) + O(1) = \\ &= (1+\epsilon)\log(-\log \nu_n) + O(1). \end{aligned} \tag{7.6}$$

Из предварительной оценки (7.5) следует, что можно считать, что $\nu_n = O\left(\sqrt{\frac{\log n}{n}}\right)$. Тогда оценка (7.6) превращается в оценку

$$\mathrm{KP}(m) \leqslant (1+\epsilon)\log\log n + O(1)$$

для некоторого $\epsilon > 0$.

Уточним оценку (7.2) при $p'_n = \frac{1}{2} + \nu'_n$:

$$\begin{aligned} -\log B_{p'_n}(\omega^n) &= n(-p_n \log p'_n - (1-p_n)\log(1-p'_n)) \leqslant \\ &\leqslant n(-p'_n \log p'_n - (1-p'_n)\log(1-p'_n)) = nH(p'_n). \end{aligned} \tag{7.7}$$

Для получения этих неравенств мы использовали неравенства $p_n \geqslant p'_n > \frac{1}{2}$ и $-\log p'_n < -\log(1-p'_n)$ при $p'_n > \frac{1}{2}$. Переход от верхнего неравенства в (7.7) к нижнему объясняется тем, что мы увеличили вес второй скобки за счет той же доли от меньшей скобки, т.е. мы использовали неравенство $1-p'_n > 1-p_n$.

Повторяем предыдущие оценки типа (7.3) теперь уже для p'_n. Для всех достаточно малых $\epsilon > 0$ выполнены следующие оценки.

Каждая оценка, приведенная ниже, следует из вышестоящей

оценки:

$$
\begin{aligned}
n - O(1) \leqslant \mathrm{KM}(\omega^n) &\leqslant -\log B_{p'_n}(\omega^n) + \mathrm{KP}(p'_n) + O(1),\\
n - O(1) \leqslant \mathrm{KM}(\omega^n) &\leqslant -\log B_{p'_n}(\omega^n) + \mathrm{KP}(\nu'_n) + O(1),\\
n &\leqslant nH(p'_n) + (1+\epsilon)\log\log n + O(1),\\
\frac{2}{\ln 2}(\nu'_n)^2 n &\leqslant (1+\epsilon)\log\log n + O(1),\\
\nu'_n &\leqslant (1+\epsilon)\sqrt{\frac{\ln 2}{2n}\log\log n + O\left(\frac{1}{n}\right)},\\
\nu_n &\leqslant (1+3\epsilon)\sqrt{\frac{\ln 2}{2n}\log\log n + O\left(\frac{1}{n}\right)},\\
\nu_n &\leqslant (1+3\epsilon)\sqrt{\frac{\ln\ln n}{2n} + O\left(\frac{1}{n}\right)},\\
\left|p_n - \frac{1}{2}\right| &\leqslant (1+3\epsilon)\sqrt{\frac{\ln\ln n}{2n} + O\left(\frac{1}{n}\right)},\\
\left|S_n(\omega) - \frac{n}{2}\right| &\leqslant (1+4\epsilon)\sqrt{\frac{1}{2}n\ln\ln n + O(n)}.
\end{aligned}
$$

Здесь мы использовали неравенство $\nu_n \leqslant (1-\epsilon)^{-1}\nu'_n \leqslant (1+2\epsilon)\nu'_n$ для $\epsilon < \frac{1}{2}$.

Отсюда

$$
\left|S_n(\omega) - \frac{n}{2}\right| \leqslant (1+5\epsilon)\sqrt{\frac{1}{2}n\ln\ln n}
$$

для всех достаточно больших n.

Поскольку $\epsilon > 0$ – произвольное достаточно малое, из этого неравенства следует утверждение теоремы – оценка (7.1). □

Первую часть закона повторного логарифма можно записать в терминах равенства для монотонной сложности.

Следствие 7.1. *Для любой бесконечной двоичной последова-*

тельности ω выполнена импликация:

$$\mathrm{KM}(\omega^n) = n + O(1) \Rightarrow \limsup_{n\to\infty} \frac{|S_n(\omega) - \frac{n}{2}|}{\sqrt{\frac{1}{2} n \ln\ln n}} \leqslant 1.$$

В работе [2] также приведено доказательство второй части закона повторного логарифма (неравенства $\geqslant$) в алгоритмической теории случайности.

Глава 8

Эргодическая теорема Биркгофа

Доказательства большинства законов теории вероятностей прямо транслируются в конструктивную форму. Первое затруднение с получением конструктивного аналога возникло в связи с эргодической теоремой Биркгофа. Известные из учебников доказательства этой теоремы прямо не транслируются в конструктивную форму. По-видимому, это связано с отсутствием алгоритмически эффективных оценок скорости сходимости по вероятности, а также почти всюду в этой теореме. Невозможность построения таких оценок будет показана в разделе 8.3.

Тем не менее, конструктивный анализ этой теоремы, приведенный в монографии Бишопа [36], позволяет доказать эргодическую теорему для случайных по Мартин-Лефу последовательностей. Мы докажем этот конструктивный вариант эргодической теоремы в разделе 8.5. Для этого нам придется ввести новый тип тестов случайности – интегральные тесты, с помощью которых можно дать новую эквивалентную формулировку случайности по Мартин-Лефу.

8.1. Эргодическая теория

Мы рассматриваем вероятностное пространство $(\Omega, \mathcal{F}, P))$, где Ω – множество всех бесконечных двоичных последовательностей, $\mathcal{F}$ – борелевское поле, которое порождается интервалами

$$\Gamma_x = \{\omega \in \Omega : x \subset \omega\},$$

где x – произвольная конечная последовательность, P – вычислимая вероятностная мера на Ω.

Произвольное измеримое отображение $T : \Omega \to \Omega$ называется *преобразованием* пространства Ω. Преобразование T *сохраняет меру*, если $P(T^{-1}(A)) = P(A)$ для всех измеримых подмножеств $A \subseteq \Omega$. Измеримое множество A называется *инвариантным* относительно преобразования T, если $T^{-1}A = A$ с точностью до множества меры 0. [1] Преобразование T называется *эргодическим*, если каждое инвариантное относительно него множество имеет меру 0 или 1.

Задаем k-ю итерацию преобразования T рекурсивно: определим $T^0\omega = \omega$ и $T^k\omega = T(T^{k-1}\omega)$ при $k \geqslant 1$, где $\omega \in \Omega$.

Преобразование T порождает бесконечную траекторию

$$\omega, T\omega, T^2\omega, \ \ldots, \ T^k\omega, \ \ldots$$

произвольной точки $\omega \in \Omega$.

Пример преобразования пространства Ω, сохраняющего равномерную меру L – сдвиг T, определенный условием

$$T(\omega_1\omega_2\omega_3\ldots) = \omega_2\omega_3\ldots$$

Если сдвиг сохраняет меру P, то такая мера называется *стационарной*. Легко доказать, что мера является стационарной тогда и только тогда, когда

$$P\{\omega : \omega_i = x_1, \ldots \ , \omega_{i+k-1} = x_k\} = P\{\omega : \omega_1 = x_1, \ldots \ , \omega_k = x_k\}$$

[1] То есть $P((T^{-1}(A) \setminus A) \cup (A \setminus T^{-1}(A))) = 0$.

для любой конечной последовательности $x = x_1 \dots x_k$ и для любого i.

Если сдвиг сохраняет меру P и к тому же является эргодическим преобразованием, то и сама мера P называется *эргодической*.

Эргодическая теорема Биркгофа формулируется следующим образом.

Теорема 8.1. *Пусть P – произвольная вероятностная мера на Ω и f – произвольная интегрируемая функция типа $\Omega \to \mathcal{R}$ (которая называется наблюдаемой). Тогда для любого сохраняющего меру P преобразования T для P-почти всех ω выполнено*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = \tilde{f}(\omega),$$

где $\tilde{f}$ – интегрируемая инвариантная относительно T функция такая, что $\int \tilde{f}(\omega)dP = \int f(\omega)dP$. [2] *Кроме того, если преобразование T является эргодическим, то $\tilde{f}(\omega) = \int f(\omega)dP$ для P-почти всех ω.*

Эргодическая теорема утверждает, что среднее по времени значение некоторой наблюдаемой величины f вдоль траектории почти любой точки ω равно среднему значению этой наблюдаемой по всему пространству.

Если мера P является стационарной, т.е. инвариантной относительно сдвига, а наблюдаемая имеет вид $f(\omega) = \omega_1$, то эргодическая теорема Биркгофа превращается в усиленный закон больших чисел:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \omega_k = \tilde{f}(\omega),$$

где $\int \tilde{f}(\omega)dP = P(\Gamma_1) = P\{\omega : \omega_1 = 1\}$.

[2] Функция $\tilde{f}$ инвариантная относительно T, если $\tilde{f}(T\omega) = \tilde{f}(\omega)$.

8.2. Теорема Пуанкаре о возвращении

Приведем пример конструктивного аналога широко известного утверждения эргодической теории – теоремы Пуанкаре о возвращении, которая утверждает следующее.

Пусть T – преобразование, сохраняющее меру P, и E – измеримое множество. Тогда множество всех $\omega \in E$ таких, что $T^n\omega \notin E$ для всех $n > 0$, имеет меру 0. Эквивалентно, для почти всех $\omega \in E$ будет $T^n\omega \in E$ для некоторого $n > 0$, т.е. траектория ω вновь посетит E. Более того, это происходит бесконечно много раз.

Данное утверждение является нетривиальным, когда мера множества E положительная.

Мы сформулируем алгоритмически эффективный аналог этого утверждения для равномерной меры на пространстве Ω всех бесконечных двоичных последовательностей и сдвига T на нем:

$$T(\omega_1\omega_2\dots) = \omega_2\omega_3\dots$$

Множество E называется эффективно замкнутым, если оно является дополнением для некоторого эффективно открытого множества.

Поточечный вариант теоремы Пуанкаре о возвращении представлен в виде следующей теоремы.

Теорема 8.2. *Пусть T – сдвиг на Ω и E – эффективно замкнутое множество положительной меры. Тогда для любой случайной по Мартин-Лефу последовательности $\omega \in E$ будет $T^n\omega \in E$ для некоторого $n > 0$. Более того, это верно для бесконечно многих n.*

Теорема 8.2 будет прямым следствием следующего утверждения, которое было впервые доказано Кучерой [45].

Предложение 8.1. *Пусть U – эффективно открытое множество равномерной меры $L(U) < 1$. Тогда для любой бесконечной последовательности ω, случайной по мере L, и произволь-*

ного натурального числа N найдется число $n \geqslant N$ такое, что $T^n\omega \notin U$.

Доказательство. Пусть N – произвольное натуральное число. Обозначим посредством U^* множество всех $\omega \in \Omega$ таких, что выполнено $T^n\omega \in U$ для всех $n \geqslant N$.

Пусть $L(U) < r$ для некоторого рационального $r < 1$. Представим эффективно открытое множество U в виде объединения вычислимой последовательности попарно непересекающихся интервалов

$$U = \cup_i \Gamma_{x_i},$$

где конечные последовательности x_i и x_j попарно не продолжают друг друга. Обозначим $U_1 = U$. Определим

$$U_2 = \cup_{i,j} \Gamma_{x_i x_j},$$
$$U_3 = \cup_{i,j,s} \Gamma_{x_i x_j x_s}$$

и т.д. Здесь $x_i x_j$ – конкатенация строк x_i и x_j. Аналогичным образом понимается $x_i x_j x_s$.

Имеем

$$L(U_2) = \sum_{i,j} L(\Gamma_{x_i x_j}) = \sum_{i,j} 2^{-l(x_i)-l(x_j)} =$$
$$= \sum_i 2^{-l(x_i)} \sum_j 2^{-l(x_j)} < r^2$$

и т.д. Аналогично имеем $L(U_n) < r^n$ для всех n.

Пусть $\omega \in U^*$. Обозначим $\omega' = \omega_{N+1}\omega_{N+2}\dots$

Тогда $\omega' = T^N\omega \in U$, значит, $\omega' = x_i\omega''$ для некоторого i и $\omega'' \in \Omega$. Так как $\omega'' = T^{N+l(x_i)}\omega \in U$, имеет место $\omega'' = x_j\omega'''$ для некоторого j и $\omega''' \in \Omega$. Теперь мы знаем, что $\omega' = x_i x_j \omega'''$, а значит, $\omega' \in U_2$. Аналогичным образом имеем $\omega' \in U_3$ и т.д.

Равномерно перечислимая система множеств

$$\{U_m : m = 1, 2, \dots\}$$

определяет некоторый тест Мартин-Лефа. Ранее мы доказали, что $\omega' \in \cap_m U_m$, т.е. ω' не случайная. Легко видеть, что в этом случае исходная последовательность ω также не случайная [3] . В частности, $U^* \subseteq \cap_m U_m$. Теорема доказана. $\Box$

Для доказательства теоремы 8.2 надо взять в предложении 8.1 $U = \Omega \setminus E$. Так как $L(E) > 0$, выполнено неравенство $L(U) < 1$. По предложению 8.1 $T^n\omega \notin U$ для бесконечно многих n. $\Box$

8.3. Отсутствие вычислимой оценки скорости сходимости в эргодической теореме

В этом разделе мы покажем, что в некоторых случаях не существует вычислимой оценки скорости сходимости средних в эргодической теореме.

Сначала мы дадим необходимые определения. В дальнейшем нам потребуются понятия различной степени алгоритмической эффективности функций, определенных на бесконечных последовательностях.

Функция $f : \Omega \to \mathcal{R} \cup \{-\infty, +\infty\}$ называется *перечислимой снизу*, если существуют вычислимые функции $r = r(i)$ и $x = x(i)$ такие, что неравенство $r < f(\omega)$ выполнено тогда и только тогда, когда $r = r(i)$ и $x(i) \subset \omega$ для некоторого i.

Можно также сказать, что для любого рационального r множество

$$U_r = \{\omega : r < f(\omega\}$$

является эффективно открытым и семейство $\{U_r, r \in \mathcal{Q}\}$ этих множеств является равномерно перечислимым семейством эффективно открытых множеств.

Другими словами, существует алгоритм, на вход которому подаются рациональное r и начальные фрагменты бесконечной последовательности ω. Данный алгоритм обладает следующим

[3] Это утверждение было предметом задачи из раздела 4.3.

свойством: если $r < f(\omega)$, то этот факт рано или поздно будет обнаружен этим алгоритмом, при этом алгоритм использует только некоторый начальный фрагмент последовательности ω; если $r \geqslant f(\omega)$, то такой алгоритм может работать бесконечно долго и не выдаст никакого ответа.

Заметим, что перечислимая снизу функция может принимать бесконечные значения.

Функция f *перечислима сверху*, если функция $-f$ перечислима снизу.

Функция f, принимающая рациональные значения, а также значения $-\infty$ и $+\infty$, называется *простой*, если множество Ω можно представить в виде объединения конечного числа интервалов так, что $f(\omega)$ постоянна на каждом из них. Простая функция описывается конечным набором конструктивных объектов, поэтому сама является конструктивным объектом.

Простые функции позволяют дать удобную характеризацию перечислимых снизу функций.

Предложение 8.2. *Для любой перечислимой снизу функции $f(\omega)$ существует вычислимая последовательность простых функций $f_n(\omega)$ такая, что*

- $f_n(\omega) \leqslant f_{n+1}(\omega)$ *для всех n и ω;*
- $f(\omega) = \lim\limits_{n\to\infty} f_n(\omega)$.

Доказательство. Пусть $x(i)$ и $r(i)$ – вычислимые функции из определения перечислимой снизу функции. Положим

$$f_n(\omega) = \sup\{r : r = r(i), x(i) \subset \omega, i \leqslant n\}.$$

Пусть $\sup \emptyset = -\infty$. Тогда

$$f(\omega) = \lim_{n\to\infty} f_n(\omega).$$

Предложение доказано. □

Функция $f(\omega)$ называется *вычислимой*, если она перечислима снизу и сверху. В этом случае существует алгоритм, который, используя в своей работе рациональное $\epsilon > 0$ и последовательность $\omega \in \Omega$, выдает рациональное приближение к $f(\omega)$ с точностью до ϵ. При этом для получения результата алгоритм использует только некоторой начальный фрагмент последовательности ω.

Сформулируем некоторые понятия конструктивной теории вероятностей. Задано вероятностное пространство $(\Omega, \mathcal{F}, P)$, где P – вычислимая вероятностная мера. Функция типа $f : \Omega \to \mathcal{R}$ будет называется *случайной функцией*.

Последовательность случайных функций $f_n(\omega)$ *сходится по вероятности* к случайной функции $f(\omega)$, если для всех $\epsilon > 0$ и $\delta > 0$ выполнено

$$P\{\omega : |f_n(\omega) - f(\omega)| > \delta\} < \epsilon \tag{8.1}$$

при всех достаточно больших n.

Сходимость по вероятности называется *алгоритмически эффективной*, если существует вычислимая функция $m(\epsilon, \delta)$, принимающая неотрицательные целые значения, такая, что для всех рациональных $\epsilon > 0$ и $\delta > 0$ выполнено (8.1) при всех $n \geqslant m(\epsilon, \delta)$. Эффективность заключается в том, что существует алгоритм, который по $\epsilon > 0$ и $\delta > 0$ вычисляет тот номер случайной функции, начиная с которого выполнено неравенство (8.1).

Функция $m(\epsilon, \delta)$ называется *регулятором сходимости*. Последовательность случайных функций $f_n(\omega)$ сходится по вероятности к случайной функции $f(\omega)$ алгоритмически эффективно, если для этой сходимости существует вычислимый регулятор сходимости.

Нетрудно проверить, что определение эффективной сходимости последовательности функций $f_n(\omega)$ по вероятности эквивалентно тому, что для любых рациональных $\epsilon > 0$ и $\delta > 0$ выполнено

$$P\{\omega : |f_n(\omega) - f_{n'}(\omega)| > \delta\} < \epsilon \tag{8.2}$$

при всех $n, n' \geqslant m(\epsilon, \delta)$.

Последовательность случайных функций $f_n(\omega)$ сходится к некоторой функции $f(\omega)$ *почти всюду*, если $\lim\limits_{n\to\infty} f_n(\omega) = f(\omega)$ для P-почти всех ω. Это определение эквивалентно тому, что существует функция $m(\epsilon, \delta)$, принимающая неотрицательные целые значения, такая, что для всех рациональных $\epsilon > 0$ и $\delta > 0$ выполнено

$$P\{\omega : \sup_{k \geqslant n} |f_k(\omega) - f(\omega)| > \delta\} < \epsilon$$

при всех $n \geqslant m(\epsilon, \delta)$ (см. [25]).

Легко видеть, что из сходимости почти всюду следует сходимость по вероятности.

Естественно называть сходимость почти всюду алгоритмически эффективной, если регулятор сходимости $m(\epsilon, \delta)$ является вычислимой функцией. Легко видеть, что если последовательность $f_n(\omega)$ сходится к некоторой функции почти всюду алгоритмически эффективно, то она сходится к ней и по вероятности алгоритмически эффективно.

Вычислимый регулятор для сходимости по вероятности в законе больших чисел для равномерной меры P строится с помощью неравенства Хефдинга:

$$P\left\{\omega : \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| > \delta\right\} \leqslant 2e^{-2n\delta^2},$$

где $S_n(\omega) = \sum\limits_{i=1}^{n} \omega_i$. В данном случае можно определить

$$m(\epsilon, \delta) = \lfloor \frac{1}{2\delta^2} \ln \frac{2}{\epsilon} \rfloor + 1.$$

Из этого неравенства также легко получить соответствующее неравенство для построения регулятора эффективной сходимости почти всюду.

Мы приведем пример вычислимой стационарной меры, для которой сходимость средних по вероятности и почти всюду в теореме Биркгофа не является алгоритмически эффективной.

Теорема 8.3. *Существует вычислимая стационарная мера P, для которой не существует вычислимого регулятора для сходимости средних по вероятности*

$$P\{\omega : |S_n(\omega) - f(\omega)| > \delta\} \to 0$$

при $n \to \infty$, где $S_n(\omega) = \frac{1}{n}\sum_{k=1}^{n} \omega_k$ и $f(\omega) = \lim_{n\to\infty} S_n(\omega)$ (этот предел существует P-почти всюду по эргодической теореме Биркгофа).

Доказательство. Мы построим необходимую вычислимую стационарную меру P в виде смеси однородных стационарных марковских мер P_i, $i = 1, 2, \ldots$. Каждая мера P_i будет вычислимой и будет содержать в себе информацию о проблеме остановки универсального алгоритма.

Пусть $U(i, \delta, \epsilon)$ – вычислимая функция, универсальная для всех вычислимых функций от двух аргументов $m(\delta, \epsilon)$.[4]

Для любой вычислимой функции $m(\delta, \epsilon)$ существует число i такое, что $m(\delta, \epsilon) = U(i, \delta, \epsilon)$ для всех δ и ϵ. Пусть также

$$U^s(i, \delta, \epsilon) = \begin{cases} U(i, \delta, \epsilon), \text{ если это значение было вычислено за} \\ \leqslant s \text{ шагов,} \\ \text{неопределено в противном случае.} \end{cases}$$

Для произвольного i определим действительное число α_i путем задания битов его двоичного разложения:

$$\alpha_i = 0.\alpha_{i1}\alpha_{i2}\ldots,$$

где

$$\alpha_{is} = \begin{cases} 1, \text{ если } u = U^s(i, \frac{1}{4}, 2^{-(i+1)}) \text{ определено и } s > u, \\ 0 \text{ в противном случае.} \end{cases}$$

Легко видеть, что значение каждого бита α_{is} алгоритмически вычислимо по i и s. Кроме того, $\alpha_i > 0$ тогда и только тогда,

[4]Напомним, что в данном случае мы отождествляем все положительных рациональные числа и все натуральные числа.

когда значение $U(i, \frac{1}{4}, 2^{-(i+1)})$ определено. Таким образом, действительное число α_i является индикатором проблемы остановки на входах $\delta = \frac{1}{4}$ и $\epsilon = 2^{-(i+1)}$.

По определению если $\alpha_i > 0$, то двоичное разложение числа α_i состоит из блока нулей, после которого идут единицы. В этом случае $\alpha_i = 2^{-k(i)}$, где $k(i)$ – длина начального блока из нулей.

Определим для произвольного i однородную марковскую цепь путем задания начальных вероятностей:

$$P_i\{\omega_1 = 0\} = P_i\{\omega_1 = 1\} = \frac{1}{2}$$

и переходных вероятностей:

$$P_i\{\omega_{s+1} = 0 | \omega_s = 1\} = P_i\{\omega_{s+1} = 1 | \omega_s = 0\} = \alpha_s$$

для произвольного $s = 1, 2, \ldots$

С помощью задачи 6 из раздела 8.6 легко показать, что вероятностная мера P_i, порожденная заданными начальными и переходными вероятностями, является стационарной. Кроме того, она является вычислимой.

Согласно теории из монографии [25] при $\alpha_i > 0$ эта мера также является эргодической. Если $\alpha_i = 0$, мера P_i сосредоточена только на двух бесконечных последовательностях: $P_i(0^\infty) = P_i(1^\infty) = \frac{1}{2}$. Множества $\{0^\infty\}$ и $\{1^\infty\}$ являются инвариантными относительно сдвига. Поэтому в случае $\alpha_i = 0$ мера P_i не является эргодической.

Каждая мера P_i является вычислимой, и, более того, существует алгоритм, который равномерно по i и x вычисляет значение $P_i(x)$. Определим меру

$$P(x) = \sum_{i=1}^{\infty} 2^{-i} P_i(x).$$

Нетрудно доказать, что мера P является вычислимой. Так как каждая мера P_i является стационарной, мера P также является

стационарной. Из определения видно, что эта мера не является эргодической.

По эргодической теореме Биркгофа, примененной для сдвига, для P-почти всех ω существует предел $\lim\limits_{n\to\infty} S_n(\omega)$ при $n \to \infty$, где $S_n(\omega) = \frac{1}{n}\sum\limits_{s=1}^{n} \omega_i$.

Пусть $m(\delta, \epsilon)$ – произвольная всюду определенная вычислимая функция – кандидат на регулятор сходимости средних по вероятности для меры P. Тогда существует i такое, что $m(\delta, \epsilon) = U(i, \delta, \epsilon)$ для всех δ, ϵ. В этом случае $\alpha_i > 0$.

По эргодической теореме для марковских процессов стационарное распределение для марковского процесса, порожденного мерой P_i при $\alpha_i > 0$, есть $\pi_0 = \frac{1}{2}$ и $\pi_1 = \frac{1}{2}$. Для этого распределения выполнен закон больших чисел. В частности,

$$P_i\{\omega : |S_n(\omega) - \frac{1}{2}| < 0.01\} \to 1 \tag{8.3}$$

при $n \to \infty$.

По определению число $k(i)$ равно номеру позиции последнего нуля в двоичном представлении числа α_i, после которого в этом представлении стоят единицы. Легко видеть, что $\alpha_i = 2^{-k(i)}$.

Оценим вероятности:

$$P_i(0^{k(i)}) = P_i(1^{k(i)}) = \frac{1}{2}(1 - \alpha_i)^{k(i)-1} > \frac{2}{5}$$

для всех достаточно больших значений $k(i)$ [5] . Следовательно,

$$P_i\{\omega : S_{k(i)}(\omega) = 0 \text{ или } 1\} > \frac{4}{5}.$$

По определению $k(i) > m(\frac{1}{4}, 2^{-(i+1)})$. Отсюда и из (8.3) следует, что найдется достаточно большое $n > m(\frac{1}{4}, 2^{-(i+1)})$, для которого

$$P_i\{\omega : |S_{k(i)}(\omega) - S_n(\omega)| > \frac{1}{4}\} > \frac{1}{2}.$$

[5] Без потери общности можно предположить, что все шаги s, на которых впервые определилось какое-либо значение универсальной функции, больше некоторого фиксированного значения s_0.

Поэтому P-мера этого множества больше $2^{-i} \cdot \frac{1}{2} = 2^{-(i+1)} = \epsilon$, т.е. числа $k(i)$ и n не удовлетворяют условию (8.2) для регулятора сходимости.

Полученное противоречие доказывает теорему. □

Поскольку из алгоритмически эффективной сходимости почти всюду следует алгоритмически эффективная сходимости по вероятности, получаем следующее следствие из теоремы 8.3.

Следствие 8.1. *Существует вычислимая стационарная мера P, для которой не существует вычислимого регулятора для сходимости средних почти всюду*

$$P\{\omega : \sup_{k \geqslant n} |S_k(\omega) - f(\omega)| > \delta\} \to 0$$

при $n \to \infty$.

8.4. Интегральные тесты случайности

Для дальнейшего изложения нам потребуется еще один вид тестов случайности по Мартин-Лефу – интегральные тесты.

Пусть P – вычислимая мера на Ω. Перечислимая снизу функция $f : \Omega \to \mathcal{R}_+ \cup \{+\infty\}$ называется *интегральным тестом случайности* относительно меры P или интегральным P-тестом, если

$$E_P(f) = \int f(\omega) dP \leqslant 1.$$

Здесь E_P – символ математического ожидания.

Из определения следует, что $f(\omega) < \infty$ для P-почти всех ω.

Интегральные тесты были впервые введены Левиным; в современной форме они впервые изучались в работе Гача [42].

Из определения для любого интегрального теста выполнено неравенство Маркова:

$$P\{\omega : f(\omega) > r\} < \frac{1}{r}$$

для любого r. В частности, $f(\omega) < \infty$ для P-почти всех ω.

Имеет место теорема о существовании максимального с точностью до мультипликативной константы интегрального теста.

Теорема 8.4. *Пусть P – вычислимая мера. Существует интегральный P-тест $p(\omega)$ такой, что для любого интегрального P-теста $f(\omega)$ существует константа c такая, что $cp(\omega) \geqslant f(\omega)$ для всех ω.*

Доказательство этой теоремы использует лемму о возможности построить равномерно эффективно перечислимую последовательность всех перечислимых снизу интегральных тестов.

Лемма 8.1. *Существует такая последовательность функций $p_i(\omega)$, что*

- *множество $\{\omega \in \Omega : r < p_i(\omega)\}$ является равномерно (относительно i и r) перечислимым семейством эффективно открытых множеств;*
- *для любого интегрального теста $f(\omega)$ существует i такое, что $f(\omega) = p_i(\omega)$ для всех ω.*

Доказательство этой леммы аналогично доказательству подобных утверждений и предоставляется читателю в качестве упражнения.

Последовательность функций $p_i(\omega)$, для которой выполнено первое из условий, приведенных выше, называется *равномерно перечислимой снизу.*

Доказательство теоремы. Определим

$$p(\omega) = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} p_n(\omega).$$

Легко видеть, что по теореме Лебега

$$\int p(\omega) dP = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \int p_n(\omega) dP \leqslant \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$$

Функция $p(\omega)$ перечислима снизу. $\square$

Фиксируем один из интегральных тестов, удовлетворяющих теореме 8.4, обозначим его $\hat{p}(\omega)$ и назовем *универсальным интегральным P-тестом.*

С помощью интегральных тестов можно дать эквивалентное определение случайной по Мартин-Лефу последовательности.

Теорема 8.5. *Пусть P – вычислимая мера и $\hat{p}(\omega)$ – универсальный интегральный P-тест. Бесконечная двоичная последовательность ω является случайной по Мартин-Лефу относительно меры P тогда и только тогда, когда $\hat{p}(\omega) < \infty$.*

Доказательство. Полагаем

$$U_m = \{\omega : \hat{p}(\omega) > 2^m\}$$

для любого m. По неравенству Маркова $P(U_m) \leqslant 2^{-m}$ для всех m. Кроме того, из перечислимости интегрального теста снизу следует, что множество U_m является эффективно открытым.

Если $\hat{p}(\omega) = \infty$, то $\omega \in \cap_n U_n$.

Докажем обратное утверждения. Пусть $\{U_m\}$ – произвольный тест Мартин-Лефа. Рассмотрим последовательность характеристических функций

$$p_m(\omega) = \begin{cases} 1, \text{ если } \omega \in U_m, \\ 0 \text{ в противном случае.} \end{cases}$$

Так как $\{U_m\}$ – это равномерно перечислимая последовательность эффективно открытых множеств, последовательность функций $p_m(\omega)$ является равномерно перечислимой снизу. Определим

$$p(\omega) = \sum_{m=1}^{\infty} p_m(\omega).$$

Функция $p(\omega)$ перечислима снизу и

$$\int p(\omega) dP = \sum_{m=1}^{\infty} \int p_m(\omega) dP = \sum_{m=1}^{\infty} P(U_m) \leqslant \sum_{m=1}^{\infty} 2^{-m} = 1.$$

Значит, функция $p(\omega)$ является интегральным P-тестом.

Если $\omega \in \cap_m U_m$, то по определению $p(\omega) = \infty$. Отсюда следует, что $\hat{p}(\omega) = \infty$. Теорема доказана. □

8.5. Эффективная эргодическая теорема

Эффективная эргодическая теорема будет рассматриваться для вычислимой меры, вычислимого преобразования и для вычислимой наблюдаемой.

Преобразование T называется вычислимым, если оно совпадает с некоторой вычислимой операцией.

Формулировка эргодической теоремы Биркгофа для случайных последовательностей получается из оригинальной формулировки заменой выражения «для P-почти всех» на «для случайных по мере P».

Теорема 8.6. *Пусть P – произвольная вычислимая мера на Ω и f – произвольная вычислимая интегрируемая функция типа $\Omega \to \mathcal{R}$. Тогда для любого сохраняющего меру P вычислимого преобразования T и для любой случайной по Мартин-Лефу последовательности ω выполнено*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = \tilde{f}(\omega), \tag{8.4}$$

где $\tilde{f}$ – интегрируемая инвариантная относительно T функция такая, что $\int \tilde{f}(\omega)dP = \int f(\omega)dP$.

Кроме того, если преобразование T является эргодическим, то $\tilde{f}(\omega) = \int f(\alpha)dP$ для такой ω.

Доказательство. Для произвольной бесконечной последовательности ω обозначим среднее значение

$$s_m(\omega) = \frac{1}{m+1} \sum_{k=0}^{m} f(T^k\omega).$$

Для удобства в дальнейших рассуждениях считаем, что выполнено $s_{-1}(\omega) = 0$.

Пусть $\int |f(\omega)|dP \leqslant M$, где M – положительное целое число.

Если предел $\lim\limits_{m\to\infty} s_m(\omega)$ не существует, то найдутся два рациональных числа $\alpha < \beta$ такие, что $-M < \alpha < \beta < M$ и

$$\liminf_{m\to\infty} s_m(\omega) < \alpha < \beta < \limsup_{m\to\infty} s_m(\omega).$$

Обратное утверждение также верно.

Пусть α и β – рациональные числа такие, что

$$-M < \alpha < \beta < M.$$

Определим функцию пересечения границ $\sigma_n(\omega|\alpha,\beta)$ следующим образом.

Значение $\sigma_n(\omega|\alpha,\beta)$ равно числу пересечений снизу вверх интервала (α,β) последовательностью $s_0(\omega), s_1(\omega), \ \ldots, \ s_n(\omega)$. Точнее, определим

$$\begin{aligned}
u_0 &= 0,\\
u_1 &= \min\{m : m \geqslant u_0, s_m(\omega) < \alpha\},\\
v_1 &= \min\{m : m > u_1, s_m(\omega) > \beta\},\\
&\ldots\\
u_i &= \min\{m : m > v_{i-1}, s_m(\omega) < \alpha\},\\
v_i &= \min\{m : m > u_i, s_m(\omega) > \beta\},\\
&\ldots\\
u_k &= \min\{m : m > v_{k-1}, s_m(\omega) < \alpha\},\\
v_k &= \min\{m : m > u_k, s_m(\omega) > \beta\}.
\end{aligned}$$

Определим функцию

$$\sigma_n(\omega|\alpha,\beta) = \begin{cases} 0, \text{ если } v_1 > n,\\ \max\{k : v_k \leqslant n\}, \text{ если } v_1 \leqslant n. \end{cases}$$

Значение функции $\sigma_n(\omega|\alpha,\beta)$ равно максимальному числу пересечений снизу вверх интервала (α,β) средними $s_m(\omega)$ при

$m = 0, 1, \ldots n$. Функция $\sigma_n(\omega|\alpha, \beta)$ является перечислимой снизу равномерно относительно параметров n, α и β.

Легко видеть, что предел $\lim\limits_{m\to\infty} s_m(\omega)$ не существует тогда и только тогда, когда $\sup_n \sigma_n(\omega|\alpha, \beta) = \infty$ для некоторых $\alpha < \beta$.

Временно фиксируем бесконечную последовательность ω, а также натуральное число n и рациональные числа α и β такие, что $\alpha < \beta$.

Введем безотносительные отклонения

$$a(u, \omega) = \sum_{s=0}^{u} (f(T^s\omega) - \alpha),$$
$$b(v, \omega) = \sum_{s=0}^{v} (f(T^s\omega) - \beta).$$

Нам будет удобно считать, что $a(-1, \omega) = 0$.

В дальнейшем будет использоваться следующее свойство: из $s_u(\omega) < \alpha$ и $s_v(\omega) > \beta$ следует, что $a(u, \omega) < b(v, \omega)$.

Осцилляция относительных частот влечет осцилляцию безотносительных величин отклонений.

Последовательность $d = \{u_1, v_1, \ \ldots, \ u_k, v_k\}$ целых чисел называется допустимой, если

$$-1 \leqslant u_1 < v_1 \leqslant u_2 < v_2 \leqslant \ldots \leqslant u_k < v_k \leqslant n.$$

Число пар в допустимой последовательности d обозначаем m_d ($m_d = k$) и назовем ее длиной.

Для каждой допустимой последовательности

$$d = \{s_1, t_1, \ldots, s_k, t_k\}$$

рассмотрим кумулятивную сумму разностей безотносительных отклонений:

$$S(d, \omega) = \sum_{j=1}^{k} (b(t_j, \omega) - a(s_j, \omega)).$$

Ключевую роль в доказательстве теоремы играет следующая комбинаторная лемма об удлинении допустимой последовательности без уменьшении кумулятивной суммы.

Лемма 8.2. *Для каждой допустимой последовательности q существует допустимая последовательность d длины не меньше, чем максимальное число пересечений интервала (α, β), иными словами $m_d \geqslant \sigma_n(\omega|\alpha, \beta)$, и такая, что $S(d, \omega) \geqslant S(q, \omega)$.*

Доказательство. Обозначим $N = \sigma_n(\omega|\alpha, \beta)$ – максимальное число пересечений интервала (α, β) последовательностью средних $s_0(\omega), \ldots, s_n(\omega)$. Пусть

$$p = \{-1 < u_1 < v_1 < u_2 < v_2 < \cdots < u_N < v_N \leqslant n\}$$

есть та допустимая последовательность длины N, по которой было определено значение функции $N = \sigma_n(\omega|\alpha, \beta)$.

Достаточно доказать, что для произвольной допустимой последовательности q с длиной $m_q < N$ существует допустимая последовательность d, для которой $m_d = m_q + 1$ и $S(d, \omega) \geqslant S(q, \omega)$.

Пусть допустимая последовательность q имеет вид

$$-1 \leqslant s_1 < t_1 \leqslant s_2 < t_2 \leqslant \ldots \leqslant s_m < t_m \leqslant n,$$

где $m = m_q$.

Расширим ее на одну пару элементов. Введем вспомогательный элемент $s_{m+1} = n$. Так как $m + 1 \leqslant N$, v_{m+1} присутствует в последовательности p. Кроме того, $v_{m+1} \leqslant n = s_{m+1}$. Следовательно, существует наименьшее i такое, что $v_i \leqslant s_i$. Если $i = 1$, то положим

$$d = \{u_1, v_1, s_1, t_1, \ \ldots, \ s_m, t_m\}. \tag{8.5}$$

Длина допустимой последовательности увеличилась на единицу.

Рассмотрим случай $i > 1$. Тогда $v_{i-1} > s_{i-1}$, и мы имеем неравенство

$$s_{i-1} < v_{i-1} < u_i < v_i \leqslant s_i.$$

Если $u_i < t_{i-1}$, то положим

$$d = \{s_1, t_1, \ \ldots, \ s_{i-1}, v_{i-1}, u_i, t_{i-1}, \ \ldots, \ s_m, t_m\}. \tag{8.6}$$

Если $u_i \geqslant t_{i-1}$, то положим при $i \leqslant m$

$$d = \{s_1, t_1, \ldots \ , s_{i-1}, t_{i-1}, u_i, v_i, s_i, t_i, \ldots \ , s_m, t_m\}, \tag{8.7}$$

и полагаем при $i = m + 1$

$$d = \{s_1, t_1, \ \ldots, \ s_m, t_m, s_{m+1}, t_{m+1}\}. \tag{8.8}$$

Построенная последовательность d допустима, и ее длина увеличилась на единицу: $m_d = m_q + 1$. Остается проверить, как изменились кумулятивные суммы для различных вариантов определения последовательности d.

При определениях (8.5), (8.7) и (8.8)

$$S(\omega, d) = S(\omega, q) + b(v_i, \omega) - a(u_i, \omega).$$

При определении (8.6)

$$S(\omega, d) = S(\omega, q) + b(v_{i-1}, \omega) - a(u_i, \omega).$$

По определению последовательности $\{u_1, v_1, \ \ldots, \ u_N, v_N\}$ прибавленные члены положительны. Значит, в обоих случаях кумулятивные суммы увеличиваются:

$$S(\omega, d) > S(\omega, q).$$

Лемма доказана. □

Пусть $d = \{s_1, t_1, \ \ldots, \ s_m, t_m\}$ – допустимая последовательность длиной $m_d = m$ и $S(\omega, d)$ – соответствующая кумулятивная сумма.

Применим преобразование T к последовательности ω и посмотрим, как при этом изменится кумулятивная сумма. Во-первых, при $s_i \geqslant 0$ происходят следующие изменения:

$$\begin{aligned} a(s_i, \omega) &= a(s_i - 1, T\omega) + f(\omega) - \alpha, \\ b(t_i, \omega) &= b(t_i - 1, T\omega) + f(\omega) - \beta. \end{aligned}$$

Отсюда и из определения кумулятивной суммы получаем

$$S(\omega, d) = S(T\omega, d') + a - (\beta - \alpha)m_d, \tag{8.9}$$

где

$$d' = \{s_1 - 1, t_1 - 1, \ \ldots, \ s_m - 1, t_m - 1\},$$

если $s_1 \geqslant 0$, и

$$d' = \{-1, t_1 - 1, s_2 - 1, t_2 - 1, \ \ldots, \ s_m - 1, t_m - 1\},$$

если $s_1 = -1$ и $t_1 > 0$. Если же $s_1 = -1$ и $t_1 = 0$, то

$$d' = \{s_2 - 1, t_2 - 1, \ \ldots, \ s_m - 1, t_m - 1\}.$$

В сумме (8.9) $a = 0$, если $s_1 \geqslant 0$, и $a = f(\omega) - \alpha$, если $s_1 = -1$.

Введем перечислимую снизу функцию:

$$\lambda_n(\omega) = \sup\{S(\omega, d) : d \text{ – допустимая последовательность}\}.$$

Тогда из (8.9) следует, что

$$S(\omega, d) \leqslant \lambda_n(T\omega) + (f(\omega) - \alpha)^+ - (\beta - \alpha)m_d, \tag{8.10}$$

где использовано обозначение $h^+ = \max\{h, 0\}$.

По лемме 8.2 для каждой допустимой последовательности q существует такая допустимая последовательность d, что выполнено $m_d \geqslant \sigma_n(\omega|\alpha, \beta)$ и $S(\omega, q) < S(\omega, d)$. Отсюда и из (8.10) имеем

$$\begin{aligned} S(\omega, q) < S(\omega, d) \leqslant \\ \leqslant \lambda_n(T\omega) + (f(\omega) - \alpha)^+ - (\beta - \alpha)\sigma_n(\omega|\alpha, \beta), \end{aligned} \tag{8.11}$$

Берем в (8.11) максимум по q и получаем

$$\lambda_n(\omega) \leqslant \lambda_n(T\omega) + (f(\omega) - \alpha)^+ - (\beta - \alpha)\sigma_n(\omega|\alpha, \beta).$$

Следовательно,

$$(\beta - \alpha)\sigma_n(\omega|\alpha, \beta) \leqslant (f(\omega) - \alpha)^+ + \lambda_n(T\omega) - \lambda_n(\omega). \tag{8.12}$$

Интегрируя неравенство (8.12), получим

$$\int (\beta - \alpha)\sigma_n(\omega|\alpha, \beta)dP \leqslant \int (f(\omega) - \alpha)^+ dP. \tag{8.13}$$

Здесь мы впервые используем предположение о том, что преобразование T сохраняет меру. Из этого предположения следует, что

$$\int \lambda_n(T\omega)dP = \int \lambda_n(\omega)dP.$$

Поскольку интеграл от функции $|f(\omega)|$ ограничен числом M,

$$\int (f(\omega) - \alpha)^+ dP \leqslant 2M.$$

Полагаем

$$\sigma(\omega|\alpha, \beta) = \sup_n \sigma_n(\omega|\alpha, \beta).$$

Легко видеть, что эта функция перечислима снизу. Кроме этого, из того, что

$$\sigma_n(\omega|\alpha, \beta) \leqslant \sigma_{n+1}(\omega|\alpha, \beta)$$

для всех n, эта функция является интегрируемой, и по (8.13) получаем

$$\int (2M)^{-1}(\beta - \alpha)\sigma(\omega|\alpha, \beta)dP \leqslant 1$$

для всех $\alpha < \beta$.

Путем усреднения величины $\sigma(\omega|\alpha, \beta)$ мы можем определить интегральный тест случайности. Пусть вычислимые функции $\alpha(i)$ и $\beta(i)$ перечисляют множество всех пар рациональных чисел

$$\{(\alpha, \beta) : -M < \alpha < \beta < M\}.$$

Определим

$$p(\omega) = \frac{1}{2M} \sum_{i=1}^{\infty} \frac{1}{i(i+1)} (\beta(i) - \alpha(i))\sigma(\omega|\alpha(i), \beta(i)).$$

По своему определению функция $p(\omega)$ перечислима снизу и

$$\int p(\omega)dP \leqslant 1,$$

т.е. она является интегральным тестом случайности, корректным относительно меры P. Кроме того, как ранее было замечено, если предел средних $\lim\limits_{n\to\infty} s_n(\omega)$ не существует, то найдутся рациональные $\alpha < \beta$ такие, что $\sigma(\omega|\alpha,\beta) = \infty$.

Отсюда следует, что для любой бесконечной двоичной последовательности ω верна импликация:

$$p(\omega) < \infty \Rightarrow \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) \text{ существует.}$$

Основная часть теоремы доказана.

Пусть $\tilde{f}(\omega)$ обозначает предел средних (8.4). Легко видеть, что $\tilde{f}(T\omega) = \tilde{f}(\omega)$ для всех ω. Если преобразование T – эргодическое, то $\tilde{f}(\omega) = c$ для P-почти всех ω, где $c = \int f(\omega)dP$ – константа.

Нам надо доказать следующее утверждение.

Лемма 8.3. *$\tilde{f}(\omega) = \int f(\omega)dP$ для любой последовательности ω случайной по мере P.*

Доказательство. Допустим, что это не так – существует случайная по мере P последовательность ω, для которой выполнено $\tilde{f}(\omega) = d \neq c$. Выберем рациональные числа r_1 и r_2 такие, что $r_1 < d < r_2$ и $c \leqslant r_1$ или $c \geqslant r_2$, и определим

$$S_n = \{\alpha : r_1 < s_n(\alpha) < r_2\},$$
$$\bar{S}_n = \{\alpha : r_1 \leqslant s_n(\alpha) \leqslant r_2\}.$$

Так как предел (8.4) равен c почти всюду, выполнено $P(\bar{S}_n) \to 0$ при $n \to \infty$. Функция $P(\bar{S}_n)$ перечислима сверху (как функция от n), так как

$$r > P(\bar{S}_n) \Leftrightarrow 1 - r < P\{\alpha : r_1 > s_n(\omega) \text{ или } r_1 < s_n(\omega)\}.$$

Поэтому мы можем по произвольному m алгоритмически эффективно найти $n \geqslant m$, такое, что $P(\bar{S}_n) < 2^{-m}$.

По определению множество S_n эффективно открытое. Для него $P(S_n) \leqslant P(\bar{S}_n) < 2^{-m}$. Полагаем $U_m = S_n$ для такого n. Семейство множеств $\{U_m\}$ представляет собой тест Мартин-Лефа, корректный относительно меры P.

Последовательность $\omega \in \cap U_m$, т.е. не является случайной. Полученное утверждение доказывает лемму и теорему 8.6. $\square$

8.6. Задачи и упражнения

1. Привести полное доказательство леммы 8.1.

2. Привести нижние оценки для универсального теста $p(\omega)$.

3. Доказать, что множество Ω является компактом в топологии, порожденной интервалами Γ_x.

4. Доказать, что всякая вычислимая функция $f(\omega)$ является непрерывной в топологии, порожденной интервалами Γ_x.

5. Доказать, что если перечислимая снизу функция $f(\omega)$ является интегрируемой, то $f(\omega) < \infty$ для любой случайной последовательности ω.

6. Доказать, что вероятностная мера P на двоичных последовательностях является стационарной тогда и только тогда, когда выполнены условия $P(0x) + P(1x) = P(x)$ для всех x.

7. Доказать, что смесь стационарных мер P_i

$$P(x) = \sum_{i=1}^{\infty} \frac{1}{i(i+1)} P_i(x)$$

также является стационарной мерой.

8. Многие законы теории вероятностей выполнены не только для алгоритмически случайных последовательностей, но и при более общих предположениях.

Проверить, что усиленный закон больших чисел выполнен для любой бесконечной последовательности ω такой, что выполнено $\mathrm{K}(\omega^n) \geqslant n - \alpha(n)$, где $\alpha(n) = o(n)$ при $n \to \infty$. Закон повторного логарифма верен при $\mathrm{KM}(\omega^n) \geqslant n - \alpha(n)$, где

$\alpha(n) = o(\log\log n)$ при $n \to \infty$. Провести анализ доказательства этих законов.

Глава 9

Случайные по Колмогорову конечные последовательности

А.Н.Колмогоров придавал большое значение изучению свойств конечных последовательностей. В этой главе мы рассмотрим стохастические свойства конечных последовательностей.

9.1. (α, β)-нестохастические по Колмогорову конечные последовательности

Понятие дефекта случайности конечной последовательности x относительно конечного множества D, где $x \in D$, было определено в разделе 3.3:

$$d(x|D) = \log |D| - \mathrm{K}(x|D),$$

где $\mathrm{K}(x)$ – простая колмогоровская сложность конечной последовательности x.

Последовательность x называется (α, β)-стохастической, если $d(x|D) \leqslant \beta$ (или $\mathrm{K}(x|D) \geqslant \log |D| - \beta$) для некоторого конечного

множества D такого, что $x \in D$ и $K(D) \leqslant \alpha$. Соответственно, посседовательность x называется (α, β)-нестохастической, если $d(x|D) \geqslant \beta$ для любого конечного множества D такого, что $x \in D$ и $K(D) \leqslant \alpha$.

Эти понятия были предложены Колмогоровым, определение Колмогорова изложено в работе [24].

Для произвольного конечного множества D и числа $\beta \geqslant 1$ число всех $x \in D$, для которых $d(x|D) > \beta$ (или $\mathrm{K}(x|D) < \log|D| - \beta$) не превосходит $2^{-\beta+1}|D|$; для не менее чем $(1 - 2^{-\beta+1})|D|$ последовательностей $x \in D$ будет $d(x|D) \leqslant \beta$. Таким образом, при $\mathrm{K}(D) \leqslant \alpha$ не менее чем $(1 - 2^{-\beta+1})|D|$ последовательностей являются (α, β)-стохастическими.

Покажем, что (α, β)-нестохастические последовательности существуют.

Предложение 9.1. *Для каждого n и для любых неотрицательных натуральных чисел α и β таких, что $2\alpha + \beta \leqslant n - 4\log n - c$, где c – положительная константа. существует (α, β)-нестохастическая последовательность длины n.*

Доказательство. Пусть $D_1, \ldots, D_k$ – все конечные множества, префиксная сложоть которых не превосходит α. Тогда $k \leqslant 2^{\alpha+1}$. Для того, чтобы задать все эти множества достаточно знать число α и номер i множества D_i, для задания которого оптимальному алгоритму требуется больше всего времени. Грубая оценка соответствующго кода не превосходит $\alpha + 4\log\alpha + O(1)$,

Пусть x есть наименьшая (в лексиграфическом порядке) последовательность длины n такая, что $x \notin \cup\{D_i : |D_i| < 2^{n-\alpha-1}\}$. Для задания x достаточно использовать α, $i \leqslant \alpha$ и n, поэтому

$$\mathrm{K}(x) \leqslant \alpha + 2\log\alpha + 2\log n + O(1). \tag{9.1}$$

Допустим, что для некоторого D такого, что $\mathrm{K}(D) \leqslant \alpha$ и $x \in D$ выполнено $\mathrm{K}(x|D) \geqslant \log|D| - \beta$. Так как $x \in D$, будет $D = D_i$ для некоторого $1 \leqslant i \leqslant k$ и $|D_i| \geqslant 2^{n-\alpha-1}$. Отсюда $\mathrm{KP}(x|D) \geqslant n - \alpha - 1$. Учитывая неравенство (9.1) и неравенство $\mathrm{K}(x) \geqslant$

$\mathrm{K}(x|D) - O(1)$, получим

$$\alpha + 2\log\alpha + 2\log n + O(1) \geqslant n - \alpha - \beta - O(1).$$

В частности,

$$2\alpha + \beta \geqslant n - 2\log\alpha - 2\log n - O(1).$$

Из условия предположения следует, что $\alpha \leqslant n$. Следовательно, при $2\alpha + \beta < n - 4\log n - O(1)$ последовательность x является (α, β)-нестохастической. $\square$

В последующих разделах мы рассморим различные модификации и обобщения понятия (α, β)-нестохастической последовательности.

9.2. Минимальная достаточная статистика

Будем рассматривать префиксную сложность конечных объектов. При использовании префиксной сложности основные неравенства имеют место с точностью до аддитивной константы. Простую колмогоровскую сложность также можно использовать в дальнейших определениях, но в этом случае неравенства будут выполнены с логарифмической точностью.

Пусть D – конечное множество бинарных последовательностей и $x \in D$. Представим описание x в виде кода, состоящего из двух частей:

$$\mathrm{KP}(x) \leqslant \mathrm{KP}(x|D) + \mathrm{KP}(D) + O(1). \tag{9.2}$$

Неравенство (9.2) следует из неравенства 5.1

Используя множество D, заданное в виде списка его элементов, можно к качестве программы для произвольного $x \in D$ использовать бинарную запись порядкового номера этого элемента в списке. Отсюда $\mathrm{KP}(x|D) \leqslant \log|D| + O(1)$.[1] В результате получаем оценку (9.2) сложности последовательности x.

$$\mathrm{KP}(x) \leqslant \log|D| + \mathrm{KP}(D) + O(1). \tag{9.3}$$

[1]Пусть $D = \{x_1, \ldots, x_k\} \subseteq \Xi_n$. Числа $1 \leqslant i \leqslant k$ могут быть закодированы

Величина $\log|D| + \mathrm{KP}(D) + O(1)$ равна верхней границе длины некоторого кода для x, состоящего из двух частей – порядкового номера последовательности x внутри множества D, состоящего из "равновозможных" элементов (модели) и кода для задания этой множества.

Программа, задающая множество D, называется достаточной статистикой.

Согласно принципу "бритва Оккама"[2], предпочтительными являются простые модели, поэтому сложность $\mathrm{KP}(D)$ модели надо выбирать как можно меньше, например, порядка $O(\log l(x))$.

Можно изменять баланс между двумя частями двухчастичного кода. Два случая:

- При $D = \{x\}$, $\log|D| = 0$ и $\mathrm{KP}(D) = \mathrm{KP}(x) + O(1)$.
- При $x \in D = \Xi_n$, $\log|D| = n$ и $\mathrm{KP}(D) = O(\log n)$;

Точность оценки (9.2) измеряется величиной

$$\delta(x|D) = \log|D| - \mathrm{KP}(x) + \mathrm{KP}(D), \tag{9.4}$$

которая называется дефектом оптимальности (см. [27]).

Ранее, в разделе 3.3 по формуле (3.16) было определено понятие дефекта случайности конечной последовательности x относительно конечного множества D, содержащего x. Сформулируем аналогичное понятие с использованием префиксной сложности:

$$d(x|D) = \log|D| - \mathrm{KP}(x|D). \tag{9.5}$$

в виде двоичных последовательностей одинаковой длины $\lceil\log|D|\rceil$. Функция

$$B(i, D) = \begin{cases} x_i, \text{ если } 1 \leqslant i \leqslant k \\ \text{неопределено, в противном случае} \end{cases}$$

определяет (условный) префиксно-корректный способ задания конечных последовательностей, при котором $\mathrm{KP}_B(x|D) \leqslant \log|D| + 1$ при $x \in D$.

[2]Этот принцип кратко формулируется как "сущности не следует умножать без необходимости".

Предложение 9.2. *Дефект случайности произвольной конечной последовательности x относительно конечного множества D не превосходит дефект оптимальности с точностью до константы:*

$$d(x|D) \leqslant \delta(x|D) + O(1).$$

Доказательство. Согласно определениям (9.5) и (9.4) нам нужно доказать, что

$$\log|D| - \mathrm{KP}(x|D) \leqslant \log|D| - \mathrm{KP}(x) + \mathrm{KP}(D) + O(1).$$

Это неравенство эквивалентно неравенству

$$\mathrm{KP}(x) \leqslant \mathrm{KP}(x|D) + \mathrm{KP}(D) + O(1),$$

которое следует из неравенства 5.1. □

Как было замечено выше, величина $\delta(x|D)$ характеризует точность оценки (9.3). Входящая в нее величина $\mathrm{KP}(x)$ – постоянная, поэтому для повышения точности этой оценки необходимо минимизировать величину $\log|D| + \mathrm{KP}(D)$. В связи с этим, рассматривается структурная функция Колмогорова

$$\delta_x(\alpha) = \min\{\log|D| : x \in D \& \mathrm{KP}(D) \leqslant \alpha\}, \qquad (9.6)$$

где α – натуральное число. Программа, задающая множество D, на котором достигается минимум в (9.6), называтся колмогоровской достаточной статистикой (или минимальной достаточной статистикой).[3]

Для каждого x имеет смысл рассматриват функцию $\delta_x(\alpha)$ при $\alpha \geqslant c_1$, где c_1 – положительная константа (зависящая от x) такая, что множество из правой части определения (9.6) являются непустым.

[3] В [39] было отмечено, что А.Н.Колмогоров предложил в 1973 на Таллинской конференции по теории информации вариант этой функции и поставил проблему изучения возможных форм кривых (9.6). См. также [27].

Со статистической точки зрения удобно рассматривать аналогичную функцию, основанную на минимизации дефекта случайности:

$$\beta_x(\alpha) = \min\{d(x|D) : x \in D \& \mathrm{KP}(D) \leqslant \alpha\}, \tag{9.7}$$

где α – натуральное число.[4] Очевидно $\beta_x(\alpha) \geqslant -c$, где c – константа.

Из определения следует, что

$$\delta_x(\alpha) \leqslant \beta_x(\alpha) + \alpha. \tag{9.8}$$

Функции $\delta_x(\alpha)$ и $\beta_x(\alpha)$ характеризует возможность описания слова x с помощью статистической модели, состоящей из множеств сложности $\leqslant \alpha$. Для любого x функции $\delta_x(\alpha)$ и $\beta_x(\alpha)$ не возрастают по α и $\delta_x(\alpha) > 0$ при $\alpha \geqslant c_1$ и $\beta_x(\alpha) \geqslant -c$ для всех α, где c – неотрицательная константа.

Нижняя и верхняя граничные прямые для функции $\delta_x(\alpha)$ указаны в следующем утверждении.

Предложение 9.3. *Существует такие положительные константы c_1 и c_2, что для любой конечной последовательности x и для всех $\alpha \geqslant c_1$ и $k \geqslant 0$ выполнено*

$$\delta_x(\alpha) \geqslant \mathrm{KP}(x) - \alpha - O(1),. \tag{9.9}$$

$$\delta_x(\alpha + k) \leqslant \delta_x(\alpha) - k + O(1). \tag{9.10}$$

Неравенства (9.9)–(9.10) указывают, что график функции $\delta_x(\alpha)$ изображается кривой, которая начинается слева приблизительно (с точностью до констант) в точке (c_1, c_2) и расположена между двумя прямыми $y(\alpha) = \mathrm{KP}(x) - \alpha$ и $y(\alpha) = c_2 - \alpha$.

Доказательство. Докажем (9.9). Для любого конечного множества D такого, что $x \in D$, имеет место неравенство $\mathrm{KP}(x) \leqslant$

[4] Функция $\beta_x(\alpha)$ впервые была введена и изучалась в работе [6]. См. также [39] и [27].

$\log|D| + \mathrm{K}(D) + O(1)$. Пусть $\delta_x(\alpha) = \log|D|$ и $\mathrm{KP}(D) \leqslant \alpha$ для некоторого D. Отсюда $\delta_x(\alpha) \geqslant \mathrm{KP}(x) - \alpha - O(1)$.

Докажем (9.10). Пусть $\delta_x(\alpha) = \log|D|$ и $\mathrm{KP}(D) \leqslant \alpha$ для некоторого D. При $k \leqslant \log|D|$ эффективным образом представим множество D в виде объединения 2^k подмножеств вида D_i так, чтобы каждое подмножество D_i содержало $2^{-k}|D|$ элементов.[5] Для этого, используя программу для D, расположим все элементы множества в соответствующем порядке и последовательно откладываем по $2^{-k}|D|$ элементов в качестве таких подмножеств. Сложность каждого такого подмножества не превосходит $\alpha + k + O(1)$. Существует i такое, что $x \in D_i$. Отсюда

$$\delta_x(\alpha + k + O(1)) \leqslant \delta_x(\alpha) - k.$$

Отсюда получаем (9.10). □

Следствие 9.1. *Для каждого x существуют константы c_1 и c_2 такие, что*

$$\beta_x(\alpha) \leqslant c_2 - \alpha + O(1) \tag{9.11}$$

для всех $\alpha \geqslant c_1$.

Доказательство. Пусть c_2 – максимальное значение величины $\log|D|$ при $x \in D$ и $\mathrm{KP}(D) \leqslant c_1$. Отсюда $\delta_x(c_1) \leqslant c_2$ и $\beta_x(\alpha) \leqslant c_2 - \alpha + O(1)$ для всех $\alpha \geqslant c_1$. □

Примеры верхних оценок:

Так как при $D = \{x\}$ будет $\log|D| = 0$ и $\mathrm{KP}(D) = \mathrm{KP}(x) + O(1)$,

$$\delta_x(\mathrm{KP}(x) + O(1)) = 0,$$
$$\beta_x(\mathrm{KP}(x) + O(1)) \leqslant -\mathrm{KP}(x|D).$$

Так как при $x \in D = \Xi_n$, $\log|D| = n$ и $\mathrm{KP}(D) = \mathrm{KP}(n) = O(\log n)$,

$$\delta_x(\mathrm{KP}(n) + O(1)) \leqslant n,$$
$$\beta_x(\mathrm{KP}(n) + O(1)) \leqslant n - \mathrm{KP}(x|n) + O(1).$$

[5] Для простоты изложения считаем, что это число целое. При более строгом рассуждении брать его целое прилижение.

Далее будем рассматривать величины аналогичные (9.6) и (9.7), где для описания (статистической модели) конечной последовательности вместо конечных множеств используются вычислимые распределения вероятностей.

Вычислимое распределение вероятностей на дискретном множестве всех конечных последовательностей это вычислимая функция $P : \Xi \to \mathcal{R}_+$ такая, что $\sum_x P(x) = 1$.[6] Под префиксной сложностью $\mathrm{KP}(P)$ вычислимого распределения P понимаем префиксную сложность программы, с помощью которой вычисляются значения P. Определим

$$\delta_x^p(\alpha) = \min\{-\log P(x) : \mathrm{KP}(P) \leqslant \alpha\},$$
$$\beta_x^p(\alpha) = \min\{-\log P(x) - \mathrm{KP}(x|P) : \mathrm{KP}(P) \leqslant \alpha\}.$$

Область определения этих функций также имеет вид $\{\alpha : \alpha \geqslant c_1\}$ для соответствующей положительной константы c_1.

В следующем предложении устанавливается связь этих величин с величинами (9.6) и (9.7).

Предложение 9.4. *Для произвольного x имеют место соотношения*

$$\delta_x^p(\alpha + O(1)) \leqslant \delta_x(\alpha) \ u \tag{9.12}$$
$$\delta_x(\alpha + \log \delta_x^p(\alpha) + O(1)) \leqslant \delta_x^p(\alpha), \tag{9.13}$$
$$\beta_x^p(\alpha + O(1)) \leqslant \beta_x(\alpha) \ u \tag{9.14}$$
$$\beta_x(\alpha + \log(\beta_x^p(\alpha) + \mathrm{KP}(x|P)) + O(1)) \leqslant \beta_x^p(\alpha). \tag{9.15}$$

Доказательство. Для призвольного подмножества $D \subseteq \Xi$ определим равномерную меру, сосредоточенную на этом множестве:

$$P(x) = \begin{cases} \frac{1}{|D|}, \text{ если } x \in D, \\ 0, \text{ в противном случае.} \end{cases}$$

Тогда $-\log P(x) = \log |D|$ при $x \in D$ и $\mathrm{KP}(P) \leqslant \mathrm{KP}(D) + O(1)$ Отсюда следуют неравенства (9.12) и (9.14).

[6] Можно также рассматривать меры, согласованные со структурой множества Ξ.

Докажем неравенства (9.13) и (9.15). Пусть $x \in \Xi$, P – вычислимое распределение на Ξ такое, что $\mathrm{KP}(P) \leqslant \alpha$ и $\delta_x^p(\alpha) = -\log P(x)$ в случае (9.13) или $\beta_x^p(\alpha) = -\log P(x) - \mathrm{KP}(P)$ в случае (9.15).

Пусть $P(x) > 0$. В противном случае, правые части неравенств (9.13) и (9.15) равны бесконечности и утверждение выполнено автоматически. Пусть число k такое, что $2^{-k+1} \leqslant P(x) < 2^{-k}$.

Рассмотрим конечное множество $D = \{z : P(z) \geqslant 2^{-(k+1)}\}$. Из определения $x \in D$ и $|D| \leqslant 2^{k+1}$. Для описания множества D достаточно знать программу для вычислимой меры P и число $k \leqslant -\log P(x) = \delta_x^p(\alpha)$. Отсюда $\mathrm{KP}(D) \leqslant \alpha + \log k + O(1) \leqslant \alpha + \log \delta_x^p(\alpha) + O(1)$. Отсюда следуют неравенство (9.13).

Для докательства (9.15) используем неравенство $\mathrm{KP}(D) \leqslant \alpha + \log k + O(1) \leqslant \alpha + \log(\beta_x^p(\alpha) + \mathrm{KP}(x|P)) + O(1)$. □

В качестве примера достаточной статистики для произвольной бинарной конечной последовательности x длины n рассмотрим множество $\Xi_n^k = \{y : l(y) = n, \sum_{i=1}^n y_i = k\}$.

Из представлениий (3.12) и (3.13) из раздела 3.3, скорректированных для префиксной сложности, получим

$$\mathrm{KP}(\Xi_n^k) \leqslant 2\log n + O(1).$$

Отсюда получаем

$$\mathrm{KP}(x) \leqslant nH\left(\frac{k}{n}\right) + O(\log n).$$

9.3. Случайность относительно разбиения

Разбиения конечного множества. Определим понятие случайной конечной последовательности относительно произвольного разбиения множества $\Xi_n = \{x : l(x) = n\}$ всех конечных двоичных последовательностей длины n. Под разбиением множества Ξ_n понимается представление этого множества в виде объединения $\mathcal{D}_n = \{D_1, \ldots, D_q\}$ попарно непересекающихся

непустых множеств: $\Xi_n = \cup_{i=1}^{q} D_i$, где q – число элементов разбиения, $D_i \subseteq \Xi_n$ и $D_i \cap D_j \neq \emptyset$ при $i \neq j$.

Разбиение $\mathcal{D}_n$ можно представить в виде набора пар (x, i), где $x \in D_i$. Поэтому будем рассматривать его как конструктивный объект.[7]

Пример разбиения – представление $\Xi_n = \cup_{k=0}^{n} \Xi_n^k$, где

$$\Xi_n^k = \{x \in \Xi_n : \sum_{i=1}^{n} x_i = k\}.$$

Это разбиение использовал А.Н.Колмогоров [12] в своем определении конечной бернуллиевской последовательности. Согласно Колмогорову последовательность $x \in \Xi_n$ называется m-бернуллиевской, если $\mathrm{K}(x|n,k) \geqslant \log \binom{n}{k} - m$. Здесь m – произвольное натуральное число. Напомним, что $|\Xi_n^k| = \binom{n}{k}$. В разделе 3.3 было показано, что $\mathrm{K}(x|n,k) \leqslant \log \binom{n}{k} + O(1)$.

Мы будем рассматривать аналогичное определение с использованием префиксной модификации колмогоровской сложности: последовательность $x \in \Xi_n$ называется m-бернуллиевской, если $\mathrm{KP}(x|k,n) \geqslant \log \binom{n}{k} - m$. Здесь m – произвольное натуральное число. Так как $\mathrm{KP}(x|D) \leqslant \log |D| + O(1)$ (см. примечание в начале раздела 9.2), выполнено неравенство

$$\mathrm{KP}(x|k,n) \leqslant \log \binom{n}{k} + O(1),$$

где $n = l(x)$ – длина и $k = \sum_{i=1}^{n} x_i$ – число единиц в последовательности x.

Пусть $D(x)$ – тот элемент разбиения $\mathcal{D}_n$, который содержит x. Величина

$$d(x|\mathcal{D}_n) = \log |D(x)| - \mathrm{KP}(x|D(x))$$

[7] Заметим, что число n можно вычислить по этому конструктивному объекту.

называется дефектом случайности (по Колмогорову) конечной последовательности $x \in \Xi_n$ относительно разбиения $\mathcal{D}_n$. Назовем последовательность $x \in \Xi_n$ m-случайной относительно разбиения $\mathcal{D}_n = \{D_1, \ldots, D_q\}$, если $d(x|\mathcal{D}_n) \leqslant m$.

Случайность относительно класса вероятностных вер. Покажем, что понятие случайности относительно разбиения эквивалентно понятию случайности относительно некоторого класса вычислимых вероятностных мер.

Вероятностная мера на множестве Ξ_n это функция $P : \Xi_n \to \mathcal{R}_+$ такая, что $\sum_{x:l(x)=n} P(x) = 1$. Таким образом, такая мера задается набором 2^n вещественных чисел, а если мера вычислимая, то эти числа вычислимые, т.е. существует алгоритм, который выдает рациональные приближения этих чисел с любой наперед заданной степенью точности.

Мера P на Ξ_n называется инвариантной относительно разбиения $\mathcal{D}_n = \{D_1, \ldots, D_q\}$, если $P(x) = P(y)$ для любых x и y лежащих в одном элементе разбиения. В этом случае, для задания инвариантной меры P достаточно задать вероятностную меру $R_n = \{r_n(D_1), \ldots r_n(D_q)\}$ на $\mathcal{D}_n$: $\sum_{i=1}^q r_n(D_i) = 1$ и $r_n(D_i) \geqslant 0$ при $1 \leqslant i \leqslant q$. Набор R вычислимых вещественных чисел задается программой, т.е. конструктивным объектом.

Инвариантная мера P представляется в виде $P(x) = \frac{r_n(D(x))}{|D(x)|}$. Назовем $r_n(D)$ весом элемента D разбиения $\mathcal{D}_n$. Существует взаимно-однозначное соответствие между вычислимыми инвариантными мерами и вычислимыми весами разбиения и алгоритм, который переводит веса в значения меры и наоборот.

В частности, мера P инвариантна относительно разбиения $\mathcal{C}_n = \{\Xi_n^k : k = 0, \ldots, n\}$, если $P(x)$ зависит только от длины n и числа единиц k в последовательности x. В этом случае, инвариантная мера представляется в виде $P(x) = \frac{r_n(k)}{\binom{n}{k}}$, где $k = \sum_{i=1}^n x_i$, $r_n(k) = r_n(\Xi_n^k)$ и $\sum_{k=0}^n r_n(k) = 1$.

Каждая вычислимая мера P на Ξ_n задается набором вычислимых вещественных чисел, который задается программой (конструктивным объектом). Фиксируем некоторый способ задания

таких наборов вычислимых вещественных чисел. Рассмотрим условную префиксную сложность $\mathrm{KP}(x|P)$, где P – вычислимая вероятностная мера на Ξ_n. Понимаем условие P в $\mathrm{KP}(x|P)$ как условие относительно программы для вычисления значений P.

Дефект случайности последовательности $x \in \Xi_n$ относительно распределения P на Ξ_n определяется

$$d(x|P) = -\log P(x) - \mathrm{KP}(x|P).$$

Пусть $\mathcal{P}_n$ – класс всех вычислимых мер инвариантных относительно разбиения $\mathcal{D}_n$. Определим дефект инвариантности (дефект случайности относительно класса мер $\mathcal{P}_n$):

$$d(x|\mathcal{P}_n) = \inf_{P \in \mathcal{P}_n} d(x|P). \tag{9.16}$$

Более точно, оптимальный алгоритм, задающий условную сложность $\mathrm{KP}(x|P)$, использует в своей работе произвольные строки p. Некоторые строки кодируют программы для вычисления вычислимых мер. Инфинум в (9.16) берется по тем строкам p, которые определяют вычислимые меры P. Более вычислительно инвариантное определение дефекта случайности относительно вычислимой меры см. в разделе 11.2.

Теорема 9.1. *Дефект инвариантности с точностью до аддитивной константы совпадает с дефектом случайности по Колмогорову:*

$$d(x|\mathcal{P}_n) = \log |D(x)| - \mathrm{KP}(x|D(x)) + O(1),$$

где $D(x)$ – элемент разбиения $\mathcal{D}_n$ содержащий x.

Доказательство. Легко видеть, что дефект инвариантности можно также представить в виде

$$d(x|\mathcal{P}_n) = \log |D(x)| + \inf_{R \in \mathcal{R}_n} (-\log r_n(D(x)) - \mathrm{KP}(x|R)). \tag{9.17}$$

Для произвольного $x \in \Xi_n$ рассмотрим вычислимую инвариантную меру

$$P(z) = \begin{cases} \frac{1}{|D(x)|}, \text{ если } z \in D(x), \\ 0, \text{ в противном случае,} \end{cases}$$

т.е. элементу разбиения $D(x)$ приписан вес 1, остальным элементам разбиения приписаны нулевые веса. Обозначим R' соответствующее распределение весов. Тогда $\mathrm{KP}(x|R') = \mathrm{KP}(x|D(x)) + O(1)$. Данная тривиальная инвариантная мера определяет верхнюю оценку (9.17)

$$d(x|\mathcal{P}_n) \leqslant \log|D(x)| - \mathrm{KP}(x|D(x))) + O(1).$$

Докажем обратное неравенство. Пусть P – произвольная вычислимая инвариантная вероятностная мера на Ξ_n и $R = \{r_n(D_1), \dots r_n(D_q)\}$ – соответствующие веса элементов разбиения.

Рассмотрим перечислимую снизу функцию

$$Q(x|R) = \begin{cases} \sum_{i=1}^{q} r_n(D_i) 2^{-\mathrm{KP}(x|D_i)}, \text{ если } x \in \Xi_n, \\ 0, \text{ в противном случае.} \end{cases}$$

Эта функция является полумерой на Ξ. Действительно, так как для любых n и D будет $\sum_{x\in\Xi_n} 2^{-\mathrm{KP}(x|D)} \leqslant 1$,

$$\sum_{x\in\Xi_n} Q(x|R) = \sum_{D\in\mathcal{D}_n} \sum_{x\in\Xi_n} r_n(D) 2^{-\mathrm{KP}(x|D)} \leqslant \sum_{D\in\mathcal{D}_n} r_n(D) = 1.$$

Функция $Q(x|n, R)$ перечислима снизу (как условная относительно n и R полумера).

По свойству априорной перечислимой полумеры $P(x|R)$ из раздела 5.1.4 (см. также раздел 5.1.5) будет $cP(x|R) \geqslant Q(x|R)$, Отсюда будет

$$cP(x|R) \geqslant Q(x|R) \geqslant r_n(D) 2^{-\mathrm{KP}(x|D)}$$

для любого элемента разбиения и $x \in \Xi_n$, где c -константа. В частности, это неравенство имеет место и для $D = D(x)$. Переходим к префиксной сложности (логарифмируем обе части этого неравенства) и получаем при $D = D(x)$ неравенство $-\mathrm{KP}(x|R) \geqslant \log r_n(D(x)) - \mathrm{KP}(x|D(x)) - O(1)$ или $-\log r_n(D(x)) - \mathrm{KP}(x|R) \geqslant -\mathrm{KP}(x|D(x)) - O(1)$. Так как в последнем неравенстве весовое

распределение R – произвольное, получаем в (9.17) $d(x|n,\mathcal{P}_n) \geqslant \log|D(x)| - \mathrm{KP}(x|D(x)) - O(1)$. $\square$

Для произвольного натурального числа m будем говорить, что конечная последовательность $x \in \Xi_n$ является m-случайной относительно вероятностного распределения P если $d(x|P) \leqslant m$. Аналогичным образом, x является m-случайной относительно класса инвариантных мер $\mathcal{P}_n$, если $d(x|\mathcal{P}_n) \leqslant m$. Из определения следует, что последовательность $x \in \Xi_n$ является m-случайной относительно класса вероятностных мер тогда и только тогда, когда она является m-случайной относительно одной из мер из этого класса.

Из теоремы 9.1 следует

Следствие 9.2. *Последовательность x является m-случайной по Колмогорову тогда и только тогда она m-случайная относительно какой-либо вычислимой инвариантной вероятностной меры.*

Это следствие устанавливает связь между колмогоровским определением бернуллиевости и вероятностным определением.

Часть IV

Последовательные предсказания

Глава 10

Предсказательная сложность

10.1. Задача последовательного прогнозирования

Задача универсального вероятностного прогнозирования в алгоритмической теории информации рассматривалась в разделе 6. В этом разделе мы рассмотрим задачу онлайн прогнозирования в общем случае.

Пусть некоторый процесс последовательно порождает последовательность битов $\omega_1, \omega_2, \ldots$. Для простоты будем считать, что $\omega_i \in \{0, 1\}$ для всех i. Рассмотрим следующую задачу: для любого $n > 1$ по известным исходам $\omega_1, \omega_2, \ldots, \omega_{n-1}$ выдать прогноз p_n будущего n-го исхода. Предполагаем, что $p_n \in [0, 1]$.[1] После того как исход ω_n будет предъявлен, вычисляем расхождение между прогнозом p_n и исходом ω_n с помощью функции потерь $\lambda(\omega_n, p_n)$. Мы будем предполагать, что функция потерь

[1] Можно интерпретировать число p_n как оценку вероятности события $\omega_n = 1$, в этом случае $1 - p_n$ есть вероятность события $\omega_n = 0$. В этом случае, число p_n также равно математическому ожиданию значения бинарного исхода ω_n. Эта интерпретация была использована в разделе 6.

принимает неотрицательные значения и ее значения могут быть вычислены с помощью некоторого алгоритма.

Прогнозирующая стратегия – это вычислимая функция $S : \Xi \to [0,1]$, которая по любой последовательности бинарных исходов $\omega_1\omega_2\ldots\omega_{i-1}$ выдает прогноз $p_i = S(\omega_1\omega_2\ldots\omega_{i-1})$ будущего исхода ω_i, где $i = 1, 2, \ldots$ (здесь $\omega_0 = \lambda$).[2]

Примеры функций потерь – логарифмическая функция потерь и квадратичная функция потерь.

Логарифмическая функция потерь применяется в том случае, когда в качестве прогноза выдается распределение вероятностей $p(\omega)$ на множестве $\{0,1\}$.[3] Логарифмическая функция потерь определяется как

$$\lambda(\omega, p) = \log_\beta p(\omega) = \begin{cases} \log_\beta p \text{ если } \omega = 1 \\ \log_\beta(1-p) \text{ в противном случае,} \end{cases}$$

где $\beta = e^{-\eta}$ – параметр, $\eta > 0$. Здесь $\omega \in \{0,1\}$ и $p = p(1) \in [0,1]$.

В дальнейшем для логарифмической функции потерь часто полагаем $\beta = \frac{1}{2}$, в этом случае $\lambda(\omega, p) = -\log p$ при $\omega = 1$ и $\lambda(\omega, p) = -\log(1-p)$ при $\omega = 0$.

Кумулятивные потери в результате предсказаний $p_1, \ldots, p_n$ исходов $\omega_1, \ldots, \omega_n$ измеряются суммой

$$\mathrm{Loss}(\omega_1\ldots\omega_n) = \sum_{i=1}^{n} \lambda(\omega_i, p_i).$$

В том случае, когда исходы $\omega_1, \ldots, \omega_n$ генерируются некоторой мерой P, можно рассмотреть прогнозирующую систему

$$S(\omega_1\ldots\omega_{i-1}) = P(\sigma|\omega_1\ldots,\omega_{i-1}) = \frac{P(\omega_1\ldots,\omega_{i-1}\sigma)}{P(\omega_1\ldots,\omega_{i-1})}$$

при $\sigma \in \{0,1\}$. Здесь мы предполагаем, что $P(\omega_1\ldots,\omega_i) \neq 0$ для всех i.

[2]Понятие вычислимой функции типа $\Omega \to \mathcal{R}$ или $\mathcal{R} \to \mathcal{R}$ определяется в разделе 11.1.

[3]Аналогичным образом можно рассматривать распределение $p(\omega)$ на конечном множестве исходов.

В этом случае естественно рассматривать логарифмическую функцию потерь. Тогда кумулятивные потери на последовательности исходов $\omega_1, \ldots, \omega_n$ равны логарифму функции правдоподобия

$$\mathrm{Loss}_S(\omega_1 \ldots \omega_n) = \sum_{i=1}^{n} \log_\beta P(\omega_i | \omega_1 \ldots, \omega_{i-1}) = \log_\beta P(\omega_1 \ldots, \omega_n).$$

Квадратичная функция потерь

$$\lambda(\omega, p) = (\omega - p)^2,$$

будет рассматриваться при $\omega \in \{0, 1\}$ (а также при $\omega \in \mathcal{R}$) и $p \in \mathcal{R}$.

Для квадратичной функции потерь кумулятивные потери представляют собой квадратичную ошибку регрессии

$$\mathrm{Loss}_S(\omega_1 \ldots \omega_n) = \sum_{i=1}^{n} (\omega_i - p_i)^2,$$

где $p_i = S(\omega_1 \ldots \omega_{i-1})$.

В этой главе мы рассмотрим задачу прогнозирования конечных последовательностей. В этом случае мы будем строить такую прогнозирующую систему S, для которой кумулятивная сумма

$$\mathrm{Loss}_S(\omega_1 \ldots \omega_n) = \sum_{i=1}^{n} \lambda(\omega_i, S(\omega_1 \omega_2 \ldots \omega_{i-1})) \tag{10.1}$$

является минимально возможной при произвольных возможных исходах $\omega_1, \ldots, \omega_n$ для всех n.

Цель прогнозирующего алгоритма – построить такую последовательность предсказаний $p_1, \ldots, p_n$, чтобы для произвольного n минимизировать кумулятивные потери каковы бы ни были исходы $\omega_1, \ldots, \omega_n$. В случае логарифмической функции потерь эта задача эквивалентна задаче максимизации функции правдоподобия. Для квадратичной функции потерь это задача минимизации суммарной квадратичной ошибки регрессии.

10.2. Перемешиваемые функции потерь

Для некоторых функций потерь задача минимизации кумулятивной суммы (10.1) решается достаточно эффективно.

Рассмотрим класс так называемых перемешиваемых функций потерь. Пусть задано число $\eta > 0$ – параметр обучения, $\beta = e^{-\eta}$. Функция потерь $\lambda(x, p)$ называется η-перемешиваемой, если для каждой последовательности предсказаний $\mathbf{p} = p_1, p_2, \ldots$ и каждой последовательности неотрицательных весов $\mathbf{r} = r_1, r_2, \ldots$, сумма которых не превосходит 1, существует предсказание γ такое, что

$$\lambda(\omega, \gamma) \leqslant \log_\beta \sum_{i=1}^{\infty} r_i \beta^{\lambda(\omega, p_i)} \tag{10.2}$$

для всех ω. Неравенство (10.2) эквивалентно

$$\beta^{\lambda(\omega,\gamma)} \geqslant \sum_{i=1}^{\infty} r_i \beta^{\lambda(\omega, p_i)} \tag{10.3}$$

Фиксируем некоторый способ вычисления γ, удовлетворяющего (10.2), и пишем $\gamma = \mathrm{Subst}(\mathbf{p}, \mathbf{r})$. Называем эту функцию функцией подстановки.

Функция потерь называется перемешиваемой, если она η-перемешиваема для некоторого $\eta > 0$. Мы будем предполагать алгоритмическую эффективность свойства η-перемешиваемости. Это означает, что предсказание γ может быть вычислено с любой наперед заданной степенью точности по достаточно точным приближениям всех параметров, входящих в это определение.

Далее мы покажем, что логарифмическая и квадратичная функции потерь являются перемешиваемыми. Функции постановки для квадратичной и логарифмической функций потерь будут приведены ниже.

Свойство перемешиваемости используется при построении алгоритмов, оптимально усредняющих наборы предсказательных алгоритмов. Один из таких алгоритмов был предложен Вовком [56, 58].

Предложение 10.1. *Пусть задана вычислимая (конечная или бесконечная) последовательность прогнозирующих стратегий $S_1, S_2, \ldots$ и вычислимая последовательность неотрицательных вещественных чисел (весов) $r_1, r_2, \ldots$, сумма которых не превосходит 1. Тогда существует вычислимая прогнозирующая стратегия S такая, что для любого $i = 1, 2, \ldots$*

$$\mathrm{Loss}_S(x) \leqslant \mathrm{Loss}_{S_i}(x) + \log_\beta r_i$$

для всех конечных последовательностей x.

Доказательство. Построение усредняющей прогнозирующей стратегии S происходит по схеме байесовского метода.

Полагаем $r_0(i) = r_i$, $i = 1, 2, \ldots$. После появления очередного исхода x_t веса перестраиваются по формуле

$$r_t(i) = \beta^{\lambda(x_t, S_i(x_1 \ldots x_{t-1}))} r_{t-1}(i).$$

Легко видеть, что при таком определении

$$r_t(i) = \beta^{\mathrm{Loss}_{S_i}(x_1 \ldots x_t)} r_0(i).$$

Определим "экспоненциальную смесь" $g_t(\omega)$, $\omega \in \{0, 1\}$ потерь по всем стратегиям

$$g_t(\omega) = \log_\beta \sum_i \beta^{\lambda(\omega, S_i(x_1 \ldots x_{t-1}))} r^*_{t-1}(i),$$

где

$$r^*_{t-1}(i) = \frac{r_{t-1}(i)}{\sum_s r_{t-1}(s)}$$

нормированные веса.

Так как функция потерь η-перемешиваемая, для любого $t = 1, \ldots, n$ существует (и может быть эффективно вычислено с любой наперед заданной степенью точности) вещественное число p_t такое, что

$$\lambda(\omega, p_t) \leqslant g_t(\omega) \tag{10.4}$$

для всех $\omega \in \{0,1\}$). Определим

$$S(x_1 \dots x_{t-1}) = p_t.$$

Из (10.4) будет следовать, что

$$\mathrm{Loss}_S(x_1 \dots x_T) = \sum_{t=1}^{T} \lambda(x_t, p_t) \leqslant \sum_{t=1}^{T} g_t(x_t).$$

Плэтому важно исследовать сумму $\sum_{t=1}^{T} g_t(x_t)$. Математической индукцией по $t = 1, 2, \dots n$ докажем, что

$$\sum_{t=1}^{T} g_t(x_t) = \log_\beta \sum_i \beta^{\mathrm{Loss}_{S_i}(x_1 \dots x_T)} r_0(t). \tag{10.5}$$

При $T = 1$ имеем

$$g_1(x_1) = \log_\beta \sum_i \beta^{\lambda(x_1, S_i(\lambda))} r_0(i).$$

При $T > 1$ замечая, что $\mathrm{Loss}_{S_i}(x_1 \dots x_T) = \mathrm{Loss}_{S_i}(x_1 \dots x_{T-1}) + \lambda(x_T, S_i(x_1 \dots x_{T-1}))$, представим в виде суммы величину

$$\log_\beta \sum_i \beta^{\mathrm{Loss}_{S_i}(x_1 \dots x_T)} r_0(i) =$$

$$\log_\beta \sum_i \beta^{\mathrm{Loss}_{S_i}(x_1 \dots x_{T-1})} r_0(i) +$$

$$\log_\beta \frac{\sum_i \beta^{\mathrm{Loss}_{S_i}(x_1 \dots x_{T-1}) + \lambda(x_T, S_i(x_1 \dots x_{T-1}))} r_0(i)}{\sum_i \beta^{\mathrm{Loss}_{S_i}(x_1 \dots x_{T-1})} r_0(i)} =$$

$$\log_\beta \sum_i \beta^{\mathrm{Loss}_{S_i}(x_1 \dots x_{T-1})} r_0(i) +$$

$$\log_\beta \sum_i \beta^{\lambda(x_T, S_i(x_1 \dots x_{T-1}))} r^*_{T-1}(i) =$$

$$\log_\beta \sum_i \beta^{\mathrm{Loss}_{S_i}(x_1 \dots x_{T-1})} r_0(i) + g_T(x_T).$$

Продолжаяя раскрывать соответствующую сумму, получим равенство (10.5).

Суммируем неравенство (10.4), объединяем его с неравенством (10.5) и получаем

$$\mathrm{Loss}_S(x) \leqslant \log_\beta \sum_i r_i \beta^{\mathrm{Loss}_{S_i}(x)} \tag{10.6}$$

для всех x.[4]

Так как сумма больше или равна любого слагаемого, из (10.6) получаем

$$\mathrm{Loss}_S(x) \leqslant \log_\beta \sum_i r_i \beta^{\mathrm{Loss}_{S_i}(x)} \leqslant \mathrm{Loss}_{S_i}(x) + \log_\beta r_i$$

для любого i. Предложение 10.1 доказано. □

Процесс построения оптимальной усредняющей стратегии можно представить в виде онлайн протокола.

FOR $t = 1, 2, \ldots$

- Наблюдаем прогнозы $p_{i,t} = S_i(x_1, \ldots x_{i-1})$ при $i = 1, 2, \ldots$.
- Вызываем функцию

  $$g_t(\omega) = \log_\beta \sum_i \beta^{\lambda(\omega, S_i(x_1 \ldots x_{t-1}))} r^*_{t-1}(i)$$

  , где $r^*_{t-1}(i) = \frac{r_{t-1}(i)}{\sum_s r_{t-1}(s)}$ – нормированные веса, а веса $r_{t-1}(i)$ были вычислены на предыдущем щаге.
- Вычисляем прогноз $S(x_1 \ldots x_{t-1}) = p_t$, где p_t – решение системы неравенств

  $$\begin{cases} \lambda(0, p_t) \leqslant g_t(0) \\ \lambda(1, p_t) \leqslant g_t(1) \end{cases}$$
- Получаем очередной бит x_t.
- Адаптируем веса $r_t(i) = \beta^{\lambda(x_t, S_i(x_1 \ldots x_{t-1}))} r_{t-1}(i)$.

[4]Напомним, что $r_0(i) = r_i$ исходные веса.

ENDFOR

Смешиваемость логарифмической функции потерь. Проверим выполнение неравенства (10.3) – определение смешиваемости для логарифмической функции потерь. В этом случае прогноз – это распределение вероятностей $p(\omega)$ на множестве $\{0,1\}$. В этом случае $\lambda(\omega,p)=\log_\beta p(\omega)$ и $\beta^{\lambda(\omega,p)}=p(\omega)$. Неравенство (10.3) превращается в равенство (10.7), а функция подстановки имеет вид байесовской смеси предсказаний $p_i(\omega)$ прогнозирующих стратегий

$$p(\omega)=\sum_i r_i p_i(\omega),. \tag{10.7}$$

где $\sum_i r_i=1$ и $r_i\geqslant 0$ для всех i.

Заметим, что логарифмическая функции потерь является перемешивамой при любом β, так как этот параметр входит в ее определение.

Смешиваемость квадратичной функции потерь. Проверим свойство перемешиваемости и построим функцию подстановки для квадратичной функции потерь $\lambda(\omega,p)=(\omega-p)^2$ при $\omega\in\{0,1\}$ и $p\in[0,1]$.

Предварительно рассмотрим общий случай при $\beta=e^{-\eta}$. Для произвольной функции потерь $\lambda(\omega,\gamma)$ рассмотрим параметрическую кривую на плоскости

$$\left(e^{-\eta\lambda(0,\gamma)},e^{-\eta\lambda(1,\gamma)}\right), \tag{10.8}$$

где γ – параметр, $\omega\in\{0,1\}$. Изучим, при каком η эта кривая является вогнутой по γ. Условие вогнутости имеет вид

$$x'(\gamma)y''(\gamma)-x''(\gamma)y'(\gamma)\geqslant 0 \tag{10.9}$$

для всех γ, где $x(\gamma)=e^{-\eta\lambda(0,\gamma)}$ и $y(\gamma)=e^{-\eta\lambda(1,\gamma)}$.

В частности, для квадратичной функции потерь $\lambda(\omega,\gamma)=(\omega-\gamma)^2$ будет $x(\gamma)=e^{-\eta\gamma^2}$ и $y(\gamma)=e^{-\eta(1-\gamma)^2}$. После элементарных преобразований неравенство (10.9) эквивалентно неравенству

$$\eta\gamma(1-\gamma)\leqslant\frac{1}{2}.$$

Легко видеть, что так как $\gamma(1-\gamma)$ принимает максимальное значение $\frac{1}{4}$, условие вогнутости выполнено для всех γ при $0 < \eta \leqslant 2$.

Условие η-смешиваемости означает, что для любого распределения $\mathbf{w} = (w_1, \ldots, w_k)$ на множестве k прогнозирующих стратегий и любых их прогнозов $\mathbf{f} = (f_1, \ldots, f_k)$ найдется γ такое, что выполнены два неравенства

$$e^{-\eta(\lambda(\omega,\gamma)} \geqslant \sum_{i=1}^{k} w_i e^{-\eta\lambda(\omega,f_i)} \tag{10.10}$$

при $\omega = 0, 1$. Точки $\left(e^{-\eta\lambda(0,f_i)}, e^{-\eta\lambda(1,f_i)}\right)$, $i = 1, \ldots, N$ лежат на кривой (10.8), а их выпуклая комбинация – точка M, лежит внутри выпуклой области, ограниченной этой кривой. Условие (10.10) означает, что абсцисса и ордината точки $N = \left(e^{-\eta\lambda(0,\gamma)}, e^{-\eta\lambda(1,\gamma)}\right)$ не меньше чем абсцисса и ордината точки M. Прямая, проходящая через точку M, отмечает точку $N = (e^{-\eta\lambda(0,\gamma)}, e^{-\eta\lambda(1,\gamma)})$ на кривой (10.8) (см. рис. 10.1). Предсказание γ вычисляется из условия

$$\frac{e^{-\eta\lambda(1,\gamma)}}{e^{-\eta\lambda(0,\gamma)}} = \frac{\sum\limits_{i=1}^{k} w_i e^{-\eta\lambda(1,f_i)}}{\sum\limits_{i=1}^{N} w_i e^{-\eta\lambda(0,f_i)}}. \tag{10.11}$$

В частности, для квадратичной функции потерь левая часть равенства (10.11) имеет вид $e^{-\eta(2\gamma-1)}$. Отсюда получаем выражение для γ:

$$\gamma = \mathrm{Subst}(\mathbf{f}, \mathbf{w}) = \frac{1}{2} - \frac{1}{2\eta} \ln \frac{\sum\limits_{i=1}^{k} w_i e^{-\eta f_i^2}}{\sum\limits_{i=1}^{N} w_i e^{-\eta(f_i-1)^2}}.$$

Условие смешиваемости квадратичной функции потерь позволяет взять $\eta = 2$.

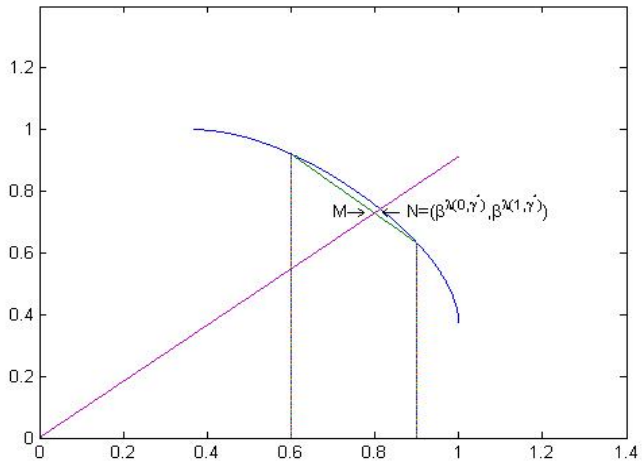

Рис. 10.1: Прямая, проходящая через точку M, отмечает точку $(\beta^{\lambda(0,\gamma)}, \beta^{\lambda(1,\gamma)})$ на кривой, где $\beta = e^{-\eta}$, по которой вычисляется предсказание $\gamma = \mathrm{Subst}(\mathbf{f}, \mathbf{w})$.

Свойство смешиваемости является обобщением свойства экспоненциальной вогнутости. Нетрудно проверить, что для любого $\omega \in [0,1]$ функция $f(\gamma) = e^{-\eta(\omega-\gamma)^2}$ является вогнутой по $\gamma \in [0,1]$ при $0 < \eta \leqslant \frac{1}{2}$. В этом случае величина $\gamma = \sum\limits_i w_i f_i$ удовлетворяет неравенству (10.10) при любом $0 < \eta \leqslant \frac{1}{2}$ по определению вогнутости.

10.3. Конструкция предсказательной сложности

Сумма (10.1) представляет собой некоторую меру сложности последовательного предсказания элементов последовательности $\omega = \omega_1\omega_2\ldots\omega_n$. Основная задача предсказания заключается в нахождении такой прогнозирующей стратегии f, которая мини-

мизирует общие потери (10.1) для всех ω в том же смысле, как в случае колмогоровской сложности. Решение этой задачи возможно в рамках идеи "смешивания", которая была изложена в предложении 10.1. Эта идея в определенном смысле также используется для построения колмогоровской сложности.

Разъясним эту идею на примере префиксной колмогоровской сложности. Рассматриваются префиксно-корректные методы декодирования и соответствующий им префиксный вариант колмоговской сложности. Каждый такой метод F определяет меру сложности $\mathrm{KP}_F(x) = \min\{l(p) : F(p) = x\}$ конечных последовательностей. Легко видеть, что функция $\mathrm{KP}_F(x)$ перечислима сверху. "Смешивание по Колмогорову" заключается в построении вычислимой последовательности F_i всех методов декодирования средствами теории алгоритмов и последующем объединении их в один универсальный метод декодирования $U(\langle i, p\rangle) = F_i(p)$, где i – программа вычисляющая F_i, а $\langle i, p\rangle$ – некоторый код пары программ i и p, например, построенный в разделе 3.1.1.

По свойству нумерации пар для любого вычислимого метода декодирования F выполнено $\mathrm{KP}_U(x) \leqslant \mathrm{KP}_F(x) + O(1)$ для всех x, где константа $O(1)$ зависит от F. Одна из возможных функций $\mathrm{KP}_U(x)$ фиксируется и обозначается $\mathrm{KP}(x)$.

Префиксную колмогоровскую сложность можно также определить "аналитическим" способом

$$\mathrm{KP}(x) = -\log \sum_{i=1}^{\infty} r_i 2^{-\mathrm{KP}_{F_i}(x)}, \tag{10.12}$$

где $r_1, r_2, \ldots$ – произвольная перечислимая снизу последовательность рациональных чисел с суммой 1. Из определения (10.12) непосредственно следует, что для произвольного i имеет место неравенство

$$\mathrm{KP}(x) \leqslant \mathrm{KP}_{F_i}(x) + \log \frac{1}{r_i}$$

для всех x.

Идея аналитического метода, по видимому, восходит к Р.Соломонову [53]. Такой же метод был применен Левиным [16]

при построении универсальной перечислимой снизу полумеры на дискретном множестве (см. раздел 5.1.3) и априорной полумеры на дереве всех двоичных последовательностей (см. раздел 6.1).

Предсказательная сложность была введена Вовком в работах [59] и [60]. Предсказательная сложность будет строиться вторым из приведенных выше способов. Будет построена перечислимая сверху последовательность мер "предсказательной сложности" $\mathrm{KG}_i(x)$ и, в случае η-перемешиваемой функции потерь, можно будет определить для $\beta = e^{-\eta}$

$$\mathrm{KG}(x) = \log_\beta \sum_{i=1}^{\infty} r_i \beta^{\mathrm{KG}_i(x)} \tag{10.13}$$

для всех $x \in \Xi$. Свойство η-перемешиваемости позволит доказать, что функция $\mathrm{KG}(x)$ также является мерой предсказательной сложности.

Функция $\mathrm{KG}(x)$ называется мерой предсказательной сложности, если выполнены следующие условия:

- (i) $K(\lambda) = 0$ и для каждого x существует предсказание p такое, что $\mathrm{KG}(x\sigma) - \mathrm{KG}(x) \geqslant \lambda(\sigma, p)$ для всех $\sigma \in \{0, 1\}$.[5]
- (ii) KG перечислима сверху, т.е. существует вычислимая неубывающая последовательность неотрицательных простых функций KG^t типа $\Xi \to \mathcal{R}_+$ такая, что для всех x выполнено $\mathrm{KG}(x) = \inf_t \mathrm{KG}^t(x)$ (под простой понимается функция, принимающая только рациональные значения или ∞, которая равна ∞ для почти всех значений аргумента).

Из условия (i) следует, что существует (не обязательно вычислимая) прогнозирующая стратегия $p = S(x)$ такая, что $\mathrm{Loss}_S(x) \leqslant \mathrm{KG}(x)$ для всех конечных последовательностей x. В том случае,

[5] Это неравенство является обобщением равенства $\mathrm{Loss}_S(x\sigma) - \mathrm{Loss}_S(x) = \lambda(\sigma, p)$ для кумулятивных потерь, где S – прогнозирующая стратегия и $p = S(x)$.

когда в (i) выполняется равенство, мы получаем определение функции потерь (10.1). Условие (ii) означает, что $\mathrm{KG}(x)$ вычислима лишь "в пределе".

Предложение 10.2. *Для логарифмической функции потерь при $\beta = \frac{1}{2}$ ($\eta = \ln 2$), условия (i)–(ii) эквивалентны требованиям, чтобы функция $P(x) = 2^{-\mathrm{KG}(x)}$ являлась перечислимой снизу полумерой.*

Доказательство. Для логарифмической функции потерь будет $\lambda(\sigma, p) = -\log p$, если $\sigma = 1$ и $\lambda(\sigma, p) = -\log(1-p)$, если $\sigma = 0$.

Пусть функция $\mathrm{KG}(x)$ удовлетворяет условию (i). Тогда, так как $\mathrm{KG}(x1) - \mathrm{KG}(x) \geqslant -\log p$ и $\mathrm{KG}(x0) - \mathrm{KG}(x) \geqslant -\log(1-p)$ для некоторого $p \in [0,1]$, будет $2^{-\mathrm{KG}(x1)} \leqslant p2^{-\mathrm{KG}(x)}$ и $2^{-\mathrm{KG}(x0)} \leqslant (1-p)2^{-\mathrm{KG}(x)}$ Отсюда функция $P(x) = 2^{-\mathrm{KG}(x)}$ удовлетвряет условию $P(x1) + P(x0) = 2^{-\mathrm{KG}(x1)} + 2^{-\mathrm{KG}(x0)} \leqslant 2^{-\mathrm{KG}(x)} = P(x)$, т.е. является полумерой. Пречислимость снизу следует из условия (ii).

В обратную сторону, пусть $P(x) \geqslant P(x0) + P(x1)$ для всех $x \in \Xi$. Тогда $\frac{Px1)}{P(x)} + \frac{Px0)}{P(x)} \leqslant 1$ и существует вещественное число p такое, что $\frac{Px1)}{P(x)} \leqslant p$ и $\frac{Px0)}{P(x)} \leqslant 1-p$. Таким образом, (i) выполнено для функции $\mathrm{KG}(x) = -\log P(x)$. Условие (ii) следует из перечислимости снизу функции $P(x)$. $\square$.

Построение оптимальной предсказательной сложности будет аналогично построению универсальной перечислимой снизу (априорной) полумеры.

Теорема 10.1. *Пусть функция потерь $\lambda(\omega, p)$ является перемешиваемой и непрерывной по p. Тогда существует мера предсказательной сложности $\mathrm{KG}(x)$ такая, что для любой меры предсказательной сложности $\overline{\mathrm{KG}}(x)$ найдется константа c так, что*

$$\mathrm{KG}(x) \leqslant \overline{\mathrm{KG}}(x) + c \tag{10.14}$$

для всех x

Предварительно докажем следующую лемму.

Лемма 10.1. *Можно построить равномерно перечислимую сверху последовательность* $\mathrm{KG}_i(x)$ *всех мер предсказательной сложности. Это означает, что существует вычислимая последовательность неотрицательных простых функций* $\mathrm{KG}_i^t(x)$*, для которой*

- $\mathrm{KG}_i^{t+1}(x) \leqslant \mathrm{KG}_i^t(x)$ *для всех* i, t, x*;*
- $\mathrm{KG}_i(x) = \inf_t \mathrm{KG}_i^t(x)$ *для всех* i, t, x*;*
- *для каждой меры предсказательной сложности* $\overline{\mathrm{KG}}(x)$ *существует* i *такое, что* $\overline{\mathrm{KG}}(x) = \mathrm{KG}_i(x)$ *для всех* x*.*

Доказательство. Последовательность $\mathrm{KG}_i(x)$ может быть построена следующим образом. Мы будем рассматривать перечислимые множества пар (x, r), где x – конечная двоичная последовательность и r – неотрицательное рациональное число. Пусть W_i – равномерно перечислимая последовательность всех перечислимых множеств, состоящих из пар (r, x), где $x \in \Xi$, $r \in \mathcal{Q}$, такая, что любое перечислимое множество пар равно одному из W_i.[6] По определению равномерная перечислимость означает, что множество $W = \{(i, r, x) : (r, x) \in W_i\}$ перечислимое.

Из вычислимости функции $\lambda(\sigma, p)$ следует, что для некоторой вычислимой последовательности простых функций $\lambda^t(\sigma, p)$ выполнено $\lambda^{t+1}(\sigma, p) \leqslant \lambda^t(\sigma, p)$ для всех t, σ, p и $\lambda(\sigma, p) = \inf_t \lambda^t(\sigma, p)$.

Пусть W^t – конечная часть W, перечисленная за t шагов. Определим

$$W_i^t = \{(x, r) | ((i, x, r) \in W^t)\} \cup (\Xi \times \{+\infty\}).$$

С помощью математической индукции по t определим вычислимую последовательность неотрицательных простых функций $\mathrm{KG}_i^t(x)$ такую, что $\mathrm{KG}_i^0(x) = \infty$ и $\mathrm{KG}_i^{t+1}(x) \leqslant \mathrm{KG}_i^t(x)$ для всех x,

[6] Последовательность W_i была построена в разделе 3.1.2.

для любых i и t для функции $\mathrm{KG}_i^t(x)$ выполнено условие (i) определения меры предсказательной сложности и график функции $\mathrm{KG}_i^t(x)$ является подмножеством W_i^t.

Для построения такой последовательности для каждого t используя пары $(x,r) \in W_i^t$ ищем максимальное расширение графика функции $\mathrm{KG}_i^{t-1}(x)$ до графика функции $\mathrm{KG}_i^t(x)$, для которой $\mathrm{KG}_i^t(x) \leqslant \mathrm{KG}_i^{t-1}(x)$ для всех x и для любого x существует рациональное p такое, что

$$\mathrm{KG}_i^t(x\sigma) - \mathrm{KG}_i^t(x) \geqslant \lambda^t(\sigma, p) \tag{10.15}$$

для всех $\sigma = 0, 1$.[7]

Определим $\mathrm{KG}_i(x) = \inf_t \mathrm{KG}_i^t(x)$ для всех i и x. Из (10.15) и непрерывности $\lambda(\sigma, p)$ по p следует, что для любого i функция $\mathrm{KG}_i(x)$ удовлетворяет условию (i) определения меры предсказательной сложности , т.е. является мерой предсказательной сложности.

Допустим, что функция $\overline{\mathrm{KG}}(x)$ удовлетворяет условиям (i), (ii) определения меры предсказательной сложности и

$$W_i = \{(x,r) | r > \overline{\mathrm{KG}}(x)\},$$

где r – рациональное. При указанной выше процедуре построения простых функций всегда будут получаться функции $\mathrm{KG}_i^t(x) \geqslant \mathrm{KG}(x)$, которые будут иметь пределом функцию $\overline{\mathrm{KG}}(x)$, т.е. $\overline{\mathrm{KG}}(x) = \mathrm{KG}_i(x)$ для всех x. □

Доказательство теоремы 10.1. Пусть r_i – перечислимая снизу последовательность действительных чисел такая, что $\sum\limits_{i=1}^{\infty} r_i \leqslant 1$. Например, можно взять $r_i = 2^{-\mathrm{KP}(i)}$, где $\mathrm{KP}(i)$ – префиксная сложность i.

[7] Если нетривиальное расширение отсутствует, то полагаем $\mathrm{KG}_i^t(x) = \mathrm{KG}_i^{t-1}(x)$, если имеется несколько таких расширений, то выбираем первое в порядке перечисления.

Пусть функция потерь является η-перемешиваемой для некоторого $\eta > 0$. Определим при $\beta = e^{-\eta}$

$$\mathrm{KG}(x) = \log_\beta \sum_{i=1}^{\infty} \beta^{\mathrm{KG}_i(x)} r_i, \tag{10.16}$$

где $r_i = 2^{-\mathrm{KP}(i)}$.[8] Нам необходимо доказать, что $\mathrm{KG}(x)$ является мерой предсказательной сложности. По определению $\mathrm{KG}(x)$ перечислима сверху, т.е. (ii) определения меры предсказательной сложности выполнено. Остается проверить условие (i).

Действительно, так как каждая функция $\mathrm{KG}_i(x)$ удовлетворяет условию (i) определения меры предсказательной сложности, для любого i существует предсказание p_i, удовлетворяющее

$$\mathrm{KG}_i(x\sigma) - \mathrm{KG}_i(x) \geqslant \lambda(\sigma, p_i),$$

при всех $\sigma \in \{0,1\}$. Отсюда для всех x и σ

$$\mathrm{KG}(x\sigma) - \mathrm{KG}(x) = \log_\beta \sum_i r_i \beta^{\mathrm{KG}_i(x\sigma)} - \log_\beta \sum_i r_i \beta^{\mathrm{KG}_i(x)} =$$

$$\log_\beta \frac{\sum\limits_i r_i \beta^{\mathrm{KG}_i(x\sigma)}}{\sum\limits_i r_i \beta^{\mathrm{KG}_i(x)}} = \log_\beta \sum_i \frac{r_i \beta^{\mathrm{KG}_i(x)}}{\sum\limits_s r_s \beta^{\mathrm{KG}_s(x)}} \beta^{\mathrm{KG}_i(x\sigma) - \mathrm{KG}(x)} =$$

$$\log_\beta \sum_i q_i \beta^{\mathrm{KG}_i(x\sigma) - \mathrm{KG}_i(x)} \geqslant \log_\beta \sum_i q_i \beta^{\lambda(\sigma, p_i)} \geqslant \lambda(\sigma, p),$$

для некоторого p, где новые нормированные веса имеют вид

$$q_i = \frac{r_i \beta^{\mathrm{KG}_i(x)}}{\sum\limits_s r_s \beta^{\mathrm{KG}_s(x)}}.$$

Предсказание p существует, так как функция $\lambda(\sigma, p)$ является η-перемешиваемой, а последнее неравенство и есть условие перемешиваемости.

[8] Как известно из раздела 5.1.3 $\sum_i 2^{-\mathrm{KP}(i)} \leqslant 1$.

Для любой меры предсказательной сложности $\overline{\mathrm{KG}}(x)$ найдется i такое, что $\overline{\mathrm{KG}}(x) = \mathrm{KG}_i(x)$ для всех x. Тогда из (10.16) получаем, что для любого i будет

$$\mathrm{KG}(x) \leqslant \overline{\mathrm{KG}}(x) + \frac{\ln 2}{\eta}\mathrm{KP}(i) \tag{10.17}$$

для всех x. $\square$

Фиксируем одну из мер предсказательной сложности $\mathrm{KG}(x)$, удовлетворяющей теореме 10.1, назовем ее (оптимальной) предсказательной сложностью конечной последовательности x.

Можно показать, что в общем случае не существует вычислимой предсказательной стратегии S такой, что $\mathrm{KG}(x) = \mathrm{Loss}_S(x)$ для всех x.

Выделим неравенство (10.17) в виде следствия.

Следствие 10.1. *Пусть выполнено условие теоремы 10.1. Тогда*

$$\mathrm{KG}(x) \leqslant \mathrm{KG}_i(x) + \frac{\ln 2}{\eta}\mathrm{KP}(i) \tag{10.18}$$

для всех i и x, где $\mathrm{KP}(i)$ – префиксная сложность числа i.

В частности, для любой (вычислимой) прогнозирующей стратегии S и любого x

$$\mathrm{KG}(x) \leqslant \mathrm{Loss}_S(x) + \frac{\ln 2}{\eta}(\mathrm{KP}(S) + c), \tag{10.19}$$

где c – константа, $\mathrm{KP}(S)$ – префиксная сложность вычислимой функции S.

Доказательство. Утверждение (10.18) прямо следует из определения (10.16).

Пусть S – вычислимая прогнозирующая стратегия и p – программа, вычисляющая значения $S(x)$ с любой наперед заданной степенью точности. Тогда существует вычислимая функция f, переводящая p в программу перечисления $S(x)$ сверху так, что

$$\mathrm{Loss}_S(x) = \mathrm{KG}_{f(p)}(x). \tag{10.20}$$

В частности, для любого x

$$\mathrm{KG}(x) \leqslant \mathrm{Loss}_S(x) + \frac{\ln 2}{\eta}(\mathrm{KP}(S) + c),$$

где c – положительная константа и $\mathrm{KP}(S)$ – сложность S. □

По предложению 10.2 в случае логарифмической функции потерь и $\beta = \frac{1}{2}$ любая мера предсказательной сложности совпадает с минус логарифмом некоторой перечислимой снизу полумеры на дереве всех двоичных последовательностей. Так как оптимальная предсказательная сложность $\mathrm{KG}(x)$ минимальна с точностью до константы, она должна совпадать (с точностью до константы) с минус логарифмом априорной полумеры $\mathrm{KA}(x) = -\log M(x)$:

$$\mathrm{KG}(x) = -\log M(x) + O(1),$$

где $M(x)$ – универсальная (априорная) полумера на дереве всех двоичных последовательностей, определенная в разделе 6.1 главы 6. Функция $\mathrm{KA}(x) = -\log M(x)$ была определена в разделе 6.1.

В следующем утверждении представлено неравенство между предсказательной и префиксной сложностями.

Предложение 10.3. *Пусть функция потерь $\lambda(\omega, p)$ является η-перемешиваемой и удовлетворяет условию $\lambda(0,0) = \lambda(1,1) = 0$. Тогда существует константа c такая, что для всех x*

$$\mathrm{KG}(x) \leqslant \frac{\ln 2}{\eta}\mathrm{KP}(x) + c.$$

Доказательство. Пусть $x = x_1 \ldots x_n$. Рассмотрим прогнозирующую стратегию S, которая для каждой последовательности z длины $i-1$, $i = 1, \ldots, l(x) - 1$, выдает i-ый элемент x_i. Такую стратегию можно задать по самой короткой программе p последовательности x (считаем, что $S(z) = 0$ для всех z длины $> l(x)$). $l(p) = \mathrm{KP}(x)$, поэтому $\mathrm{KP}(S) \leqslant \mathrm{KP}(x) + c$ для некоторой константы c. По определению $\mathrm{Loss}_S(x) = 0$. Утверждение следует из (10.18). □

10.4. Просто и сложно предсказуемые конечные последовательности

В этом разделе будем изучать вопрос о сложности $\mathrm{KP}(S)$ вычислимой предсказательной стратегии S, которая несет потери близкие к минимальным: $\mathrm{Loss}_S(x) \approx \mathrm{KG}(x)$.[9] Для каких x это выполнено когда $\mathrm{KP}(S)$ достаточно мала? Существуют ли такие конечные последовательности x, для которых $\mathrm{Loss}_S(x) \approx \mathrm{KG}(x)$ только для тех S, для которых $\mathrm{KP}(S) \approx \mathrm{KP}(x)$?[10]

Полумеры на множествах конечных последовательностей. В дальнейшем мы будем оценивать априорную полумеру $M(D)$ множеств D, состоящих их конечных последовательностей одной и той же длины. Приведем соответствующие определения.

Напомним, что $\Gamma_x = \{\omega \in \Omega : x \subseteq \omega\}$ – интервал (шар) в пространстве Ω. Для любой перечислимой снизу полумеры P можно определить вероятностную машину Тьюринга (МТ) (L, F) такую, что

$$P(x) = L\{\omega \in \Omega : x \subseteq F(\omega)\}.$$

Величина $P(x)$ равна вероятности того, что эта МТ напечатает последовательность, префиксом которой является x. Для любого множества $D \subseteq \Xi$, состоящего из конечных последовательностей, определим полумеру этого множества

$$P(D) = L\{\omega \in \Omega : (\exists x \in D)(x \subseteq F(\omega))\}.$$

Легко видеть, что если D состоит из попарно несравнимых последовательностей, то $P(D) = \sum_{x \in D} P(x)$. Более того, для любого множества D, состоящего из конечных последовательностей, существуют попарно несравнимые последовательности $x_1, x_2, \cdots \in$

[9] Как было замечено выше в разделе 10.3, сама предсказательная сложность может не порождаться никакой вычислимой предсказательной стратегией.

[10] Равенство $\mathrm{KP}(S) = \mathrm{KP}(x) + O(1)$ выполнено для предсказательной стратегии $S_x(x_1 \ldots x_{i-1}) = x_i$ при $1 \leqslant i \leqslant n$, которая использует информацию об $x = x_1 \ldots x_n$.

D такие, что $P(D) = \sum_i P(x_i)$. Можно выбрать $x_1, x_2, \cdots \in D$ так, чтобы для любого $x \in D$ существует i такое, что $x_i \subseteq x$.

(α, γ)-предсказуемость. Допустим, что задана некоторая η-перемешиваемая функция потерь $\lambda(\omega, \gamma)$, где $\eta > 0$ – параметр обучения.

Соответствующая предсказательная сложность $\mathrm{KG}(x_1 \ldots x_n)$ определяет асимптотическую нижнюю границу кумулятивных потерь

$$S(x_1 \ldots x_n) = \sum_{i=1}^{n} \lambda(x_i, S(x_1 \ldots x_{i-1}))$$

для любой как угодно сложной вычислимой прогнозирующей стратегии S.

Префиксная сложность $\mathrm{KP}(S)$ вычислимой стратегии S равна длине самой короткой двоичной строки, представляющей программу для вычисления значений $S(x)$, при универсальном префиксно-корректным способе программирования.

Согласно неравенству 10.19 следствия 10.1, предсказательная сложность $\mathrm{KG}(x)$ конечной последовательности x определяет нижнюю границу суммарных потерь $\mathrm{Loss}_S(x)$ при последовательном предсказании всех битов x с помощью произвольного вычислимого метода предсказания S.

В этом разделе мы приведем метод построения последовательностей $x = x_1 \ldots x_n$, для которых любая "простая" (т.е. "малой" сложности α) вычислимая прогнозирующая стратегия S несет кумулятивные потери близкие к предсказательной сложности. Такие последовательности x будем неформально называть "α-сложно предсказуемыми".

Мы оценим сверху и снизу априорную полумеру множества всех "α-сложно предсказуемых" последовательностей заданной длины и покажем, что априорная полумера M множества всех "α-сложно предсказуемых" последовательностей заданной длины n имеет порядок $2^{-\alpha+O(\log n)}$.

Также будет доказано, что существует простая вычислимая прогнозирующая стратегия S, которая для "подавляющего боль-

шинства" (по полумере M) последовательностей x несет потери близкие к асимптотически нижней границе – предсказательной сложности $\mathrm{KG}(x)$. Это значит, что "подавляющее большинство" конечных последовательностей предсказуемы с помощью "простой" пронозирующей стратегии, которая несет потери близкие к минимально возможным.

Переходим к точным определениям. Пусть α и γ – неотрицательные целые числа. Конечная последовательность x называется (α,γ)-предсказуемой, если существует прогнозирующая стратегия S такая, что $\mathrm{KP}(S) \leqslant \alpha$ (α-простая стратегия) и

$$\mathrm{Loss}_S(x) - \mathrm{KG}(x) \leqslant \gamma.$$

Свойство (α,γ)-предсказуемости конечной последовательности x означает, что существует α-простая прогнозирующая стратегия, потери которой при последовательном прогнозировании битов последовательности x с точностью до γ близки к нижней границе таких потерь, а именно, к предскзательной сложности $\mathrm{KG}(x)$ этой последовательности.

В случае логарифмической функции потерь и $\beta = \frac{1}{2}$

$$\mathrm{Loss}(\omega_1 \ldots \omega_n) = -\log p(\omega_1) \ldots p_n(\omega_n),$$

где $p(\omega_i)$ – предсказание (условная вероятность события $\omega_{=}1$) и

$$\mathrm{KG}(\omega_1 \ldots \omega_n) = -\log M(\omega_1 \ldots \omega_n) = \mathrm{KA}(\omega_1 \ldots \omega_n).$$

Величина

$$d_P(\omega_1 \ldots \omega_n) = -\log P(\omega_1 \ldots \omega_n) - \mathrm{KA}(\omega_1 \ldots \omega_n) \qquad (10.21)$$

называется дефектом случайности последовательности $\omega_1 \ldots \omega_n$ относительно меры $P(\omega_1 \ldots \omega_n) = p(\omega_1) \ldots p_n(\omega_n)$.[11]

Следующее ниже предложение 10.4 показывает, что для широкого класса функций потерь "сложно предсказуемые" последовательности существуют.

Пусть функция потерь $\lambda(\omega, p)$ является вычислимой и удовлетворяет условиям:

[11] См. аналогичные определения (5.13) и (6.3).

1) Область определения функции $\lambda(\omega, p)$: $\omega \in \{0,1\}$, $p \in [0,1]$,

2) $\lambda(0,0) = \lambda(1,1) = 0$,

3) существует вычислимое вещественное число $b > 0$ такое, что для каждого p будет $\lambda(0,p) \geqslant b$ или $\lambda(1,p) \geqslant b$,

4) функция потерь $\lambda(\omega, p)$ является η-перемешиваемой.

Логарифмическая функция потерь удовлетворяют этим условиям при $b = \log_\beta \frac{1}{2} = \frac{\ln 2}{\eta}$ и при любом $\eta > 0$, где $\beta = e^{-\eta}$. При $\beta = \frac{1}{2}$ будет $b = 1$. Для квадратичной функции потерь подходят $b = \frac{1}{4}$ и $0 < \eta \leqslant 2$ (подробнее см. [58]).

Оценка полумеры множества всех сложно предсказуемых последовательностей. В следующем предложении дается нижняя оценка вероятности того, что некоторая вероятностная машина Тьюринга выдаст сложно предсказуемую конечную последовательность.

Предложение 10.4. *Для любой функции потерь, удовлетворяющей условиям (1)–(4), существует положительная константа c такая, что для любых n и $\alpha \geqslant \log n + 2\log\log n + c$ существует последовательность x длины n, для которой*

1) $\mathrm{Loss}_S(x) - \mathrm{KG}(x) \geqslant bn - (2\ln 2/\eta)\alpha - (\ln 2/\eta)(\log n + 2\log\log n) - c$ *для каждой прогнозирующей стратегии S такой, что* $\mathrm{KP}(S) \leqslant \alpha$,[12]

2) $M(x) \geqslant 2^{-\alpha - \log n - 2\log\log n - c}$.

Доказательство этого предложения приведено в разделе 10.4.1.

Наложим ограничение на сложность прогнозирующих систем: $\mathrm{KP}(S) \leqslant \alpha$, где α отражает ограничение на величину допустимых вычислительных ресурсов. Мы покажем, что даже в

[12] В случае логарифмической функции потерь согласно определению (10.21) это неравенство превращается в неравенство для дефекта случайности $d_P(x) \geqslant bn - (2\ln 2/\eta)\alpha - (\ln 2/\eta)(\log n + 2\log\log n) - c$.

том случае, когда α мало по сравнению с длиной n (например, $\alpha = O(\log n)$), большую часть последовательностей длины n можно прогнозировать с почти минимально возможными потерями.

Пусть $D^n_{\alpha,\gamma}$ – множество всех двоичных последовательностей длины n, которые не являются (α,γ)-предсказуемыми. Для любого $x \in D^n_{\alpha,\gamma}$ имеем

$$\mathrm{Loss}_S(x) - \mathrm{KG}(x) > \gamma$$

для каждой прогнозирующей стратегии S такой, что $\mathrm{KP}(S) \leqslant \alpha$. В следующем предложении мы получим верхнюю оценку величины $M(D^n_{\alpha,\gamma}) = \sum_{x\in D^n_{\alpha,\gamma}} M(x)$, где M – априорная полумера.

Напомним, что согласно неравенству (6.5) из раздела 6.1 для произвольного множества D, состоящего из попарно несравнимых конечных последовательностей, величина $M(D) = \sum_{x\in D} M(x)$ определяет асимптотически максимальную вероятность генерации последовательности из D на вероятностной машине Тьюринга.

Следующее предложение утверждает, что вероятность генерации "α-сложно предсказуемой" конечной последовательности на любой вероятностной машине Тьюринга экспоненциально убывает с ростом α.

Предложение 10.5. • *Для любой перемешиваемой функции потерь, удовлетворяющей условию 2), существует положительная константа c такая, что для всех n, α и $1 \leqslant \gamma \leqslant n$*

$$M(D^n_{\alpha,\gamma}) = \sum_{x\in D^n_{\alpha,\gamma}} M(x) \leqslant 2^{-\alpha+2\log n+2\log\log n-\log\gamma+c}; \tag{10.22}$$

• *В случае логарифмической функции потерь $\lambda(\omega,p) =$*

$-\log p(\omega)$ *(при* $\beta = \frac{1}{2}$*) выполнено*

$$M(D^n_{\alpha,\gamma}) = \sum_{x \in D^n_{\alpha,\gamma}} M(x) \leqslant 2^{-\alpha + \log n + 2\log\log n - \log\gamma + c}. \tag{10.23}$$

Доказательство этого предложения приведено в разделе 10.4.2.

Это утверждение обобщает теорему 3 из работы [5] на более широкий класс функций потерь.

Неравенства (10.22) и (10.23) представляют верхнюю оценку на вероятность генерации трудно предсказуемых последовательностей с помощью произвольного вероятностного алгоритма.

По предложению 10.5 для любого m вероятность генерации (α, γ)-предсказуемых последовательностей будет больше или равна $1 - 2^{-m}$ при

$$\alpha + \log\gamma > 2\log n + 2\log\log n + c + m.$$

Последнее условие заведомо имеет место для всех $1 \leqslant \gamma \leqslant n$, если

$$\alpha \geqslant 2\log n + 2\log\log n + c + m. \tag{10.24}$$

Предложение 10.5 показывает, что если верхняя граница сложности прогнозирующих стратегий α удовлетворяет (10.24), то большинство (по полумере M) последовательностей x являются просто предсказуемыми с помощью некоторой вычислимой прогнозирующей стратегии S такой, что $\mathrm{KP}(S) \leqslant \alpha$. Для таких x суммарные потери прогнозирования $\mathrm{Loss}_S(x)$ близки к нижней границе $\mathrm{KG}(x)$.

Таким образом, по предложению 10.4 трудно предсказуемые последовательности существуют, однако по предложению 10.5 любая вероятностная машина может выдавать такие последовательности только с асимптотически малой вероятностью.

Мы суммируем основные результаты этого раздела в виде следующей теоремы.

Теорема 10.2. *Для произвольной функции потерь, удовлетворяющей условиям 1)–4), существует константа c такая, что*

$$2^{-\alpha-\log n-2\log\log n-c} \leqslant M(D^n_{\alpha,\gamma}) \leqslant 2^{-\alpha+2\log n+2\log\log n-\log\gamma+c}$$

для всех n, $\alpha \geqslant \log n + 2\log\log n + c$ и $0 < \gamma \leqslant bn - (2\ln 2/\eta)\alpha - (\ln 2/\eta)(\log n + 2\log\log n) - c$.

10.4.1. Доказательство предложения 10.4

Фиксируем некоторую достаточно высокую точность вычисления всех функций, принимающих действительные значения. Для произвольного α пусть $p_1, p_2, \ldots, p_k$ – все программы длины $\leqslant \alpha$, которые при данной точности вычисления останавливаются и выдают результат для всех $z, l(z) \leqslant n$. Для произвольного $j = 1, \ldots k$ пусть $S_j(z)$ – результат, выдаваемый p_j на z.

Имеем $k < 2^{\alpha+1}$. Используя предложение 10.1, построим усредняющую стратегию S_α такую, что

$$\mathrm{Loss}_{S_\alpha}(x) \leqslant \log_\beta \sum_{i=1}^{k} k^{-1} \beta^{\mathrm{Loss}_{S_i}(x)}. \tag{10.25}$$

Пусть p – программа, имеющая максимальное время работы (на z, $l(z) \leqslant n$) среди всех программ $p_1, p_2, \ldots, p_k$. Используя p, мы можем восстановить все значения $S_\alpha(z)$, $l(z) \leqslant n$, с заданной степенью точности. После этого, определим последовательность $x = x_1 x_2 \ldots x_n$ следующим образом. Вычисляем рациональные приближения для действительных чисел b сверху и рациональные приближения к числам $\lambda(1, S_\alpha(x_1 \ldots x_{s-1}))$ и $\lambda(0, S_\alpha(x_1 \ldots x_{s-1}))$ снизу до тех пор, пока для одного из них не будет выполнено

$$\lambda(1, S_\alpha(x_1 \ldots x_{s-1})) \geqslant b - 2^{-(s+1)} \tag{10.26}$$

или

$$\lambda(0, S_\alpha(x_1 \ldots x_{s-1})) \geqslant b - 2^{-(s+1)} \tag{10.27}$$

(мы предполагаем, что $x_1 \ldots x_{s-1} = \lambda$ при $s = 1$). По свойству 3) функции потерь хотя бы одно из неравенств, (10.26) или (10.27), обязательно будет иметь место. Если (10.26) было вычислено первым, определим $x_s = 1$, и определим $x_s = 0$, в противном случае. По определению $\mathrm{Loss}_{S_\alpha}(x) \geqslant bn - 1$. По (10.25)

$$\mathrm{Loss}_{S_i}(x) \geqslant bn - (\ln 2/\eta)\alpha - 1$$

для всех $1 \leqslant i \leqslant k$ (т.е. для всех P таких, что $\mathrm{KP}(S) \leqslant \alpha$).

По способу определения последовательности x будет $\mathrm{KP}(x|n) \leqslant \alpha + c$ для некоторой положительной константы c. Имеет место неравенство, связывающее безусловную и условную префиксную сложности

$$\mathrm{KP}(x) \leqslant \mathrm{KP}(x|n) + \log n + 2\log\log n + c \tag{10.28}$$

для некоторой константы c. Учитывая неравенство (10.28) и предложение 10.3, получим

$$\mathrm{KG}(x) \leqslant (\ln 2/\eta)\mathrm{KP}(x) \leqslant (\ln 2/\eta)(\alpha + \log n + 2\log\log n) + c$$

для некоторой константы c. Следовательно, для всех S таких, что $\mathrm{KP}(S) \leqslant \alpha$, имеем

$$\mathrm{Loss}_S(x) - \mathrm{KG}(x) \geqslant bn - (2\ln 2/\eta)\alpha - (\ln 2/\eta)(\log n + 2\log\log n) - c$$

для некоторой константы c.

Так как $\mathrm{KA}(x) \leqslant \mathrm{KP}(x) + c$ для некоторой константы c, имеем по (10.28)

$$M(x) \geqslant 2^{-\alpha - \log n - 2\log\log n - c}. \tag{10.29}$$

$\square$.

Заметим, что нижнюю оценку 2) априорной полумеры (10.29) из предложения 10.4 можно улучшить. Если в определении множества $D^n_{\alpha,\gamma}$ заменить $\mathrm{KP}(S) \leqslant \alpha$ на $\mathrm{KP}(S|n) \leqslant \alpha$, то вместо неравенства (10.28) будет выполнено $\mathrm{KP}(x) \leqslant \alpha + c$, тогда неравенство (10.29) заменяется на $M(x) \geqslant 2^{-\alpha+c}$. Оценка 1) также упрощается до $\mathrm{Loss}_P(x) - \mathrm{KG}(x) \geqslant bn - (2\ln 2/\eta)\alpha - c$. Соответственно, изменяются оценки предложения 10.4. Аналогичным образом изменяются оценки предложения 10.5. теоремы 10.2.

10.4.2. Доказательство предложения 10.5

.

Мы построим α-простую прогнозирующую стратегию в виде некоторой апроксимации универсальной "прогнозирующей полустратегии", задающей $\mathrm{KG}(x)$.

Предварительно докажем утверждение предложения для случая логарифмической функции потерь, а затем обобщим для произвольного случая. Рассмотрим определение (10.16) предсказательной сложности при $\beta = \frac{1}{2}$

$$\mathrm{KG}(x) = -\log \sum_{i=1}^{\infty} 2^{-\mathrm{KG}_i(x)} 2^{-\mathrm{KP}(i)}. \tag{10.30}$$

Также $\mathrm{KG}(x) = -\log M(x) + O(1)$, где M – априорная полумера.

Так как по предложению (10.2) функция $\beta^{\mathrm{KG}(x)}$ является полумерой, для любого n будет

$$\sum_{l(x)=n} \beta^{\mathrm{KG}(x)} \leqslant 1. \tag{10.31}$$

Пусть p – конечная двоичная последовательность, представляющая двоично-рациональное приближение действительного числа $\sum\limits_{l(x)=n} \beta^{\mathrm{KG}(x)}$ снизу с точностью до $2^{-\alpha}$. Используя p и n, можно эффективно найти натуральные числа t и k такие, что

1) $\sum\limits_{l(x)=n} \beta^{\mathrm{KG}^{t,k}(x)} > \sum\limits_{l(x)=n} \beta^{\mathrm{KG}(x)} - 2^{-\alpha}$, где $\mathrm{KG}^{t,k}(x) = \log_\beta \sum\limits_{i=1}^{k} \beta^{\mathrm{KG}_i^t(x)} 2^{-i}$ и $\mathrm{KG}_i^t(x)$ есть некоторое рациональное приближение сверху числа $\mathrm{KG}_i(x)$, вычисленное за t шагов.

2) для каждого x длины $\leqslant n$ и для каждого $i \leqslant k$ существует рациональное γ такое, что $\mathrm{KG}_i^t(xj) - \mathrm{KG}_i^t(x) \geqslant \lambda^t(j, \gamma)$ для всех $j = 0, 1$, где $\lambda^t(j, \gamma)$ – невозрастающая по t последовательность простых функций, для которой $\lambda(j, \gamma) = \inf_t \lambda^t(j, \gamma)$ для всех j и t.

Смесь функций, удовлетворяющих 2), также удовлетворяет этому условию (мы проверяли это в доказательстве предложения 10.1). Тогда разность $\mathrm{KG}^{t,k}(xj) - \mathrm{KG}^{t,k}(x)$ удовлетворяет 2) для всех x длины $\leqslant n$. Легко видеть, что в этом случае существует прогнозирующая стратегия Q такая, что

$$\mathrm{Loss}_Q(x) \leqslant \mathrm{KG}^{t,k}(x) \tag{10.32}$$

для всех x длины $\leqslant n$. Так как эта конструкция алгоритмически эффективная, будет

$$\mathrm{KP}(Q|n) \leqslant \alpha + c, \tag{10.33}$$

где c – положительная константа.

Временно, в определении $D^n_{\alpha,\gamma}$ мы будем рассматривать прогнозирующие стратегии S, которые являются α-простыми относительно n, т.е. такие, что $\mathrm{KP}(S|n) \leqslant \alpha$.

По определению, для любого $x \in D^n_{\alpha,\gamma}$ имеем

$$\mathrm{Loss}_S(x) - \mathrm{KG}(x) > \gamma$$

для каждой прогнозирующей стратегии S такой, что $\mathrm{KP}(S|n) \leqslant \alpha$.

Следовательно, по (10.32) и по пункту 1) выше для каждого $x \in D^n_{\alpha+c,\gamma}$ мы получили

$$\beta^\gamma \sum_{x \in D^n_{\alpha+c,\gamma}} \beta^{\mathrm{KG}(x)} > \sum_{x \in D^n_{\alpha+c,\gamma}} \beta^{\mathrm{Loss}_Q(x)} \geqslant$$
$$\sum_{x \in D^n_{\alpha+c,\gamma}} \beta^{\mathrm{KG}^{k,t}(x)} > \sum_{x \in D^n_{\alpha+c,\gamma}} \beta^{\mathrm{KG}(x)} - 2^{-\alpha},$$

где c такая, что (10.33) выполнено. Отсюда следует

$$(1 - \beta^\gamma) \sum_{x \in D^n_{\alpha+c,\gamma}} \beta^{\mathrm{KG}(x)} \leqslant 2^{-\alpha}. \tag{10.34}$$

В случае логарифмической функции потерь $\beta = \frac{1}{2}$ и $\beta^{\mathrm{KG}(x)}$ совпадает с $M(x)$ с точностью до мультипликативной константы. Тогда по (10.34) мы имеем

$$M(D^n_{\alpha+c,\gamma}) \leqslant c' 2^{-\alpha+1}$$

для каждого $\gamma \geqslant 1$, где c' – положительная константа. Отсюда

$$M(D^n_{\alpha,\gamma}) \leqslant 2^{-\alpha+c+1},$$

где c – положительная константа.

Для других видов предсказательной сложности сумма $\sum\limits_{l(x)=n} \beta^{\mathrm{KG}(x)}$ может быть больше единицы.[13]

В общем случае заменим неравенство (10.31) на

$$\sum_{l(x)=n} \beta_n^{\mathrm{KG}(x)} M(x) \leqslant 1,$$

где $\beta_n = e^{-\frac{1}{n}}$. Пусть p – конечная двоичная последовательность, представляющая двоично-рациональное приближение снизу числа $\sum\limits_{l(x)=n} \beta_n^{\mathrm{KG}(x)} M(x)$ с точностью до $2^{-\alpha}$. После этого, также как и выше, используя p и n, эффективно находим натуральные t и k такие, что выполнены условия 1') и 2), где

1') $\sum\limits_{l(x)=n} \beta_n^{\mathrm{KG}^{t,k}(x)} M^t(x) > \sum\limits_{l(x)=n} \beta_n^{\mathrm{KG}(x)} M(x) - 2^{-\alpha}$

используется вместо условия 1) выше. Здесь

$$\mathrm{KG}^{t,k}(x) = \log_\beta \sum_{i=1}^{k} \beta^{\mathrm{KG}^t_i(x)} 2^{-i},$$

где $\beta = e^{-\eta}$ и η такое, что наша функция потерь η-перемешиваемая, $\mathrm{KG}^t_i(x)$ – рациональное приближение к $\mathrm{KG}_i(x)$

[13] Например, можно проверить, что для квадратичной функции потерь эта сумма имеет порядок экспоненты от n.

сверху и $M^t(x)$ – рациональное приближение к $M(x)$ снизу, вычисленные за t шагов.

Существует прогнозирующая стратегия Q такая, что $\mathrm{Loss}_Q(x) \leqslant \mathrm{KG}^{t,k}(x)$ для всех x длины $\leqslant n$ и $\mathrm{KP}(Q|n) \leqslant \alpha + c$, где c – положительная константа. Тогда по 1') и $M(x) \geqslant M^t(x)$ получим

$$\sum_{l(x)=n} \beta_n^{\mathrm{Loss}_Q(x)} M(x) \geqslant \sum_{l(x)=n} \beta_n^{\mathrm{KG}^{t,k}(x)} M^t(x) > \sum_{l(x)=n} \beta_n^{\mathrm{KG}(x)} M(x) - 2^{-\alpha}. \qquad (10.35)$$

По определению для каждого $x \in D^n_{\alpha+c,\gamma}$ имеем

$$\mathrm{Loss}_Q(x) - \mathrm{KG}(x) > \gamma.$$

Следовательно, по (10.35) получаем

$$\beta_n^\gamma \sum_{x \in D^n_{\alpha+c,\gamma}} \beta_n^{\mathrm{KG}(x)} M(x) > \sum_{x \in D^n_{\alpha+c,\gamma}} \beta_n^{\mathrm{Loss}_Q(x)} M(x) > \sum_{x \in D^n_{\alpha+c,\gamma}} \beta_n^{\mathrm{KG}(x)} M(x) - 2^{-\alpha}.$$

Отсюда получаем

$$(1 - \beta_n^\gamma) \sum_{x \in D^n_{\alpha+c,\gamma}} \beta_n^{\mathrm{KG}(x)} M(x) \leqslant 2^{-\alpha}. \qquad (10.36)$$

По предложению 10.3 существует положительная константа c такая, что

$$\mathrm{KG}(x) \leqslant (\ln 2/\eta)\mathrm{KP}(x) + c.$$

Также выполнено $\mathrm{KP}(x) \leqslant n + 2\log n + c$ для всех x длины n, где c – положительная константа (см. предложение 5.1 из раздела 5.1.1). Огрубляя, получим $\mathrm{KG}(x) \leqslant cn$ для некоторого $c > 0$, где n – длина x.

Так как $\beta_n = e^{-\frac{1}{n}}$, имеем

$$\beta_n^{\mathrm{KG}(x)} = e^{-\frac{1}{n}\mathrm{KG}(x)} \geqslant e^{-c}$$

и

$$1 - \beta_n^{\gamma} \geqslant \frac{\gamma}{2n}$$

для всех $0 < \gamma \leqslant n$. Следовательно, по (10.36) получаем оценку

$$\sum_{x \in D^n_{\alpha+c,\gamma}} M(x) \leqslant 2^{-\alpha} \frac{2n}{\gamma} e^c = 2^{-\alpha + \log n - \log \gamma + c \log e + 1}. \qquad (10.37)$$

Для того, чтобы устранить условие n в $\mathrm{KP}(Q|n)$, используем следующие оценки для префиксной колмогоровской сложности:

$$\mathrm{KP}(Q) \leqslant \mathrm{KP}(Q|n) + \mathrm{K}(n) + c' \leqslant \alpha + \log n + 2 \log \log n + c$$

для некоторых положительных констант c' and c. Заменяя α в (10.37) на $\alpha - \log n - 2 \log \log n - c$ и возвращаясь к предыдущему (безусловному) определению $D^n_{\alpha,\gamma}$, мы получим необходимую оценку

$$\sum_{x \in D^n_{\alpha,\gamma}} M(x) \leqslant 2^{-\alpha + 2 \log n + 2 \log \log n - \log \gamma + c} \qquad (10.38)$$

для некоторой положительное константы c. □

Глава 11

Стохастичность конечных последовательностей

В этой главе будем изучать стохастические свойства конечных последовательностей. Это означает, что с точки зрения главы 10 мы рассматриваем только логарифмическую функцию потерь. В этом случае $\mathrm{KG}(x) = -\log M(x) + O(1)$, где $\mathrm{KG}(x)$ – предсказательная сложность конечной последовательности x, а $M(x)$ – априорная полумера на дереве всех двоичных последовательностей.

В разделе 11.2 мы введем понятие равномерного супермартингала. Предварительно в разделе 11.1 мы рассмотрим понятия перечислимой снизу (сверху) функции с более общей точки зрения.

11.1. Вычислимые вещественные функции

В этом разделе мы напомним и обобщим ранее рассмотренные в разделах 3.1.2 и 5.1.3 понятия, связанные с алгоритмами и вычислимыми операциями.

Конечные объекты естественным образом отождествляются со своими конструктивными представлениями. Поэтому в даль-

нейшем мы будем говорить о конструктивных объектах. В [26] было введено понятие ансамбля конструктивных объектов. Типичные примеры таких ансамблей: все слова в конечном алфавите, все конечные множества слов (в заданном алфавите) т.д. Ранее в разделе 3.1.2 установлено эффективное взаимнооднозначное соответствие между его элементами таких ансамблей и конечными двоичными последовательностями. Пример такого соответствия между всеми упорядоченными парами конечных двоичных последовательностями и всеми конечными двоичными последовательностями был приведен в разделе 3.1.2.

Напомним, что $\mathcal{N}$, $\mathcal{Z}$, $\mathcal{Q}$ и $\mathcal{R}$ – множества всех натуральных (включая ноль), целых, рациональных и вещественных чисел, соответственно, $\mathcal{R}_+$ – множество всех неотрицательных вещественных чисел. Символы $+\infty$ и $-\infty$ понимаются обычным образом, в частности, $\alpha < +\infty$ и $\alpha > -\infty$ для всех $\alpha \in \mathcal{R}$. Множества $\mathcal{N}$, $\mathcal{Z}$ и $\mathcal{Q}$ (но не $\mathcal{R}$ и $\mathcal{R}_+$), а также мнжество Ξ всех конечных двоичных (бинарных) последовательностей, являются ансамблями конструктивных объектов.

Элементы теории алгоритмов (рекурсивных функций) были рассмотрены в разделе 3.1.2. Алгоритмы реализуются в виде машин Тьюринга, поэтому понятия программы и шага вычисления точно определены.

Пусть A – некоторый ансамбль конструктивных объектов. Вещественнозначная функция $f : A \to \mathcal{R} \cup \{+\infty\}$ называется перечислимой снизу, если множество

$$\{(r, x) : r \in \mathcal{Q}, r < f(x)\}$$

перечислимо. Функция $f : A \to \mathcal{R} \cup \{-\infty\}$ называется перечислимой сверху, если множество

$$\{(r, x) : r \in \mathcal{Q}, r > f(x)\}$$

перечислимо.

Предложение 11.1. *Пусть A – ансамбль конструктивных объектов. Существует перечислимая снизу (сверху) вещественнозначная функция $f(i, a)$, где i – натуральное число и*

$a \in A$, такая, что последовательность функций $f_i(a) = f(i, a)$ состоит из всех перечислимых снизу (сверху) функций, определенных на множестве A (назовем такую функцию универсальной).

Доказательство. Докажем утверждение для перечислимых снизу функций $f : A \to \mathcal{R} \cup \{+\infty\}$. В разделе 3.1.2 была определена последовательность всех перечислимых множеств пар (r, a), где $r \in \mathcal{Q}$ и $a \in A$, такое, что множество $W = \{(i, r, a) : (r, a) \in W_i\}$ является перечислимым. Пусть W^s обозначает конечное подмножество множества W, состоящее из элементов перечисленных за s шагов. Определим функцию

$$f^s(i, a) = \max\{r : (i, r, a) \in W^s\} \cup \{-\infty\}.$$

Для любого s функция $f^s(i, a) = -\infty$ для всех пар (r, a), кроме может быть конечного их числа, также $f^s(i, a) \leqslant f^{s+1}(i, a)$. Определим $f(i, a) = \sup_s f^s(i, a)$. Так как $r < f(i, a)$ тогда и только тогда, когда $\exists s(r < f^s(i, a))$, функция $f(i, a)$ перечислима снизу.

Для любой перечислимой снизу функции $f(a)$ существует такое i, что $W_i = \{(r, a) : r < f(a)\}$. Тогда $f(i, a) = f(a)$ для всех $a \in A$.

Аналогично, в случае перечислимых сверху функций, определим

$$f^s(i, a) = \min\{r : (i, r, a) \in W^s\} \cup \{+\infty\}.$$

Для любого s функция $f^s(i, a) = +\infty$ для всех пар (r, a), кроме может быть конечного их числа, также $f^s(i, a) \geqslant f^{s+1}(i, a)$. Определим $f(i, a) = \inf_s f^s(i, a)$. Так как $r > f(i, a)$ тогда и только тогда, когда $\exists s(r > f^s(i, a))$, функция $f(i, a)$ перечислима сверху.

Для любой перечислимой сверху функции $f(a)$ существует такое i, что $W_i = \{(r, a) : r > f(a)\}$. Тогда $f(i, a) = f(a)$ для всех $a \in A$. $\square$

Будем называть функции, равные $-\infty$ (или $+\infty$) для почти всех a, простыми. Простые функции являются конструктивными объектами (и образуют ансамбль). Аналогичные построения можно провести для перечислимых сверху функций.

Всюду определенная функция $f : A \to \mathcal{R}$ называется вычислимой, если она перечислима снизу и сверху. Как было доказано в разделе 5.1.3, если функция f типа $A \to \mathcal{R}$ вычислимая, то существует алгоритм, который по любому a такому, что $f(a) \in \mathcal{R}$, и рациональному $\epsilon > 0$ выдает рациональное приближение к $f(a)$ с точностью до ϵ. Если $f(a) \in \mathcal{R}$ для всех $a \in A$, то верно и обратное утверждение.

Пусть $f^-(i,a)$ и $f^+(i,a)$ – функции, универсальные для всех перечислимых снизу и всех перечислимых сверху функций от a, соответственно. Для произвольной всюду определенной вычислимой функции $f : A \to \mathcal{R}$ назовем пару $\langle i,j\rangle$ (а точнее, ее двоичное представление) такую, что $f(a) = f^-(i,a) = f^+(j,a)$ для всех a, программой для вычисления f. При этом, число i называется программой перечисления функции f снизу, а число j – программой перечисления сверху. Заметим, что не каждая пара $\langle i,j\rangle$ задает вычислимую функцию.

Пару вычислимых функций $\Pi = \langle f^-, f^+\rangle$ назовем способом описания вычислимых функций типа $A \to \mathcal{R}$.

Пусть Φ – некоторый способ описания. Для произвольной вычислимой функции f типа $A \to \mathcal{R}$ определим ее сложность $\mathrm{K}_\Phi(f)$ как длину самой короткой двоичной записи пары $\langle i,j\rangle$, задающей эту функцию при данном способе описания Φ (т.е. $f(a) = f^-(i,a) = f^+(j,a)$).

Предложение 11.2. *Для любого ансамбля A существует оптимальный способ описания Π всех вычислимых функций типа $A \to \mathcal{R}$ со следующим свойством. Для любого способа описания Φ существует константа такая с такая, что $\mathrm{K}_\Pi(f) \leqslant \mathrm{K}_\Phi(f) + c$ выполнено что для любой вычислимой функции f типа $A \to \mathcal{R}$.*

Доказательство. Пусть $f^-(n,i,a)$ – функция, универсальная для всех перечислимых снизу функций от (i,a), $f^+(n,i,a)$ – функция, универсальная для всех перечислимых сверху функций от (i,a). Определим способ описания функций от a: $\tilde{f}^-(\langle n,i\rangle,a) = f^-(n,i,a)$, $\tilde{f}^+(\langle n,i\rangle,a) = f^+(n,i,a)$.

Для любого способа описания $(f^-(i,a), f^+(i,a))$ существуют m и n такие, что $f^-(i,a) = \tilde{f}^-(n,i,a)$, $f^+(i,a) = \tilde{f}^+(m,i,a)$. Утверждение следует из свойства принятого способа кодирования пар. $\square$

Заметим, что построенный в доказательстве предложения 11.2 оптимальный способ описания определяет также оптимальные способы описания для функций перечислимых сверху и для фунций перечислимых снизу.

Фиксируем некоторый оптимальный способ описания Π вычислимых функций типа $A \to \mathcal{R}$. В соответствии с общей схемой определения колмогоровской сложности определим сложность произвольной вычислимой функции f типа $A \to \mathcal{R}_+$ как длину наименьшей программы, задающей f при способе описания Π, т.е. $\mathrm{K}(f) = \mathrm{K}_\Pi(f)$.

Вычислимые функции типа $\Omega \to \mathcal{R}$. Функция f типа $\Omega \to \mathcal{R}$ называется простой, если пространство Ω представлено в виде объединения конечного множества шаров $\Gamma_{x_1}, \ldots, \Gamma_{x_k}$, на каждом из которых функция f постоянна и принимает рациональное значение или $+\infty$ $(-\infty)$. В этом случае можно писать $f(\omega) = f(x_i)$ для всех $\omega \in \Gamma_{x_i}$, $i = 1, \ldots k$, где $f(x_i)$ обозначает это постоянное значение функции f на шаре Γ_{x_i}.

Множество всех простых функций образуют ансамбль конструктивных объектов. Функция $f : \Omega \to \mathcal{R}$ называется вычислимой, если существуют две вычислимые последовательности простых функций $f_s^-(\omega)$ и $f_s^+(\omega)$ такие, что

$$f_s^-(\omega) \leqslant f_{s+1}^-(\omega),$$
$$f_s^+(\omega) \geqslant f_{s+1}^+(\omega)$$

для всех s и

$$f(\omega) = \sup_s f_s^-(\omega) = \inf_s f_s^+(\omega).$$

Функция $f : \Omega \to \mathcal{R}_+$ называется рекурсивно перечислимой снизу, если множество

$$\{(r,\omega) : r < f(\omega), r \in \mathcal{Q}\} \tag{11.1}$$

является эффективно открытым, т.е может быть представлено в виде объединения рекурсивно перечислимой последовательности множеств типа $\{r\} \times \Gamma_x$, где r – рационально и $x \in \Xi$. Аналогичным образом определяется понятие функции рекурсивно перечислимой сверху.

Предложение 11.3. *Функция $f(\omega)$ типа $\Omega \to \mathcal{R}$ вычислима тогда и только тогда, когда она перечислима снизу и сверху.*

Доказательство. Действительно, пусть $f(\omega)$ вычислимая. Тогда

$$r < f(\omega) \Longleftrightarrow \exists s : r < f_s^-(\omega) \Longleftrightarrow \exists s, x : r < f_s^-(x),\ x \subseteq \omega.$$

Аналогичным образом доказывается перечислимость $f(\omega)$ сверху.

Пусть теперь, $f(\omega)$ – перечислима снизу и сверху. Тогда множество (11.1) может быть представлено в виде $\cup_{(r,x)\in A}(\{r\} \times \Gamma_x)$, где A – некоторое рекурсивно перечислимое множество. Пусть A^s – часть множества A, перечисленная за s шагов перечисления. Определим неубывающую по s последовательность простых функций

$$f_s^-(\omega) = \max\{r : \exists x((r,x) \in A^s \& x \subseteq \omega)\}$$

(предполагаем, что $\max \emptyset = -\infty$). Из определения $f(\omega) = \sup_s f_s^-(\omega)$.

Аналогичным образом можно определить невозрастающую по s последовательность простых функций $f_s^+(\omega)$ такую, что $f(\omega) = \inf_s f_s^+(\omega)$. $\square$

Если $f(\omega)$ вычислима и $f(\omega) \in \mathcal{R}$, то того чтобы найти рациональное приближение к $f(\omega)$ с точностью до $\epsilon > 0$ (рациональное), достаточно найти $x \subset \omega$ и s такие, что $|f_s^+(x) - f_s^-(x)| < \epsilon$, и выдать $f_s^-(x)$ в качестве результата.

Как легко следует из определения, вычислимая функция типа $\Omega \to \mathcal{R}$ непрерывна в каждой точке ω.

Вычислимые функции типа $\mathcal{R} \to \mathcal{R}$ определяются аналогичным образом. Функция f типа $\mathcal{R} \to \mathcal{R}$ называется простой, если множество $\mathcal{R}$ представлено в виде объединения конечного множества интервалов $\Gamma_1, \ldots, \Gamma_k$, на каждом из которых функция f постоянна и принимает рациональное значение, $+\infty$ или $-\infty$. В этом случае можно писать $f(x) = f_i$ для всех $x \in \Gamma_i$, $i = 1, \ldots k$, где f_i обозначает это постоянное значение функции f на шаре Γ_x.

11.2. Равномерные тесты случайности

Сформулируем определение равномерного супермартингала в форме, приспособленной для прогнозирующих стратегий.

Задана система программирования, при которой каждая вычислимая функция $f(x)$ типа $\Xi \to R_+$ имеет программу для вычисления ее значений. Эта программа, в зависимости от контекста, закодирована в виде конечной двоичной последовательности p (в виде натурального числа). Однако не каждая двоичная последовательность задает некоторую функцию $f(x)$.

Неотрицательная функция $\psi(p, x)$ называется равномерным перечислимым снизу супермартингалом, если выполнено следующее:

1) Множество $\{(r, p, x) : r < \psi(p, x), r \in \mathcal{Q}\}$ является рекурсивно перечислимым,

2) $\psi(p, \lambda) \leqslant 1$ для любого p,

3) если p является программой для вычисления значения прогнозирующей стратегии f, то $\psi(p, x) \geqslant \psi(p, x0)(1 - f(x)) + \psi(p, x1)f(x)$ для всех x.

Предложение 11.4. *Существует универсальный равномерный супермартингал $\hat{\psi}$ такой, что для любого равномерного супермартингала ψ существует такая константа c, что $c\hat{\psi}(p, x) \geqslant \psi(p, x)$ для всех $x, p \in \Xi$.*

Доказательство. По предложению 11.1 существует функция $f(n,p,x)$ универсальная для всех перечислимых снизу функций типа $f : \mathcal{N} \times \Xi \to \mathcal{R} \cup \{+\infty\}$. По определению перечислимости снизу существует вычислимая последовательность простых функций $f_s(n,p,x)$ такая, что $f_s(n,p,x) \leqslant f_{s+1}(n,p,x)$ для всех s и $f(n,p,x) = \sup_s f_s(n,p,x) = \lim_{s\to\infty} f_s(n,p,x)$ для любых p и x. Полагаем $f_n(p,x) = f(n,p,x)$ и $f_{n,s}(p,x) = f_s(n,p,x)$.

Для произвольных n, s определим равномерный супермартингал $\psi(i,x)$ следующим образом:

Пусть $\langle f^-(i,x), f^+(j,x)\rangle$ – способ описания вычислимых функций из Ξ в R_+. Обозначим $f_i^-(x) = f^-(i,x)$, $f_{i,s}^-(x) = f_s^-(i,x)$ и $f_j^+(x) = f^+(j,x)$, $f_{j,s}^+(x) = f_s^+(j,x)$. Пусть π и τ – функции нумерации пар такие, что $\pi(\langle i,j\rangle) = i$ и $\tau(\langle i,j\rangle) = j$ для всех натуральных чисел i и j. Будем также говорить, что число k задает способ описания $\langle f^-(\pi(k),x), f^+(\tau(k),x)\rangle$.

Для любой простой функции f определим $\tilde{f}(x) = 0$, если $f(x) \leqslant 0$, $\tilde{f}(x) = 1$, если $f(x) \geqslant 1$, и $\tilde{f}(x) = f(x)$, в противном случае.

Временно фиксируем индексы n и s, а также функцию $f_{n,s}$. Для произвольного натурального числа t пусть ψ_t^1 и ψ_t^2 – минимальные неотрицательные функции такие, что $\psi_t^1(i,x) \geqslant f_{n,s}(i,x)$, $\psi_n^2(i,x) \geqslant f_{n,s}(i,x)$ для всех i и x, а также выполнено

$$\psi_t^1(i,x) \geqslant \psi_t^1(i,x0)(1 - \tilde{f}^+_{\tau(i),t}(x)) + \psi_t^1(i,x1)\tilde{f}^-_{\pi(i),t}(x) \qquad (11.2)$$

$$\psi_t^2(i,x) \geqslant \psi_t^2(i,x0)(1 - \tilde{f}^-_{\tau(i),t}(x)) + \psi_t^2(i,x1)\tilde{f}^+_{\pi(i),t}(x). \qquad (11.3)$$

Так как $f_{n,s}(i,x) = -\infty$ для почти всех пар (i,x), такие ψ_t^1 и ψ_t^2 легко могут быть определены. Заметим, что эти функции принимают в качестве значений только рациональные числа и равны 0 для почти всех пар (i,x). Из $f^+_{\pi(i),t+1}(x) \leqslant f^+_{\pi(i),t}(x)$ и $f^-_{\tau(i),t+1}(x) \geqslant f^-_{\tau(i),t}(x)$ следует, что $\psi^1_{t+1}(i,x) \geqslant \psi^1_t(i,x)$ и $\psi^2_{t+1}(i,x) \leqslant \psi^2_t(i,x)$ для всех t, i, x. Кроме этого, $\psi^1_t(i,x) \leqslant \psi^2_t(i,x)$ для всех t. Полагаем

$$\psi^1(i,x) = \sup_t \psi^1_t(i,x) = \lim_{t\to\infty} \psi^1_t(i,x).$$

По определению эта функция перечислима снизу. Устремляя в (11.2) и (11.3) t к ∞, получим

$$\psi^1(i,x) \geqslant \psi^1(i,x0)(1-\tilde{f}^+_{\tau(i)}(x)) + \psi^1(i,x1)\tilde{f}^-_{\pi(i)}(x)$$
$$\psi^2(i,x) \geqslant \psi^2(i,x0)(1-\tilde{f}^-_{\tau(i)}(x)) + \psi^2(i,x1)\tilde{f}^+_{\pi(i)}(x)$$

для всех i, x. Определим также $\psi^2(i,x) = \lim\limits_{t\to\infty} \psi^2_t(i,x)$.

Пусть i – задание вычислимой прогнозирующей стратегии. Тогда $f(x) = f^-_{\pi(i)}(x) = f^+_{\tau(i)}(x)$. Отсюда $\psi^1(i,x) = \psi^2(i,x)$.

Только что определенная функция $\psi^1(i,x)$ зависит от n и s. В связи с этим, переобозначим $\psi^1_{n,s} = \psi^1$ и $\psi^1_{n,s,t} = \psi^1_t$, а также $\psi^2_{n,s} = \psi^2$ и $\psi^2_{n,s,t} = \psi^2_t$. Так как $f_{n,s+1} \geqslant f_{n,s}$ имеем $\psi^1_{n,s+1} \geqslant \psi^1_{n,s}$ для всех s. Определим

$$\psi_n(i,x) = \sup_s\{\psi^1_{n,s}(i,x) : \exists t(\psi^2_{n,s,t}(i,x) \leqslant 2)\}.$$

Мы полагаем $\psi^1_n(i,x) = 0$ для всех x, если $\psi^2_{n,s,t}(i,x) > 2$ для всех s, t.

Определим

$$\hat{\psi}(i,x) = \sum_{n=1}^{\infty} \frac{1}{n(n+1)}\psi_n(i,x).$$

Легко видеть, что $\hat{\psi}$ – равномерный супермартингал.

Пусть ψ – произвольный равномерный супермартингал. Тогда существует n такое, что $\psi(i,x) = f_n(i,x)$. Пусть i – программа вычислимой прогнозирующей стратегии. Тогда для любого s будет $\psi^2_{n,s,t}(i,x) \leqslant 2$ для всех достаточно больших t. Следовательно, $\psi(i,x) = \psi_n(i,x)$ и $n(n+1)\hat{\psi}(i,x) \geqslant \psi(i,x)$ для всех x. Предложение доказано. □

Для перехода к логарифмической шкале определим равномерный дефект случайности следующим образом:

$$\hat{d}(p,x) = \log \hat{\psi}(p,x).$$

Функция $\psi(p,x)$ перечислима снизу и определена для всех двоичных строк p, даже если p не является программой для вычисления значений вычислимой меры. Если p является программой для вычисления значений вычислимой меры P, то из результата раздела 6.1 следует, что

$$\hat{d}(p,x) = \log \frac{M(x)}{P(x)} + O(1) = -\log P(x) - \mathrm{KA}(x) + O(1),$$

где M – априорная полумера на дереве всех двоичных последовательностей. Здесь константы $O(1)$ могут зависеть от P.

11.3. (α,β)-стохастические конечные последовательности

Пусть α и β – натуральные числа. По определению α-простой называется функция, значения которой могут быть вычислены с помощью программы длины $\leqslant \alpha$. Конечная последовательность x называется (α,β)-стохастической, если для некоторой α-простой прогнозирующей стратегии f выполнено $\hat{d}(p,x^j) \leqslant \beta$ для всех j, $1 \leqslant j \leqslant l(x)$, где p – некоторая программа длины $\leqslant \alpha$ для вычисления значений f.

Приводимое определение стохастичности представляет собой некоторое обобщение соответствующего определения (α,β)-стохастичности по Колмогорову.

Последовательность x называется (α,β)-нестохастической, если она не является (α,β)-стохастической. Это эквивалентно тому, что для любой программы p длины $l(p) \leqslant \alpha$, вычисляющей значения α-простой прогнозирующей стратегии, будет $\hat{d}(p,x^m) > \beta$ хотя бы для одного $m \leqslant l(x)$. Мы будем также говорить в этом случае, что любая прогнозирующая программа длины $\leqslant \alpha$ отвергается на x при уровне доверия β.

Пусть $D^n_{\alpha,\beta}$ – множество всех (α,β)-нестохастических последовательностей длины n.[1]

[1]Это определение отличается от определения множества $D^n_{\alpha,\beta}$, которое

Следующее простое утверждение выражает тот естественный факт, что вероятность генерации (α, β)-нестохастической последовательности в α-простом вероятностном процессе мала.

Предложение 11.5. *Для любых α, β, n и α-простой меры P выполнено $P(D^n_{\alpha,\beta}) < 2^{-\beta}$.*

Доказательство. Пусть p – программа длины $\leqslant \alpha$ для вычисления значений прогнозирующей стратегии f. Для любого $x \in D^n_{\alpha,\beta}$ существует начальный фрагмент $x' \subseteq x$ такой, что $\hat{\psi}(p, x') > 2^{\beta}$. Пусть $x_1, \ldots, x_s$ – все такие начальные фрагменты последовательностей из $D^n_{\alpha,\beta}$, имеющие максимальную длину. По определению, все они попарно несравнимы. Поэтому из определения супермартингала следует, что

$$1 \geqslant \sum_{j=1}^{s} \hat{\psi}(p, x_j) P(x_j) > 2^{\beta} \sum_{j=1}^{s} P(x_j).$$

Отсюда получаем $P(D^n_{\alpha,\beta}) \leqslant \sum\limits_{j=1}^{s} P(x_j) < 2^{-\beta}$. $\square$

Теорема 10.2 (и предложение 10.5) также имеет место для модифицированного определения (α, β)-нестохастической последовательности в случае логарифмической функции потерь. Нижняя оценка

$$M(D^n_{\alpha,\gamma}) \geqslant 2^{-\alpha - \log n - 2\log\log n - c}$$

прямо следует из нижней оценки теоремы 10.2. Верхняя оценка

$$M(D^n_{\alpha,\beta}) \leqslant 2^{-\alpha + \log n + 2\log\log n - \log\beta + c},$$

где c – константа, может быть получена аналогичным образом как для предложения 10.4.2.

Заметим, что логарифмы от n в этих оценках можно устранить, если использовать длину последовательностей n в качестве параметра прогнозирующих стратегий. Это соответствует случаю, когда задан конечный "горизонт" прогнозирования.

было дано в разделе 10.4. Легко видеть, что оно приводит к более широкому множеству (α, β)-нестохастических последовательностей.

Пусть $\alpha(n)$ и $\beta(n)$ – вычислимые неубывающие неограниченные функции, принимающие натуральные значения. Определим

$$I^n_{\alpha,\beta} = \bigcup_{i=n}^{\infty} D^i_{\alpha(i),\beta(i)}.$$

По определению, множество $I^n_{\alpha,\beta}$ состоит из всех последовательностей, имеющих длину m и являющихся $(\alpha(m), \beta(m))$-нестохастическими, для некоторого $m \geqslant n$.

Величина $P(I^n_{\alpha,\beta})$ равна вероятности генерации на соответствующей полумере P вероятностной машине $(\alpha(m), \beta(m))$-нестохастической последовательности длины $m \geqslant n$.

Теорема 11.1. *Пусть $\alpha(n)$ и $\beta(n)$ – две вычислимые неубывающие неограниченные функции, принимающие натуральные значения. Для любой рекурсивно перечислимой снизу полумеры P*

$$\lim_{n\to\infty} P(I^n_{\alpha,\beta}) = 0.$$

Доказательство. Для произвольного рационального $\epsilon > 0$ выберем вычислимую последовательность натуральных чисел

$$0 = n_0 < n_1 < \ldots$$

такую, что

$$\lfloor (1-\epsilon)\alpha(n_i) \rfloor < \lfloor (1-\epsilon)\alpha(n_{i+1}) \rfloor$$

и $\beta(n_i) < \beta(n_{i+1})$ для всех i.

Если P – рекурсивно перечислима снизу, то определим

$$P^s(x) = \max\{r : (r, x) \in A^s\},$$

где $A = \{(r, x) : r \in \mathcal{Q}, r < P(x)\}$ и A^s – конечная часть множества A, перечисленная за s шагов работы соответствующего алгоритма.

Для произвольной конечной последовательности q определим меру P_q следующим образом. Если найдутся i и s такие, что

$l(q) = \lfloor (1-\epsilon)\alpha(n_i) \rfloor$ и

$$\sum_{l(z)=n_{i+1}} P^s(z) \geqslant \sum_{j=1}^{l(q)} q_j 2^{-j}, \tag{11.4}$$

полагаем $P_q(x) = P^s(x)$ при $l(x) = n_{i+1}$ и $x \neq 0^{n_{i+1}}$, где s – минимальное, удовлетворяющее (11.4). Доопределим P_q на остальных x каким либо естественным образом так, чтобы сохранились условия меры. Мы будем использовать только значения $P_q(x)$ при $l(x) \leqslant n_{i+1}$. Если такие i и s не существует, то мера P_q не определена. Пусть

$$r_n = \sum_{l(x)=n} P(x).$$

Так как P является полумерой, будет $r_{n+1} \leqslant r_n$ для всех n. Если $\lim\limits_{n\to\infty} r_n = 0$, то утверждение теоремы выполнено. Допустим противное.

Пусть p_i – двоичная последовательность, представляющая двоично-рациональное приближение числа $r_{n_{i+1}}$ снизу с точностью $2^{-\lfloor (1-\epsilon)\alpha(n_i) \rfloor}$. Тогда P_{p_i} – мера. Обозначаем $P_i = P_{p_i}$.

По определению, прогнозирующая стратегия, соответствующая мере P_i, является $(\lfloor (1-\epsilon)\alpha(n_i) \rfloor + c)$-простой, для некоторой константы c. Имеем $\lfloor (1-\epsilon)\alpha(n_i) \rfloor + c \leqslant \alpha(n_i)$ для всех достаточно больших i. В дальнейшем мы будем рассматривать только такие i.

Пусть множество X состоит из попарно несравнимых и не состоящих только из нулей конечных последовательностей длины $\leqslant n_{i+1}$. Определим

$$R_j = \sum \{P(x) : l(x) = n_{j+1} \& \exists z (z \subseteq x \& z \in X)\}.$$

Имеем $R_s \leqslant R_i$ при $s \geqslant i$. Так как $P_j(X) = R_j - \epsilon_j$ при $j = i, s$, где $\epsilon_j \leqslant 2^{-\lfloor (1-\epsilon)\alpha(n_j) \rfloor}$, имеем при $s \geqslant i$

$$P_s(X) \leqslant P_i(X) + 2^{-\lfloor (1-\epsilon)\alpha(n_i) \rfloor}.$$

Пусть конечное множество D состоит из попарно несравнимых конечных последовательностей, каждая из которых $(\alpha(n), \beta(n))$-нестохастична, где n – ее длина, и не состоит из одних нулей. Мы будем называть такое D (α, β)-нестохастическим сечением.

Пусть D – некоторое (α, β)-нестохастическое сечение. Определим

$$A_i = \{x \in D : n_i < l(x) \leqslant n_{i+1}\}.$$

Пусть $s = s(D)$ – максимальное такое s, что $n_s < l(z)$ для всех $z \in D$, а r – минимальное такое, что $l(z) \leqslant n_{r+1}$ для всех $z \in D$.

Заметим, что для произвольного i мера P_i (и соответствующая ей прогнозирующая стратегия[2]) является $\alpha(n)$-простой при $n_i < n \leqslant n_{i+1}$, так как $\alpha(n_i) \leqslant \alpha(n)$ для таких n.

Пусть i – произвольное. Из определения нестохастичности следует, что для любого $x \in A_i$ существует префикс $x' \subseteq x$ такой, что $\hat{\psi}(p_i, x') > 2^{\beta(n_i)}$. Пусть x' – такой префикс x, имеющий максимальную длину. Пусть также $x_1, \ldots, x_t$ – все такие префиксы у всех $x \in A_i$. Ясно, что все они попарно несравнимы. Также, из определения супермартингала следует, что

$$\sum_{j=1}^{t} \hat{\psi}(p_i, x_j) P_i(x_j) \leqslant 1.$$

Отсюда следует, что

$$P_i(A_i) \leqslant \sum_{j=1}^{t} P_i(x_j) \leqslant 2^{-\beta(n_i)}.$$

Отсюда получаем

$$P_r(D) = \sum_{i=s}^{r} P_r(A_i) \leqslant \sum_{i=s}^{r} P_i(A_i) + \sum_{i=s}^{r} 2^{-\lfloor (1-\epsilon)\alpha(n_i) \rfloor} \leqslant 2^{-\beta(n_s)+1} + 2^{-\lfloor (1-\epsilon)\alpha(n_s) \rfloor + 1}. \quad (11.5)$$

[2]Произвольной мере P соотвествует прогнозирующая стратегия $f(x_1 \ldots x_{i-1}) = P(x_1 \ldots x_{i-1}1)/P(x_1 \ldots x_{i-1})$.

Пусть $\tilde{D}$ – множество всех последовательностей длины n_{r+1}, продолжающих последовательности из D. Из (11.5) следует

$$P(\tilde{D}) \leqslant P_r(\tilde{D}) + 2^{-\lfloor(1-\epsilon)\alpha(n_r)\rfloor} \leqslant$$
$$2^{-\beta(n_s)+1} + 2^{-\lfloor(1-\epsilon)\alpha(n_s)\rfloor+1} + 2^{-\lfloor(1-\epsilon)\alpha(n_r)\rfloor}. \qquad (11.6)$$

Допустим, что

$$\lim_{n\to\infty} P(I^n_{\alpha,\beta}) > 0.$$

Тогда существует $h > 0$ такое, что для любого n существует (α,β)-нестохастическое сечение D, для которого $s(D) > n$, $P(\tilde{D}) > h$ и, кроме этого, для любого $x \in D$ будет $x \not\subseteq 0^\infty$. Существование такого сечения является противоречием, так как по (11.6) мера $P(\tilde{D})$ должна быть как угодно мала всех для достаточно больших n. Теорема доказана. □

Для произвольного рационального $\epsilon > 0$ назовем вычислимую последовательность натуральных чисел $0 = k_0 < k_1 < \ldots$ такую, что $\lfloor(1-\epsilon)\alpha(k_i)\rfloor < \lfloor(1-\epsilon)\alpha(k_{i+1})\rfloor$ и $\beta(k_i) < \beta(k_{i+1})$ для всех i, ϵ-остовом пары функций $\alpha(n)$ и $\beta(n)$. Тогда из оценки (11.5) получаем

Следствие 11.1. *Для любой рекурсивно перечислимой снизу полумеры P, ϵ-остова $0 = k_0 < k_1 < \ldots$ и пары функций $\alpha(n)$ и $\beta(n)$ имеем*

$$P(D^n_{\alpha(n),\beta(n)}) \leqslant 2^{-\beta(k(n))+1} + 2^{-(1-\epsilon)\alpha(k(n))+2}$$

для всех достаточно больших n, где $k(n) = \max\{k_i : k_i \leqslant n\}$.

Полумеры на дереве всех двоичных последовательностей и связанные с ними меры. Пусть P – полумера на дереве всех двоичных последовательностей.[3] Определим функцию

$$\bar{P}(x) = \inf_{n\geqslant l(x)} \sum_{x\subseteq z, l(z)=n} P(z). \qquad (11.7)$$

[3]Определение перечислимой снизу полумеры на дереве всех двоичных последовательностей приведено в разделе 6.1.

Легко видеть, что $\bar{P}(x) = \bar{P}(x0) + \bar{P}(x1)$ для всех x и $\bar{P}(\lambda) \leqslant 1$. Кроме этого, $\bar{P}$ является максимальной мерой не превосходящей P, т.е. такой, что $\bar{P}(x) \leqslant P(x)$ для всех x.

Мы не можем утверждать, что $\bar{P}$ – вычислимая функция, а также, что она вероятностная мера, так как может быть $\bar{P}(\lambda) < 1$.

Определим $\bar{P}(\Gamma_x) = \bar{P}(x)$ для всех $x \in \Xi$, где $\Gamma_x = \{\omega \in \Omega : x \subseteq \omega\}$ – интервал (шар) в пространстве Ω. Пользуясь теоремой Колмогорова о распространении меры (см. [25]), можно определить меру $\bar{P}(A)$ для любого борелевского (измеримого) подмножества $A \subseteq \Omega$.

Оценки для бесконечных последовательностей. Для любого множества D конечных последовательностей определим

$$\tilde{D} = \{\omega \in \Omega : \exists z(z \in D \& z \subset \omega)\}.$$

Множество

$$I_{\alpha,\beta} = \cap_n \tilde{I}^n_{\alpha,\beta}$$

состоит из (бесконечных) последовательностей $\omega \in \Omega$, имеющих $(\alpha(n), \beta(n))$-нестохастические фрагменты длины n для бесконечно многих n.

Из теоремы 11.1 легко следует

Следствие 11.2. *Пусть $\alpha(n)$ и $\beta(n)$ – две вычислимые неубывающие неограниченные функции, принимающие натуральные значения. Тогда $\bar{M}(I_{\alpha,\beta}) = 0$.*

Если отказаться от требования вычислимости функции $\alpha(n)$, то получим противоположный утверждению теоремы 11.1 результат, который приводится в следующей теореме.

Теорема 11.2. *Пусть $\beta(n)$ – неограниченная вычислимая функция такая, что для некоторого $\gamma > 0$ выполнено неравенство $\beta(n) \leqslant (1-\gamma)n$ для всех n. Тогда для любого $\epsilon > 0$ существует перечислимая сверху неограниченная функция $\alpha(n)$ такая, что*

$$\bar{P}(I_{\alpha,\beta}) > 1 - \epsilon$$

для некоторой перечислимой снизу полумеры P. В частности, $\bar{M}(I_{\alpha,\beta}) > 0$.

Доказательство этой теоремы проводится на основе общей схемы раздела 12.2 и приводится в разделе 12.4.

Для произвольных натуральных чисел α и β множество $D^n_{\alpha,\beta}$ состоит из таких конечных последовательностей длины n, на которых любая прогнозирующая программа длины $\leqslant \alpha$ будет отвергнута на уровне доверия β. Теорема 11.1 показывает, в том случае, когда $\alpha = \alpha(n)$ и $\beta = \beta(n)$ вычислимо зависят от n и неограничены, вероятность генерации такой последовательности длины n на произвольной вероятностной машине стремится к 0.

Согласно теореме 11.2 для произвольно малого $\epsilon > 0$ на некоторой вероятностной машине можно с вероятностью $1 - \epsilon$ сгенерировать неограниченную последовательность, на которой при произвольном заданном уровне доверия β любая прогнозирующая программа когда-либо будет отвергнута на этом уровне.

Часть V

Степени рандомизированной вычислимости

Глава 12

Алгоритмически-инвариантные свойства бесконечных последовательностей

В этой главе проблема стохастических и нестохастических носителей информации рассматривается в информационном аспекте. Мы будем изучать свойства бесконечных двоичных последовательностей, как носителей определенной информации, т.е. такие свойства должны сохраняться при перекодировании их носителей. Мы будем рассматривать способы кодирования самого общего вида – алгоритмические операторы, определенные в разделе 11.1.

12.1. Алгебра инвариантных свойств

Бесконечная последовательность $\alpha \in \Omega$ алгоритмически сводится к бесконечной последовательности $\beta \in \Omega$, если $\alpha = F(\beta)$ для некоторого алгоритмического оператора F. Две бесконечные последовательности α и β алгоритмически эквивалентны,

обозначается это как $\alpha \equiv \beta$, если каждая из них сводится к другой. Множество (свойство) бесконечных последовательностей A называется алгоритмически инвариантным, если оно вместе с каждой последовательностью содержит и все алгоритмически эквивалентные ей последовательности. Иными словами, множество A представляется в виде объединения тьюгинговых степеней [32]. Для произвольного множества $A \subseteq \Omega$ обозначим $\bar{A} = \{\omega : \exists\alpha(\alpha \in A \& \alpha \equiv \omega)\}$ алгоритмическое замыкание множества A.

Случайные последовательности должны служить в качестве математических аналогов последовательностей, которые получаются в стохастических процессах. С другой стороны, некоторые бесконечные случайные по Мартин-Лефу последовательности можно определить с помощью точных математических конструкций. Например, двоичная запись α числа Чейтина $\sum_n 2^{-\mathrm{KP}(n)}$ является последовательностью случайной по Мартин-Лефу относительно равномерной меры (см. задачу из раздела 5.4).

Этот и другие примеры (см. также пример из [10]) показывают, что необходима некоторая корректировка интерпретации понятия случайной последовательности.

Кроме вероятностных процессов рассматриваются алгоритмические преобразования стохастических последовательностей, которые могут осуществляться с помощью вероятностных машин Тьюринга. Согласно теореме 6.3, $\bar{M}(\{\alpha\}) = 0$, где α – любая невычислимая последовательность. В частности, число Чейтина, а точнее, его двоичное представление α, не может быть выдано (с положительной вероятностью) никакой вероятностной машиной Тьюринга. Аналогично, случайная последовательность α, определенная математической конструкцией, не может быть получена ни в каких комбинациях случайных и алгоритмических процессов.

В общем случае, пусть некоторое свойство $\mathcal{A}$ определяет борелевское множество $A = \{\omega \in \Omega : \mathcal{A}(\omega)\}$ такое, что $\bar{M}(A) = 0$. Тогда для любой вероятностной машины Тьюринга (L, F) вероятность $P(A) = L\{\omega : F(\omega) \in A\}$ генерации с помощью этой

машины последовательности из множества A равна 0. Такие множества называются пренебрежимыми. Как следует из задач раздела 6.3, множество A пренебрежимо тогда и только тогда, когда $L(F^{-1}(A)) = 0$ для любого вычислимого оператора F, где $F^{-1}(A) = \{\omega \in \Omega : F(\omega \in A\}$.

Возможна такая интерпретация свойства пренебрежимости: бесконечные последовательности, принадлежащие пренебрежимому множеству, нельзя (с положительной вероятностью) получать ни в каких комбинациях вероятностных и алгоритмических процессов.

Как замечено выше, для любой невычислимой бесконечной последовательности α множество

$$\{\omega \in \Omega : \exists F(F(\omega) = \alpha)\} = \{\omega \in \Omega : \cup_F(F(\omega) = \alpha)\}$$

является пренебрежимым.

Рассмотрим следующую математическую структуру, которая была введена и изучалась в [47] и [3]. Пусть I – булева алгебра всех алгоритмически инвариантных борелевских подмножеств Ω. Мы не различаем два множества из I, которые отличаются на множестве априорной полумеры 0, точнее, рассматривается отношение эквивалентности

$$A \sim B \Longleftrightarrow \bar{M}((A \setminus B) \cup (B \setminus A)) = 0.$$

Пусть Υ – фактор алгебра алгебры I по отношению эквивалентности $\sim$. Изучим структуру этой фактор-алгебры Υ. Класс эквивалентности алгоритмически инвариантного множества A обозначается $\mathbf{a} = [A]$. Для любой перечислимой полумеры P можно корректным образом определить $\bar{P}(\mathbf{a}) = \bar{P}(A)$. Определим операции $\mathbf{a} \cup \mathbf{b} = [A] \cup [B]$ и $\mathbf{a} \cap \mathbf{b} = [A] \cap [B]$, где $\mathbf{a} = [A]$ и $\mathbf{b} = [B]$.

Можно также рассмотреть частичный порядок на Υ:

$$\mathbf{a} \preceq \mathbf{b} \Longleftrightarrow \bar{M}(A \setminus B) = 0.$$

В дальнейшем последовательности, алгоритмически эквивалентные последовательностям, случайным относительно вычислимых мер, будут называться стандартными. По определению

любая вычислимая мера множества всех стандартных последовательностей равна 1.

В разделе 12.9 (см. также ([10], теорема 3.1)) доказано, что любая последовательность ω, случайная относительно какой-либо вычислимой меры, либо сама вычислима, либо алгоритмически эквивалентна последовательности, случайной относительно равномерной меры. Поэтому естественным образом возникают элементы $\mathbf{r} = [\bar{R}]$ и $\mathbf{c} = [C]$ алгебры Υ, где R – множество всех последовательностей, алгоритмически случайных по равномерной мере (мере Лебега), C – множество всех вычислимых последовательностей.[1]

Из свойств априорной меры легко следует, что $\bar{M}(\mathbf{r}) > 0$ и $\bar{M}(\mathbf{c}) > 0$.

Нулевой элемент $\mathbf{0}$ алгебры Υ определяется как класс эквивалентности пустого множества. Он состоит из всех алгоритмически инвариантных подмножеств Ω априорной меры 0, поэтому $\bar{M}(\mathbf{0}) = 0$. Единица алгебры Υ определяется $\mathbf{1} = [\Omega]$.

Л.А.Левин [16, 17] заметил, что множество всех случайных по Мартин-Дефу последовательностей (относительно различных вычислимых мер) может быть разделено только на два алгоритмически инвариантных подмножества положительной априорной меры. Первое из них состоит из всех невычислимых случайных последовательностей, второе – это все вычислимые последовательности. Каждая вычислимая последовательность является случайной относительно вычислимой меры, сосредоточенной на этой последовательности.

Напомним, что элемент $\mathbf{d}$ является атомом Υ, если $\mathbf{d} \neq \mathbf{0}$ и не существует разложения $\mathbf{d} = \mathbf{a} \cup \mathbf{b}$, где $\mathbf{a} \cap \mathbf{b} = \emptyset$, $\mathbf{a} \neq \mathbf{0}$ и $\mathbf{b} \neq \mathbf{0}$.

Теорема 12.1. *Элемент* $\mathbf{r}$ *является атомом* Υ.

Доказательство. Допустим, что $\mathbf{r} = \mathbf{a} \cup \mathbf{b}$, где $\mathbf{a} \cap \mathbf{b} = \mathbf{0}$, $\mathbf{a} \neq \mathbf{0}$ и $\mathbf{b} \neq \mathbf{0}$. Тогда $\bar{R} = A \cup B$, где $\mathbf{a} = [A]$ и $\mathbf{b} = [B]$, где A и B – алгоритмически инвариантные множества бесконечных

[1] Множество $\bar{R}$ содержит все невычислимые последовательности, алгоритмически случайные по какой-либо вычислимой мере.

последовательностей. Без потери общности можно считать, что $A \cap B = \emptyset$. Напомним, что R – множество всех случайных по мере Лебега последовательностей. Пусть $A' = A \cap R$ и $B' = B \cap R$. Так как любая $\alpha \in A$ алгоритмически эквивалентна некоторой последовательности из A', из $\bar{M}(A) > 0$ следует $\bar{M}(A') > 0$. Аналогично $\bar{M}(B') > 0$.

Пусть P – произвольная вероятностная мера на Ω. Рассмотрим производную Радона – Никодима $\frac{dP}{d\bar{M}}(\omega)$ меры P по мере $\bar{M}$ (см. [9]).

Лемма 12.1. *Пусть $A \subseteq \Omega$ и для любой $\omega \in A$ выполнено $\frac{dP}{d\bar{M}}(\omega) > 0$. Тогда из $P(A) = 0$ следует $\bar{M}(A) = 0$.*

Доказательство. Из соотношения

$$0 = P(X) = \int\limits_X \frac{dP}{d\bar{M}}(\omega) d\bar{M}$$

и того, что $\frac{dP}{d\bar{M}}(\omega) > 0$ при всех $\omega \in A$, легко следует, что $\bar{M}(A) = 0$. □

Следствие 12.1. *Пусть P – вычислимая мера и A состоит из случайных по мере P последовательностей. Тогда из $P(A) = 0$ следует $\bar{M}(A) = 0$.*

Доказательство. Для любой случайной ω выполнено

$$P(\omega^n)/\bar{M}(\omega^n) \geqslant P(\omega^n)/M(\omega^n) \geqslant c > 0$$

для всех n, где c – константа, зависящая от ω. Отсюда $\frac{dP}{d\bar{M}}(\omega) \neq 0$ для $\omega \in A$. По лемме 12.1 $\bar{M}(A) = 0$. □

Продолжим доказательство теоремы. Как было замечено ранее, если бесконечная последовательность ω алгоритмически случайна по мере P, то любая последовательность ω', отличающаяся от нее в конечном числе членов, также является алгоритмически случайной по мере P. Поэтому $\omega, \omega' \in R$. Кроме этого, $\omega \equiv \omega'$. Можно выбрать алгоритмически инвариантные множества A и B так, что никакие две последовательности $\alpha \in A$ и

$\beta \in B$ алгоритмически не эквивалентны. Поэтому $\omega, \omega' \in A'$ или $\omega, \omega' \in B'$.

По следствию 12.1 из $\bar{M}(A') > 0$ следует $L(A') > 0$. Аналогично имеем $L(B') > 0$. Поэтому мы можем применить закон 0 или 1 А.Н.Колмогорова к последовательности $f_1, f_2, \ldots$ случайных величин $f_i(\omega) = \omega_i$ на вероятностном пространстве (L, Ω). Из свойства инвариантности множеств A' и B' относительно изменения последовательности в конечном числе членов следует, что каждое из них лежит в остаточной σ-алгебре последовательности $f_1, f_2, \ldots$. Поэтому мера L каждого из этих множеств равна 0 или 1. Это противоречит тому, что $A' \cap B' = \emptyset$ и $L(A') > 0$, $L(B') > 0$. Таким образом, $\mathbf{r}$ является атомом. $\square$

Тривиальным образом, $\mathbf{c}$ также является атомом Υ. Легко видеть, что $\mathbf{r}$ – единственный атом, мера Лебега которого равна 1.

Атомы $\mathbf{c}$ и $\mathbf{r}$ порождаются стандартными последовательностями. Возникает естественный вопрос, исчерпывается ли алгебра инвариантных свойств Υ этими элементами, т.е. будет ли $\mathbf{1} = \mathbf{r} \cup \mathbf{c}$?[2]

Следующая теорема дает ответ на этот вопрос.

Теорема 12.2. $\mathbf{1} \setminus \mathbf{r} \cup \mathbf{c} \neq \mathbf{0}$.

Доказательство этой теоремы приведено в разделе 12.3.

Структура алгебры Υ описывается следующими теоремами.

Теорема 12.3. *Множество всех атомов алгебры* Υ *счетно.*

Доказательство теоремы приведено в разделе 12.5.

Пусть $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \ldots$ – все атомы алгебры Υ, причем $\mathbf{a}_1 = \mathbf{c}$ и $\mathbf{a}_2 = \mathbf{r}$.

Теорема 12.4. *Имеет место* $\mathbf{1} \setminus \bigcup_{i=1}^{\infty} \mathbf{a}_i \neq \mathbf{0}$.

Доказательство теоремы приведено в разделе 12.6.

[2]Этот вопрос эквивалентен вопросу о независимости аксиомы 4.2 из [17].

Заметим, что по определению элемент $\mathbf{d} = \mathbf{1} \setminus \bigcup_{i=1}^{\infty} \mathbf{a}_i$ является бесконечно делимым, т.е. для любого ненулевого элемента $\mathbf{x} \subseteq \mathbf{d}$ существует разложение $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2$, где $\mathbf{x}_1 \cap \mathbf{x}_2 = \mathbf{0}$, $\mathbf{x}_1 \neq \mathbf{0}$ и $\mathbf{x}_2 \neq \mathbf{0}$.

Теоремы 12.3 и 12.4 определяют разложение единицы алгебры Υ.

Следствие 12.2. *Имеет место разложение* $\mathbf{1} = \cup_{i=1}^{\infty} \mathbf{a}_i \cup \mathbf{d}$, *где* $\mathbf{a}_1, \mathbf{a}_2, \ldots$ – *бесконечная последовательность всех атомов алгебры* Υ *(атом* $\mathbf{a_1} = \mathbf{r}$ *порождается случайными по мере Лебега последовательностями, атом* $\mathbf{a}_2 = \mathbf{c}$ *порождается всеми вычислимыми последовательностями),* $\mathbf{d}$ – *бесконечно делимый элемент.*

Все последовательности лежащие в множествах, определяющих атомы $\mathbf{a}_3, \mathbf{a}_4, \ldots$ и бесконечно делимый элемент $\mathbf{d}$, порождаются последовательностями, которые не являются стандартными, т.е. они не только не являются случайными по каким-либо вычислимым мерам, но и не могут быть алгоритмически эквивалентны никаким случайным последовательностям. Любая вычислимая мера множества всех нестандартных последовательностей равна 0. Так как априорная мера множества таких последовательностей положительна, то это означает, что их можно генерировать с помощью вероятностных алгоритмов (с вероятностью как угодно близкой к 1).

12.2. Сети и потоки

Для построения элементов Υ, порождаемых нестандартными последовательностями, будем определять перечислимые снизу полумеры P, для которых $\bar{P}(\Omega \setminus R) > 0$. Каждая из таких полумер будет определяться как поток по некоторой сети. Мы будем рассматривать множество Ξ всех конечных двоичных последовательностей как граф (дерево) вершинами которого являются последовательности, соединенные ребрами единичной длины

$(x, x0)$, $(x, x1)$, где $x \in \Xi$. В процессе конструкции мы будем добавлять дополнительные ребра (x, y), где $x, y \in \Xi$, $x \subset y$, длины $l(y) - l(x) > 1$. Для любого ребра $\sigma = (x, y)$ пусть $\sigma_1 = x$ – его начало, а $\sigma_2 = y$ – его конец. Функция $q(\sigma)$, определенная на всех ребрах единичной длины, а также на всех дополнительных ребрах, называется сетью, если она вычислима и для любого $x \in \Xi$

$$\sum_{\sigma:\sigma_1=x} q(\sigma) \leqslant 1. \tag{12.1}$$

В дальнейшем G^q обозначает множество всех дополнительных ребер сети q. Пусть q – сеть. Будем считать, что функция q определяет множество G^q всех дополнительных ребер, на которых она определена. q-потоком называется наименьшая полумера P, удовлетворяющая неравенству $P \geqslant R$, где функция R определятся

$$R(\lambda) = 1; \tag{12.2}$$

$$R(y) = \sum_{\sigma:\sigma_2=y} q(\sigma) R(\sigma_1) \tag{12.3}$$

при $y \neq \lambda$. Полумера P может быть вычислена по R следующим образом. Множество D называется сечением над x, если оно безпрефиксное и $x \subseteq y$ для любого $y \in D$. Пусть π_x – множество всех таких сечений. Тогда

$$P(x) = \sup_{D \in \pi_x} \sum_{x:x \in D} R(x). \tag{12.4}$$

Сеть q называется элементарной, если множество дополнительных ребер конечно и $q(\sigma) = 1/2$ для почти всех всех ребер длины 1. Элементарная сеть является конструктивным объектом.

С произвольной сетью связываем функцию задержки потока $s(x) = 1 - q(x, x0) - q(x, x1)$.

Общая часть конструкции. Мы укажем конструкцию сети, зависящую от произвольного рекурсивного отношения $B(i, q, \sigma)$, где i – натуральное число (номер задания), q – элементарная сеть, σ – дополнительное ребро. Пусть $p(\langle i, j \rangle) = i$ для

всех i. Каждое дополнительное ребро σ будет относится к некоторому заданию i так, что $p(l(\sigma_1)) = p(l(\sigma_2)) = i$. Целью задания i будет проведение дополнительных ребер σ так, чтобы каждая бесконечная последовательность ω проходила через одно из таких ребер или функция задержки была бы равна 1 на одном из начальных фрагментов ω. Ребро σ должно удовлетворять отношению $B(i, q, \sigma)$. Пусть $G^q(i)$ – множество всех дополнительных ребер, проведенных заданием i.

С отношением B связывается функция установки дополнительного ребра

$$\beta(q, x) = \min\{y : p(l(y)) = p(l(x)), B(p(l(x)), q, (x, y))\}. \quad (12.5)$$

Считаем, что значение $\min \emptyset$ не определено. Здесь пара $(x, \beta(q, x))$ является дополнительным ребром.

Индукцией по n определим последовательность q^n элементарных сетей. Полагаем $q^0(\sigma) = 1/2$ для всех путей длины 1. Пусть $n \geqslant 1$ и сеть q^{n-1} уже определена, s^{n-1} – функция задержки. G^{n-1} – множество всех ее дополнительных ребер, причем $l(\sigma_2) < n$ при $\sigma \in G^{n-1}$.

Определим сеть q^n и множество G^n. Определение распадается на два случая.

Пусть q – элементарная сеть. Предварительно введем вспомогательную функцию $w(i, q)$. Значение $w(i, q)$ равно наименьшему n такому, что $p(n) = i$ и $n > l(\sigma_2)$ для любого дополнительного ребра σ, установленного при обработке какого-либо задания $j < i$. На шагах $s \geqslant w(i, q)$ задания $j < i$ не обрабатывались. Изменение значения $w(i, q^n) \neq w(i, q^{n-1})$ может происходить из-за того, что некоторое задание $j < i$ устанавливает свое дополнительное ребро выше уровня старого значения $w(i, q^{n-1})$, и тем самым, нарушает условие определения $w(i, q^{n-1})$. Лемма 12.3 покажет, что это нарушение будет происходить не более чем на конечном число шагов конструкции.

Параметром конструкции будет функция $\rho(n)$, принимающая натуральные значения. В разделах 12.4 и 12.6 достаточно взять $\rho(n) = (n + 2)^2$, в разделе 12.5 будет $\rho(n) = 2^{2n+2}$.

Случай 1. $w(p(n), q^{n-1}) = n$ (шаг переустановки задания $i = p(n)$). В этом случае полагаем $s^n(y) = 1/\rho(n)$ при $l(y) = n$, определим $s^n(y) = s^{n-1}(y)$ для остальных y. Полагаем $G^n = G^{n-1}$.

Случай 2. $w(p(n), q^{n-1}) < n$ (шаг обработки задания $i = p(n)$). Пусть C_n обозначает множество всех x, требующих обработки, т.е. таких, что $w(i, q^{n-1}) \leqslant l(x) < n$, $0 < s^{n-1}(x) < 1$, $l(\beta(q^{n-1}, x)) = n$ (в частности $p(l(x)) = i$ и значение функции β определено) и не существует дополнительного ребра из G^{n-1}, исходящего из x.

В этом случае для каждого $x \in C_n$ определим $s^n(\beta(q^{n-1}, x)) = 0$ и определим $s^n(y) = s^{n-1}(x)/(1 - s^{n-1}(x))$ для остальных y таких, что $x \subset y$ и $l(y) = n$. Полагаем $s^n(y) = s^{n-1}(y)$ для всех остальных y. Кроме этого, определим

$$G^n = G^{n-1} \cup \{(x, \beta(q^{n-1}, x)) : x \in C_n\}.$$

Определим значение сети на ребрах σ единичной длины $q^n(\sigma) = (1/2)(1 - s^n(\sigma_1))$ и $q^n(\sigma) = s^n(\sigma_1)$ для дополнительных ребер $\sigma \in G^n$.

В результате определим $q = \lim\limits_{n\to\infty} q^n$, $s = \lim\limits_{n\to\infty} s^n$ и $G = \cup_n G^n$. Легко видеть, что q и s – вычислимые функции, а множество G – разрешимое.

Отметим свойство множества G дополнительных ребер.

Лемма 12.2. *Не может существовать двух ребер $(x, y) \in G$ и $(x', y') \in G$ таких, что $x \subset x' \subset y$ и $l(y) < l(y')$.*

Доказательство. Допустим, что такая пара ребер существует. Так как $l(y) < l(y')$, ребро (x', y') установлено позже, чем ребро (x, y). По конструкции $p(l(x')) > p(l(x))$ и $l(x') > l(y)$. Полученное противоречие доказывает утверждение. □

Следующая лемма показывает, что каждое задание приводит к установке новых дополнительных ребер только на конечном числе шагов.

Лемма 12.3. *$G(i)$ конечно и $w(i, q) < \infty$ для любого i.*

Доказательство. Заметим, что если $G(j)$ конечно для каждого $j < i$, то $w(i,q) < \infty$. Поэтому достаточно доказать, что $G(i)$ конечно для любого i. Допустим противное. Пусть i – наименьшее, для которого $G(i)$ бесконечно. Так как $G(j)$ конечно для любого $j < i$, то $w(i,q) < \infty$. Если $l(x) \geqslant w(i,q)$, то возьмем максимальное m такое, что из некоторого начального фрагмента x длины m исходит дополнительное ребро i-го типа. Если ни одного такого ребра не существует, то полагаем $m = w(i,q)$. Определим

$$K(x) = \begin{cases} 1/s(x^m) \text{ если } s(x^m) \neq 0, l(x) \geqslant w(i,q) \\ \rho(w(i,q)) \text{ если } l(x) < w(i,q) \end{cases}$$

Два случая в определении $K(x)$ исключают друг друга. Действительно, пусть $l(x) \geqslant w(i,q)$. Если $m = w(i,q)$, то по конструкции $s(x^m) = 1/\rho(m) > 0$. Если $m > w(i,q)$ – максимальное такое, что из начального фрагмента x длины m исходит дополнительное ребро i-го типа, то $s(x^m) > 0$, так иначе это ребро не было бы установлено.

По конструкции целочисленная функция $K(x)$ удовлетворяет условию: $K(x) \geqslant K(y)$ при $x \subset y$, причем, если $K(x) > K(y)$, то $K(x) > K(z)$ для любого z такого, что $x \subset z$ и $l(z) = l(y)$. Отсюда легко следует, что функция

$$\hat{K}(\omega) = \min\{n : \forall i \geqslant n(K(\omega^i) = K(\omega^n))\}$$

определена для всех $\omega \in \Omega$ и непрерывна. Она ограничена некоторым числом m, так как Ω – компакт. Тогда при $l(x) \geqslant m$ будет $K(x) = K(x^m)$. Если на шаге n было проведено хотя бы одно ребро i-го типа, то значение $K(y)$ уменьшается на 1 для некоторых y. Тогда существование такого m противоречит бесконечности $G(i)$. Лемма доказана. □

Последовательность $\alpha \in \Omega$ называется i-продолжением конечной последовательности x, если $x \subset \alpha$ и $B(i, q^{n-1}, (x, \alpha^n))$ выполнено для почти всех n. Последовательность $\alpha \in \Omega$ называется i-отброшенной, если $s(\alpha^n) = 1$ для некоторого n такого,

что $p(n) = i$. Заметим, что если $\sigma \in G(i)$ – дополнительное ребро i-го типа, то $B(i, q^{n-1}, \sigma)$ выполнено, где $n = l(\sigma_2)$.

Лемма 12.4. *Пусть для любого начального фрагмента ω^n бесконечной последовательности ω найдется i-продолжение. Тогда ω либо будет i-отброшена, либо ω пройдет через дополнительное ребро i-го типа (т.е. $\sigma_2 \subset \omega$ для некоторого $\sigma \in G(i)$).*

Доказательство. Пусть ω не является i-отброшенной. По лемме 12.3 существует максимальное m такое, что $p(m) = i$ и $s(\omega^m) > 0$. Так как ω^m имеет i-продолжение и $s(\omega^m) < 1$, по случаю 2 конструкции будет построено дополнительное ребро i-го типа (ω^m, y). По конструкции будет $s(y) = 0$ и $s(z) \neq 0$ при $\omega^m \subset z$, $l(z) = l(y)$, $z \neq y$. По выбору m будет $y \subset \omega$. □

Пусть s – функция задержки потока q, G – множество всех дополнительных ребер и Q – q-поток. Функция $w(i, q)$ была определена выше. Обозначим $w(i) = w(i, q)$.

Полумера P называется непрерывной, если $\lim\limits_{n\to\infty} P(\omega^n) = 0$ для любой бесконечной ω. Приведем одно достаточное условие непрерывности потока по сети.

Число n разделяет множество путей D, если для любого пути $\sigma \in D$ будет $l(\sigma_1) \geqslant n$ или $l(\sigma_2) < n$.

Лемма 12.5. *Произвольный r-поток непрерывен, если множество дополнительных ребер разделяется бесконечным множеством чисел, и $r(x, x0) = r(x, x1)$ для любого $x \in \Xi$.*

Доказательство. Пусть P – r-поток и число n разделяет множество дополнительных ребер. Тогда

$$P(x) = R(x) = r(x^{n-1}, x)R(x^{n-1}) \leqslant r(x^{n-1}, x)P(x^{n-1})$$

для любого x длины n. Из условия 12.1 и из условия леммы следует, что $r(x^{n-1}, x) \leqslant 1/2$ для всех x и n. Отсюда $P(\omega^n) \leqslant (1/2)P(\omega^{n-1})$ для всех n, разделяющих множество дополнительных ребер. Так как таких n бесконечно много, $\lim\limits_{n\to\infty} P(\omega^n) = 0$, т.е. P – непрерывна. □

Следствие 12.3. *Полумера Q непрерывна.*

Доказательство. Так как для любого i число $w(i)$ разделяет G, можно применить лемму 12.3. □

Лемма 12.6. *$Q(y) = 0$ тогда и только тогда, когда $q(\sigma) = 0$ для некоторого ребра σ единичной длины, лежащего на y, т.е. такого, что $\sigma_2 \subseteq y$.*

Доказательство. Доказательство необходимости очевидно. Для доказательства достаточности предположим, что $q(y^n, y^{n+1}) = 0$ для некоторого $n < l(y)$ и, тем не менее, $Q(y) \neq 0$. По определению $s(y^n) = 1$. Из $Q(y) \neq 0$ следует, что должно существовать дополнительное ребро $(x, z) \in G$ такое, что $x \subseteq y^n$ и $y^{n+1} \subseteq z$. Будем считать, что (x, z) – самое короткое такое ребро. Поэтому в множестве дополнительных ребер сети $q^{l(z)-1}$ нет ребра с таким свойством. Отсюда и из того, что $q((z)^n, (z)^{n+1}) = 0$ следует, что $\tilde{Q}(z) = 0$, где $\tilde{Q}$ есть $q^{l(z)-1}$-поток. Это равенство противоречит $(x, z) \in G$, так как для того, чтобы это ребро попало в G необходимо, чтобы $\tilde{Q}(z) \neq 0$. Полученное противоречие доказывает лемму. □

Носителем полумеры P будем называть множество

$$E = \{\omega \in \Omega : \forall n(P(\omega^n) \neq 0)\}.$$

Легко видеть, что E замкнуто в Ω и $\bar{P}(E) = \bar{P}(\Omega)$.

Из леммы 12.6 следует, что отношение $Q(y) = 0$ разрешимо и носителем Q является множество $E = \Omega \setminus \cup_{s(x)=1}\Gamma_x$.

Используя общую конструкцию этого раздела определим сеть q (полагаем в этой конструкции $\rho(n) = (n+2)^2$). Пусть Q – поток по сети q.

Лемма 12.7. *Имеем $\bar{Q}(\mathbf{1}) > 0$.*

Доказательство. Оценим снизу величину $\bar{Q}(\Omega)$. Пусть

$$S_n = \sum_{u:l(u)=n} R(u) - \sum_{\sigma:\sigma\in G, l(\sigma_2)=n} q(\sigma)R(\sigma_1).$$

Из определения функции задержки имеем

$$\sum_{u:l(u)=n+1} R(u) = \sum_{u:l(u)=n} (1-s(u))R(u) + \tag{12.6}$$

$$\sum_{\sigma:\sigma\in G,l(\sigma_2)=n+1} q(\sigma)R(\sigma_1). \tag{12.7}$$

Предварительно рассмотрим случай $w(p(n), q^{n-1}) < n$. Если не существует ни одного пути $\sigma \in G$, для которого $l(\sigma_2) = n$, то $S_{n+1} \geqslant S_n$. Пусть теперь такой путь существует. Определим

$$P(\sigma, u) \Longleftrightarrow l(\sigma_2) = l(u) \& \sigma_1 \subseteq u \& u \neq \sigma_2 \& \sigma \in G.$$

Из определения функций задержки имеем

$$\sum_{u:l(u)=n} s(u)R(u) = \sum_{\sigma:\sigma\in G,l(\sigma_2)=n} s(\sigma_2) \sum_{u:P(\sigma,u)} R(u) =$$

$$\sum_{\sigma:\sigma\in G,l(\sigma_2)=n} \frac{s(\sigma_1)}{1-s(\sigma_1)} \sum_{u:P(\sigma,u)} R(u) \leqslant \sum_{\sigma:\sigma\in G,l(\sigma_2)=n} s(\sigma_1)R(\sigma_1) =$$

$$\sum_{\sigma:\sigma\in G,l(\sigma_2)=n} q(\sigma)R(\sigma_1).$$

Здесь мы использовали неравенство

$$\sum_{u:P(\sigma,u)} R(u) \leqslant (1-s(\sigma_1))R(\sigma_1) \tag{12.8}$$

для любого $\sigma \in G$ такого, что $l(\sigma_2) = n$. Неравенство (12.8) имеет место, так как сумма слева равна величине потока через множество вершин $\{u : P(\sigma, u)\}$, а величина из правой части неравенства равна величине исходящего из вершины σ_1 потока, за исключением его части, проходящей через дополнительное ребро σ. По лемме 12.2 не может быть ребра $\sigma' \in G$ такого, что $\sigma'_1 \subset \sigma_1$ и $l(\sigma'_2) < l(\sigma_2)$. Поэтому никакая дополнительная порция потока из какой-либо вершины $\sigma'_1 \subset \sigma_1$ не может увеличить величину потока из σ_1.

Объединяя полученную оценку с (12.6)–(12.7), получим $S_{n+1} \geqslant S_n$.

Рассмотрим теперь случай $w(p(n), q^{n-1}) = n$. Тогда

$$\sum_{u:l(u)=n} s(u)R(u) \leqslant \rho(n) = 1/(n+2)^2.$$

Объединяя это неравенство с (12.6)–(12.7), получим $S_{n+1} \geqslant S_n - 1/(n+2)^2$. Отсюда и из $S_0 = 1$ получим

$$S_n \geqslant 1 - \sum_{i=1}^{\infty} 1/(i+1)^2 \geqslant \frac{1}{2}$$

для всех n. Так как $Q \geqslant R$, получим

$$\bar{Q}(\Omega) = \inf_n \sum_{l(u)=n} Q(u) \geqslant \inf_n S_n \geqslant \frac{1}{2}.$$

Лемма доказана. □

12.3. Доказательство теоремы 12.2

Приведем типичный пример применения вышеприведенной конструкции. Докажем, что $\mathbf{1} \setminus (\mathbf{r} \cup \mathbf{c}) \neq \mathbf{0}$.

Основной результат построения представлен в теореме 12.5 ниже.

Пусть F_i – вычислимая последовательность всех вычислимых операторов. Мы перестроим нумерацию F_i следующим образом. Для любого i определим $F'_{\langle i,j \rangle} = F_i$ для всех j. Новую последовательность операторов по-прежнему обозначим F_i. Таким образом, по любому номеру i вычислимого оператора F_i можно перечислить бесконечную последовательность других его номеров. Определим

$$B(i, q, \sigma) \Longleftrightarrow Q(\sigma_2) \neq 0 \& l(F_i(\sigma_2)) > \sigma_1 + i,$$

где конечная последовательность σ_1 (начало ребра σ) отождествляется с ее номером в естественной нумерации множества Ξ, Q обозначает q-поток.

Пусть $p(\langle i,j\rangle) = i$ для всех i. Каждое дополнительное ребро σ будет относится к некоторому заданию i так, что $p(l(\sigma_1)) = p(l(\sigma_2)) = i$. Целью задания i будет проведение дополнительных ребер σ так, чтобы каждая бесконечная последовательность ω проходила через одно из таких ребер или функция задержки была бы равна 1 на одном из начальных фрагментов ω. Иными словами, цель задания – найти такое продолжение σ_2 конечной последовательности $x = \sigma_1$, для которого значение $F_i(\sigma_2)$ имеет достаточно большую длину.

Теорема 12.5. *Для любой бесконечной последовательности ω из носителя полумеры Q и любого вычислимого оператора F, если $F(\omega)$ бесконечно, то $F(\omega)$ не является алгоритмически случайной по равномерной мере.*

Доказательство. Заметим, что если $F(\omega)$ бесконечна и $F_i = F$, то для любого начального фрагмента последовательности ω найдется i-продолжение. Поэтому для любого такого i найдется ребро $\sigma \in G(i)$, лежащее на ω. Для любого i определим открытое множество

$$U_i = \cup_{\sigma\in G(i)}\Gamma_{F_i(\sigma_2)}.$$

Ввиду того, что при $\sigma \in G(i)$ имеем $l(F_i(\sigma_2)) > \sigma_1 + i$,

$$L(U_i) \leqslant \sum_{\sigma\in G(i)} 2^{-\sigma_1 - i} \leqslant 2^{-i},$$

где L – равномерная мера.

Отсюда легко увидеть, что для любого рационального $\epsilon > 0$ можно эффективно найти такое i, что $L(U_i) \leqslant \epsilon$ и $F(\omega) = F_i(\omega) \in U_i$. Таким образом, последовательность $F(\omega)$ не является случайной относительно равномерной меры L. $\square$

Следствие 12.4. *$\bar{Q}$-почти любая бесконечная последовательность ω не является алгоритмически случайной по любой вычислимой мере. Таким образом, априорная мера $\bar{M}$ множества всех последовательностей, неслучайных относительно вычислимых мер, положительна.*

Доказательство. Множество всех вычислимых последовательностей счетно. Из непрерывности полумеры Q следует, что $\bar{Q}$-почти любая последовательность из ее носителя невычислимая.

Согласно теорема 12.4 любая невычислимая случайная относительно какой-либо вычислимой меры последовательность алгоритмически эквивалентна последовательности, случайной относительно равномерной меры.

Мы доказали, что любая последовательность α, которая алгоритмически сводится к последовательности ω из носителя полумеры Q не является случайной по равномерной мере. Поэтому $\bar{Q}$-почти любая последовательность не случайна по любой вычислимой мере.

Так как полумера Q перечислима снизу, $cM \geqslant Q$ для некоторой константы c. Отсюда $\bar{M}(\Omega \setminus R) \geqslant \bar{Q}(\Omega \setminus R) > 0$, где $\Omega \setminus R$ – множество всех последовательностей, которые не являются случайными по любой вычислимой мере. □

12.4. Доказательство теоремы 11.2

Формулировка теоремы и необходимые определения имеются в разделе 11.3.

Пусть $\beta(n)$ – неограниченная вычислимая функция такая, что $\beta(n) \leqslant (1-\epsilon)n$ для всех n, где $0 < \epsilon < 1$. Без потери общности можно считать, что ϵ рационально. Мы построим перечислимую сверху функцию $\alpha(n)$ такую, что $\bar{M}(I_{\alpha,\beta}) > 0$. Построение будет осуществляться на основе общей конструкции раздела 12.2.

Рассмотрим на Ξ частичный порядок:

$$\gamma \ll \gamma' \Longleftrightarrow l(\gamma) = l(\gamma') \& \forall i(\gamma_i \leqslant \gamma'_i),$$

а также линейный порядок $\preceq$, при котором все кортежи меньшей длины предшествуют всем кортежам большей длины, упорядочение кортежей равной длины является обратным к стандартному лексикографическому порядку. Нам важно, что из $\gamma \ll \gamma'$ следует $\gamma \succeq \gamma'$ Обозначим посредством $\gamma(i)$ кортеж с номером i при линейном упорядочении $\preceq$, т.е. $\gamma(i) \preceq \gamma(j)$ при $i \leqslant j$.

Пусть P_i – перечислимая снизу последовательность всех перечислимых снизу полумер. Пусть $P_i(x) = \lim\limits_{k\to\infty} P_i^k(x)$, где $P_i^k(x)$ – вычислимая последовательность простых полумер, неубывающая по k.

Определим рекурсивное отношение $B(i, q, \sigma)$, где i – порядковый номер кортежа $\gamma(i)$ (относительно линейного порядка $\preceq$), q – сеть, σ – дополнительное ребро. Данное отношение определяет дополнительное ребро σ, которое пополнит нашу сеть для того, чтобы выполнить часть i-ого задания. Кортеж $\gamma(i)$ соответствует i-ому заданию. Если $(\gamma(i))_j = 1$, то мы будем ожидать, что полумера P_j является мерой (или достаточно близка к ней с заданной степенью точности).

Пусть задана некоторая простая сеть q, s – ее функция задержки, полумера Q есть q-поток.

Номер задания i будет определять конечная последовательность x, из которой будет проводиться дополнительное ребро, а именно, $i = p(l(x))$.

Пусть $m(x, q)$ равно наименьшему m такому, что $p(m) = i$, $m > l(x)$ и для задержанной части потока выполняется неравенство

$$s(x)Q(x) \geqslant l(\gamma(i))2^{\beta(m)/(1-\epsilon/2)-m+l(x)+3}. \qquad (12.9)$$

Такое m всегда существует, по свойству $\beta(n) \leqslant (1-\epsilon)n$ для всех n.

Ребро σ будет удовлетворять i-ому заданию, т.е. $B(i, q, \sigma)$ выполнено, если $Q(\sigma_2) \neq 0$ и оно имеет достаточно большую длину $k = l(\sigma_2)$ (правого конца) так, что при $m = m(\sigma_1, q)$ для всех j, для которых $(\gamma(i))_j = 1$, верно следующее:

- вычисленная часть (на уровне m) полумеры P_j достаточно

велика так, что выполнено неравенство

$$\sum_{l(z)=m} P_j^k(z) \geqslant 1 - l(\gamma(i))2^{-m+l(\sigma_1)+1}; \tag{12.10}$$

- вычисленная величина j-ой полумеры на правом конце ребра σ достаточно мала так, что выполнено неравенство

$$P_j^k((\sigma_2^m)) \leqslant l(\gamma(i))2^{-m+l(\sigma_1)+1}. \tag{12.11}$$

Далее, по общей конструкции раздела 12.2 определим сеть q. Пусть s – функция задержки потока q, G – множество всех дополнительных ребер и Q – q-поток. Функция $w(i,q)$ была определена в разделе 12.2. Обозначим $w(i) = w(i,q)$.

Следствие 12.5. *(Из леммы 12.3). Полумера Q непрерывна.*

Для применения леммы 12.3 к полумере Q достаточно заметить, что для любого i число $w(i,q)$ разделяет Q.

Из леммы 12.6 следует, что отношение $Q(y) = 0$ разрешимо и носителем Q является множество $E = \Omega \setminus \cup_{s(x)=1}\Gamma_x$.

Ясно, что лемма 12.7 имеет место и в данном случае. По этой лемме имеем $\bar{Q}(\mathbf{1}) > 0$.

Определим $\alpha(n) = \lfloor \log l(\gamma(i) \rfloor$, где i такое, для которого $w(i) \leqslant n < w(i+1)$. По конструкции $\alpha(n)$ – перечислима сверху и неограничена.

Лемма 12.8. *Имеем $E \subseteq I_{\alpha,\beta}$.*

Доказательство. Допустим, что для некоторых $\omega \in E$ и числа n_0 при всех $n \geqslant n_0$ последовательность ω^n является $(\alpha(n), \beta(n))$-стохастической. Пусть d – наименьшее такое, что для всех i из $l(\gamma(i)) \geqslant d$ следует, что $w(i) > n_0$ и пусть i_0 – наименьшее такое i. Заметим, что по определению линейного порядка $\preceq$ будет $\gamma(i_0) = 1^d$.

Для произвольного n пусть последовательность δ^n имеет длину d и такая, что

$$\delta_j^n = 1 \Longleftrightarrow \sum_{z:l(z)=m} P_j(z) = 1,$$

где $m = m(\omega^n, q^{n-1})$. Тогда $\delta^{n+1} \ll \delta^n$ для всех n. Пусть l – максимальное число такое, что $\gamma(i_0+l) \gg \delta^{w(i_0+l)}$. Такое l существует, так как из $\gamma \ll \gamma'$ следует $\gamma \succeq \gamma'$.

Допустим, что для некоторого j такого, что

$$w(i_0+l) \leqslant j < w(i_0+l+1)$$

будет $\gamma(l_0+l) \neq \delta^j$. Для этого j будет $\gamma(l_0+l) \gg \delta^j$. Тогда, так как $\delta^j \gg \delta^{w(i_0+l+1)}$, существует $l' > l$ такое, что $\gamma(i_0+l') = \delta^j \gg \delta^{w(i_0+l+1)} \gg \delta^{w(i_0+l')}$, что противоречит выбору l. Следовательно, $\gamma(i_0+l) = \delta^j$ при $w(i_0+l) \leqslant j < w(i_0+l+1)$.

Возьмем $i = i_0 + l$. Докажем, что существует ребро $\sigma \in G(i)$ такое, что $\sigma_2 \subset \omega$ и $B(i, q^{n-1}, \sigma)$ выполнено при $n = l(\sigma_2)$. Допустим противное. Пусть n – наибольшее число такое, что $p(n) = i$ и $s(\omega^n) > 0$. Пусть $m = m(\omega^n, q^{n-1})$. Как только что было доказано, из $w(i) \leqslant n < w(i+1)$ следует $\delta^n = \gamma(i)$. Отсюда следует, что для всех достаточно больших k будет выполнено

$$(\gamma(i))_j = 1 \Longleftrightarrow \sum_{z:l(z)=m} P_j^k(z) \geqslant 1 - d2^{-m+n+1}.$$

Кроме этого, для всех достаточно больших k существует x такое, что $l(x) = m$, $\omega^n \subseteq x$ и $P_j^k(x) < d2^{-m+n+1}$ для всех $j \leqslant d$. Действительно, при $j \leqslant d$ имеется не более $2^{m-n-1}/d$ таких x длины m, что $P_j(x) > d2^{-m+n+1}$. Всего таких x по всем $j \leqslant d$ не более 2^{m-n-1}, поэтому такое x найдется из соображений мощности. Тогда по конструкции некоторое дополнительное ребро σ с началом в ω^n будет добавлено в $G(i)$. Из того, что $s(\omega^n) \neq 1$, следует, что должно быть $\sigma_2 \subset \omega$.

Из существования такого ребра σ следует, что для каждого $j \leqslant d$, если $\sum_{z:l(z)=m} P_j^k(z) = 1$, то $(\gamma(i))_j = \delta_j^n = 1$. Поэтому $P_j(\omega^n) \leqslant P_j^k(\omega^m) + d2^{-m+n+1} \leqslant d2^{-m+n+2}$, где $k = l(\sigma_2)$, $n = l(\sigma_1)$, $m = m(\omega^n, q^{n-1})$. По определению $d = \alpha(m)$. Кроме этого, по (12.9) и (12.11) имеем

$$Q(\omega^m) > 2^{\beta(m)/(1-\epsilon/2)} P_j(\omega^m), \tag{12.12}$$

где Q – q-поток. Так как для любой меры P функция $Q(x)/P(x)$ является P-супермартингалом, существует такая константа c, что $c\tilde{\psi}(j,x) \geqslant Q(x)/P_j(x)$ для всех x и для всех $j \leqslant d$ таких, что P_j является мерой.

Из (12.12) следует, что если мы выберем достаточно большое начальное значение n_0, то для любой меры P с программой p длины $\leqslant \alpha(m)$ будет выполнено

$$\tilde{d}(p,\omega^m) > \beta(m)/(1-\epsilon/2) - \log c > \beta(m),$$

т.е. ω^m не является $(\alpha(m),\beta(m))$-стохастической. Полученное противоречие доказывает лемму. $\square$

Как замечено выше $\bar{Q}(E) > 0$. Поэтому $\bar{M}(I_{\alpha,\beta}) > 0$. Теорема доказана.

12.5. Доказательство теоремы 12.3

Формулировка теоремы и необходимые определения имеются в разделе 12.1.

Нетрудно показать, что множество всех атомов алгебры Υ не более чем счетно. Для этого выберем для каждого атома $\mathbf{a} = [A]$ объединение конечного числа шаров D_a такое, что

$$\bar{M}((A \setminus D_a) \cup (D_a \setminus A)) < (1/4)\bar{M}(\mathbf{a}).$$

Тогда при $\mathbf{a} \neq \mathbf{b}$ будет $D_a \neq D_b$. Множество подобных объединений не более чем счетно.

Модификация общей части конструкции. Приведенная ниже модификация не затрагивает приведенных выше свойств общей части конструкции.

Повторим определения из общей части конструкции из раздела 12.2. Пусть $p(n)$ и $\tilde{p}(n)$ – вычислимые функции такие, что для каждой пары натуральных чисел (i,k) выполнено $p(n) = i$ и $\tilde{p}(n) = k$ для бесконечно многих n.

Каждое дополнительное ребро σ соответствует некоторому заданию i, где $p(l(\sigma_1)) = p(l(\sigma_2)) = i$. Оно также соответствует некоторому подзаданию (i,k), где $\tilde{p}(l(\sigma_1)) = \tilde{p}(l(\sigma_2)) = k$.

Целью задания i является установка некоторых дополнительных ребер σ таких, что каждая бесконечная двоичная последовательность ω продолжает одно из этих ребер или функция задержки равна 1 на одном из начальных сегментов ω.

Определим последовательность q^n элементарных сетей математической индукцией по n. Определим $q^0(\sigma) = 1/2$ для всех ребер длины 1. Пусть $n \geqslant 1$ и сеть q^{n-1} уже определена, s^{n-1} – соответствующая функция задержки и G^{n-1} – множество дополнительных ребер, кроме этого, пусть $l(\sigma_2) < n$ для всех $\sigma \in G^{n-1}$.

Пусть q – элементарный поток и G – множество дополнительных ребер. Обозначаем подмножество дополнительных ребер, соответствующих заданию i,

$$G(i) = \{\sigma \in G : p(l(\sigma_1)) = p(l(\sigma_2)) = i\}$$

Рассмотрим вспомогательную функцию $w(i, q)$. Значение $w(i, q)$ равно наименьшему n такому, что $p(n) = i$ и $n > l(\sigma_2)$ для любого дополнительного ребра σ, которое было установлено заданием $j < i$. В частности отсюда следует, что на шагах n' таких, что $w(i, q) \leqslant n' \leqslant n$ никакое задание $j < i$ не обрабатывалось. Определим

$$w(i, q) = \min\{n : p(n) = i \& \forall j \forall \sigma(j < i \& \sigma \in G(j) \to n > l(\sigma_2))\}$$

. Говорим, что $w(i, q^{n-1})$ есть начальный шаг сессии по обработке задания i.

Нарушение равенства $w(i, q^n) = w(i, q^{n-1})$ может произойти, только если некоторое задание $j < i$ установило дополнительное ребро, расположенное выше уровня $w(i, q^{n-1})$, и таким образом, нарушило условие определения $w(i, q^{n-1})$. Лемма 12.3 утверждает, что это может случиться только на конечном числе шагов конструкции.

Введем семейство отношений эквивалентности между конечными последовательностями, зависящих от параметра n:

$$x \sim_n y \iff l(x) = l(y) \& \forall (n \leqslant i \leqslant l(x) \Longrightarrow x_i = y_i);$$

для любых путей σ и σ'. По определению $\sigma \sim_n \sigma'$ тогда и только тогда, когда $\sigma_1 \sim_n \sigma'_1$ и $\sigma_2 \sim_n \sigma'_2$.

Мы будем писать $x \sim z$ вместо $x \sim_w z$, где $w = w(p(l(x)), q^{l(x)})$, причем всегда будет ясно, о какой сети q идет речь.

Несколько модифицируем общую часть конструкции из раздела 12.2. Приведем неформальные пояснения. Для того, чтобы соответствующий поток определял атом алгебоы Υ, в модифицированной конструкции величина потока через любые два ребра σ и σ' таких, что $\sigma' \sim \sigma$ должна быть одинаковой. Поэтому при установке ребра σ все ребра $\sigma' \sim \sigma$ и $\sigma' \neq \sigma$ становятся зависимыми от него и не могут также устанавливаться заданием i.

Для того, чтобы избежать коллизию при установке ребер задания i, разобъем процесс выполнения задания i на подзадания (i, k). Мы будем в процессе конструкции выполнять задание i, выполняя подзадания (i, k) в порядке их приоритета. Если ребро установлено подзаданием (i, k), то говорим, что оно также установлено заданием i.

Пусть $z_{i,1,q}, \ldots, z_{i,2^{w(i,q)},q}$ – все конечные последовательности z длины $w(i, q)$, выписанные в лексикографическом порядке.

Пока в процессе конструкции выполнено равенство $w(i, q^n) = w(i, q^{n-1})$, мы будем выполнять каждое подзадание (i, k), где $k \leqslant 2^{w(i,q^{n-1})}$, в пределах поддерева, продолжающего последовательность $z_{i,k,q^{n-1}}$. При этом, установка нового ребра в каждом таком поддереве будет разрушать все подзадания (i, s), соответствующие поддеревьям, имеющим меньший приоритет, а точнее, для которых $s > k$.

Определим при $1 \leqslant k \leqslant 2^{w(i,q)}$

$$G(i, k) = \{\sigma \in G(i) : \tilde{p}(l(\sigma_1)) = \tilde{p}(l(\sigma_2)) = k \& z_{i,k,q} \subseteq \sigma_1\}.$$

Полагаем $G(i, k) = \emptyset$ при $k > 2^{w(i,q)}$. Назовем $G(i, k)$ множеством дополнительных ребер, установленных подзаданием (i, k).

По определению, значение $w(i, k, q)$ будет равно наименьшему n такому, что выполнены следующие условия. Во-первых, $p(n) = i$ и $n > l(\sigma_2)$ для каждого дополнительного ребра σ, которое было

установлено заданием $j < i$. Во-вторых, $n > l(\sigma_2)$ для каждого дополнительного ребра σ, которое было установлено каким-либо подзаданием (i, s), где $s < k$.

Это означает, что на всех шагах n' таких, что $w(i,k,q) \leqslant n' \leqslant n$, любое подзадание (j, s), где $j < i$ или $j = i$ и $s < k$, не обрабатывалось. Приведем точное определение

$$w(i,k,q) = \min\{n : p(n) = i \& \tilde{p}(n) \leqslant 2^{w(i,q)} \& \tilde{p}(n) = k \& \\ \forall j \forall \sigma((j < i \& \sigma \in G(j) \to n > l(\sigma_2)) \& \\ \forall s \forall \sigma(1 \leqslant s < k \& \sigma \in G(i,s) \to n > l(\sigma_2))\}.$$

Полагаем $w(i,k,q) = \infty$ при $k > 2^{w(i,q)}$.

Говорим, что $w(i,k,q^{n-1})$ – начальный шаг подсессии по обработке подзадания (i,k). По определению $w(i,k,q) \geqslant w(i,q)$.

Нарушение равенства $w(i,k,q^n) = w(i,k,q^{n-1})$ может произойти из-за того, что $w(i,q^n) \neq w(i,q^{n-1})$ или при $w(i,q^n) = w(i,q^{n-1})$ из-за того, что некоторое подзадание (i,s), где $s < k$, установливает дополнительное ребро выше уровня $w(i,k,q^{n-1})$ и, таким образом, нарушает условие определения $w(i,k,q^{n-1})$. Далее будет показано, что такое нарушение будет происходить на не более чем конечном числе шагов n таких, что $w(i,q^n) = w(i,q^{n-1})$.

Пусть $G(i)$ – множество всех дополнительных ребер, установленных заданием i. Рекурсивному отношению B соответствует функция установки дополнительного ребра

$$\beta(q,x) = \min\{y : p(l(y)) = p(l(x)), B(p(l(x)), q, (x,y))\}. \quad (12.13)$$

Считаем, что $\min \emptyset$ не определено.

Полагаем $\rho(n) = 2^{2n+2}$. Определим сеть q^n и множество G^n дополнительных ребер. Это определение распадается на три случая.

Случай 1. $w(i,k,q^{n-1}) = n$ (первичная установка или переустановка подзадания (i,k), где $i = p(n)$ и $k = \tilde{p}(n)$).

Определим $s^n(y) = 1/\rho(n)$ для всех y таких, что $l(y) = n$ и $z_{i,k,q^{n-1}} \subseteq y$. Определим также $s^n(y) = s^{n-1}(y)$ для всех остальных y. Полагаем $G^n = G^{n-1}$.

Случай 2. $w(i,k,q^{n-1}) < n$ (шаг обработки подзадания (i,k), где $i = p(n)$ и $k = \tilde{p}(n)$).

Пусть C_n обозначает множество всех последовательностей x, которые должны быть обработаны, т.е. такие, что $z_{i,k,q^{n-1}} \subset x$, $w(i,k,q^{n-1}) \leqslant l(x) < n$, $0 < s^{n-1}(x) < 1$, $l(\beta(q^{n-1},x)) = n$ (также в этом случае, $p(l(x)) = i$ и $\tilde{p}(l(x)) = k$, значение функции β определено) и не существует дополнительного ребра из G^{n-1}, исходящего из x.

В этом случае определим $s^n(\beta(q^{n-1},x)) = 0$ для каждого $x \in C_n$, также определим $s^n(y) = s^{n-1}(x)/(1 - s^{n-1}(x))$ для остальных y таких, что $x \subset y$ и $l(y) = n$. Полагаем $s^n(y) = s^{n-1}(y)$ для всех остальных y. Определим

$$G^n = G^{n-1} \cup \{(x, \beta(q^{n-1},x)) : x \in C_n\}.$$

Определим значение сети на ребрах σ единичной длины $q^n(\sigma) = (1/2)(1 - s^n(\sigma_1))$ и $q^n(\sigma) = s^n(\sigma_1)$ для каждого дополнительного ребра $\sigma \in G^n$.

Случай 3. Случаи 1 и 2 не имеют места. В этом случае определим $q^n = q^{n-1}$ и $G^n = G^{n-1}$.

На этом описание шага индукции заканчивается.

Определим $q = \lim\limits_{n\to\infty} q^n$, $s = \lim\limits_{n\to\infty} s^n$ и $G = \cup_n G^n$. Из определения следует, что q и s – вычислимые функции, а множество G разрешимое.

Следующая лемма утверждает, что любое подзадание (i,k) обрабатывается только конечное число раз.

Лемма 12.9. *$G(i,k)$ конечно и $w(i,k,q) < \infty$ для любых i и $k \leqslant 2^{w(i,k,q)}$.*

Доказательство. По лемме 12.3 $w(i,q) < \infty$, поэтому $w(i,q^{n-1}) = w(i,q)$ для всех $n \geqslant n_i$ для некоторого n_i. Далее, также как в доказательстве леммы 12.3, показываем, что $w(i,k,q^n) \neq w(i,k,q^{n-1})$ только для конечного числа различных $n \geqslant n_i$. $\square$

Мы построим бесконечную последовательность р.п. полумер $P_1, P_2, \ldots$, по которой будет определена последовательность попарно различных атомов $\mathbf{d_1}, \mathbf{d_2}, \ldots$.

Пусть $\langle x_1, x_2, x_3\rangle$ – номер тройки натуральных чисел, при некотором фиксированном вычислимом взаимно-однозначном соответствии между всеми тройками $\langle x_1, x_2, x_3\rangle$ такими, что $x_2 \neq x_3$, и всеми натуральными числами. Заданы также обратные функции $[\langle x_1, x_2, x_3\rangle]_i = x_i$, $i = 1, 2, 3$. Номер $\langle x_1, x_2, x_3\rangle$ каждой такой тройки будет кодом некоторого задания, где x_1 – номер вычислимого оператора, x_2 – база задания, x_3 – мишень задания. Цель задания – обеспечить условия, при которых оператор F_{x_1} не может перевести никакую бесконечную последовательность из носителя полумеры P_{x_2} в носитель полумеры P_{x_3}. Конкурирующим требованием является то, что эти полумеры должны быть нетривиальными, т.е. должно быть $\bar{P}_{x_2}(\Omega) > 0$ и $\bar{P}_{x_3}(\Omega) > 0$.

Введем семейство отношений эквивалентности между конечными последовательностями, зависящих от параметра n:

$$x \sim_n y \Longleftrightarrow l(x) = l(y) \& \forall (n \leqslant i \leqslant l(x) \Longrightarrow x_i = y_i);$$

для любых путей σ и σ' по определению $\sigma \sim_n \sigma'$ тогда и только тогда, когда $\sigma_1 \sim_n \sigma'_1$ и $\sigma_2 \sim_n \sigma'_2$.

Мы будем писать $x \sim z$ вместо $x \sim_w z$, где $w = w(p(l(x)), q^{l(x)})$, причем всегда будет ясно, о какой сети q идет речь.

Основная причина модификации общей конструкции заключается в том, что теперь конструкция обладает следующим свойством.

Лемма 12.10. *Для любого i и любого ребра $\sigma \in G(i)$ не существует ребра $\sigma' \in G(i)$ такого, что $\sigma' \neq \sigma$ и $\sigma' \sim \sigma$.*

Доказательство. Дополнительные ребра устанавливаются в процессе обработки подзаданий. Для каждого ребра $\sigma \in G(i)$, установленного нкоторым подзаданием (i, k), будет $z_{i,k,q^{n-1}} \subseteq \sigma_1$,

при этом, все остальные подзадания (i,s) при $s > k$ переустанавливаются и обрабатываются на шагах, больших $l(\sigma_2)$. Для ребер σ', установленных подзаданиями (i,s) при $s < k$ выполнено $l(\sigma'_2) < l(\sigma_1)$. Поэтому дополнительные ребра, установленные подзаданиями (i,s), $s \neq k$, не могут быть эквивалентными σ. $\square$

Пусть р.п. множество $\hat{F}$ задает некоторый оператор F. Мы преобразуем его в другой оператор $F'(x)$ так, чтобы длина результата $F'(x)$ на конечной последовательности x не превышала длины аргумента:

$$F'(x) = \max\{u : \exists z(z \subseteq x \& (z,u) \in \hat{F} \& l(u) \leqslant l(x))\}.$$

Заметим, что на бесконечных ω будет $F(\omega) = F'(\omega)$.

В разделе 11.1 была построена вычислимая последовательность всех вычислимых операторов F_i. Будем считать, что все они преобразованы, как указано выше.

Будем говорить, что последовательность x является i-закрытой последовательностью y (или ребром σ, для которого $\sigma_2 = y$), если $l(y) = l(x)$ и найдется u, для которого $x \sim_{w(i)} u$ и $F_{[i]_1}(y) \subseteq u$, где $w(i) = w(i, q^{l(x)})$.

Мы определим эффективную операцию Ψ_m, зависящую от параметра m, которая преобрпзует произвольную сеть q и множество G дополнительных ребер в сеть q_m и множество дополнительных ребер G_m.

Если $[p(l(\sigma_1))]_2 = m$, то говорим, что сеть q_m или число m – база задания $i = p(l(\sigma_1))$. Если $[p(l(\sigma_1))]_2 = m$, то говорим, что сеть q_m – мишень задания i. $\sigma \in G_m$ и $q_m(\sigma) = 1/2$, если $[p(l(\sigma_1))]_2 \neq m$ и $[p(l(\sigma_1))]_3 \neq m$. По свойству нумерации троек число m не может быть одновременно мишенью и базой одного и того же задания.

Полумера P_m будет определяться как поток по сети q_m, $m = 1, 2, \ldots$.

Пусть q – сеть, s – ее функция задержки и G – множество дополнительных ребер. Пусть x – произвольная конечная последовательность, $i = p(l(x))$ – номер соответствующего задания. Далее приводим следующее определение.

1) Если m является базой задания i (т.е. $m = [i]_2$), то определим $s_m(x) = \min\{\sum\limits_{z \sim x} s(z), 1\}$.

2) Если m является мишенью задания i (т.е. $m = [i]_3$) и x является i-закрытой некоторым ребром из $G_m(i)$, то определим $s_m(x) = 1$.

3) Определим $s_m(x) = 0$ в остальных случаях.

****

Пусть x – произвольная конечная последовательность, пусть $n = l(x)$, $i = p(n)$ and $k = \tilde{p}(n)$ – номера соответствующих задания и подзадания. Определим

1) Если m является базой задания i (т.е. $m = [i]_2$), то определим $s_m(x) = s(x')$, где x' такое, что $x \sim x'$ и $z_{i,n,k} \subseteq x'$ (т.е. x' есть единственная последовательность, лежащая в поддереве подзадания (i, k) и эквивалентная x. Определим

$$G_m = \{\sigma : \exists \sigma' (\sigma \sim \sigma' \& \sigma' \in G)\},$$

$q_m(\sigma) = q(\sigma')$, где σ' есть единственный элемент из G, для которого $\sigma \sim \sigma'$.

2) Если m является мишенью задания i (т.е. $m = [i]_3$) и x является i-закрытой некоторым ребром из $G_{[i]_2}(i)$, то определим $s_m(x) = 1$.

3) Определим $s_m(x) = 0$ в остальных случаях.

****

Множество дополнительных ребер G_m сети q_m будет состоять из ребер σ, эквивалентных дополнительным ребрам сети q (т.е. $\sigma \sim \sigma'$ для некоторого $\sigma' \in G$), для которых m является базой задания $i = p(l(\sigma_1))$ (т.е. $[p(l(\sigma_1))]_2 = m$) и для которых $s_m(\sigma_1) < 1$.

Определим

$$q_m(\sigma) = \begin{cases} (1/2)(1 - s_m(\sigma_1)) \text{ если ребро } \sigma \text{ имеет единичную длину,} \\ \sum\limits_{\sigma':\sigma' \in G, \sigma' \sim \sigma} q(\sigma') \text{ если } \sigma \in G_m, \\ 0 \text{ в противном случае.} \end{cases}$$

Так как

$$\sum_{\sigma:\sigma \in G, \sigma_1 = x} q(\sigma) \leqslant s(x),$$

имеем

$$\sum_{\sigma:\sigma \in G_m, \sigma_1 = x} q_m(\sigma) \leqslant \sum_{z:z \sim x} s(z) = s_m(x).$$

Отсюда следует, что функция q_m удовлетворяет условию (12.1).

Теперь мы в состоянии определить рекурсивное отношение, необходимое для применения модифицированной общей схемы раздела 12.2

$$B(i, q, \sigma) \Longleftrightarrow Q_{[i]_2}(\sigma_2) \neq 0 \& \sum_z Q_{[i]_3}(z) \leqslant 2^{-\sigma_2 + 3}, \qquad (12.14)$$

где для произвольного m буквой Q_m обозначается $\Psi_m(q)$-поток, а сумма берется по всем z длины $l(\sigma_2)$, которые являются i-закрытыми последовательностью σ_2. Здесь в показателе степени мы отождествляем конечную последовательность σ_2 и ее номер.

Далее, в пределах этого раздела, буква q будет обозначать сеть, определенную по модифицированной общей схеме из этого раздела для отношения (12.14), G – множество всех дополнительных ребер, s – соответствующая функция задержки.

Определим $q_m = \Psi_m(q)$, $G_m = \Psi_m(G)$. Пусть s_m – соответствующая q_m функция задержки, R_m определена по q_m соотношениями (12.2) и (12.3), P_m обозначает q_m-поток.

По конструкции легко видеть, что $q_m(\sigma) = q_m(\sigma')$ для $\sigma, \sigma' \in G_m$ таких, что $\sigma \sim \sigma'$.

Понятие непрерывности полумеры было введено в разделе 12.4. Имеет место следующее следствие из леммы 12.3.

Следствие 12.6. *Полумера P_m непрерывна для любого m.*

Для применения этой леммы к полумере P_m достаточно заметить, что для любого i число $w(i,q)$ разделяет G_m.

Лемма 12.11. *$P_m(y) = 0$ тогда и только тогда, когда $q_m(\sigma) = 0$ для некоторого ребра σ единичной длины лежащего на y, т.е. такого, что $\sigma_2 \subseteq y$.*

Доказательство. Доказательство необходимости очевидно. Для доказательства достаточности предположим, что $q_m(y^n, y^{n+1}) = 0$ для некоторого $n < l(y)$ и тем не менее $P_m(y) \neq 0$. По определению $s_m(y^n) = 1$. Из $P_m(y) \neq 0$ следует, что должно существовать дополнительное ребро $(x,z) \in G_m$ такое, что $x \subseteq y^n$ и $y^{n+1} \subseteq z$. Будем считать, что (x,z) – самое короткое такое ребро. По определению G_m должно существовать ребро $(x',z') \in G$ такое, что $(x,z) \sim (x',z')$. Значение $s_m(y^n) = 1$ было определено по 1) или 2) определения s_m. По конструкции в этом случае существует ребро $(u,v) \in G$ такое, что $l(v) = n$ (в противном случае было бы $s(z) = 0$ для всех $z \sim y^n$). Из существования в G ребра (x',z') следует, что для любого ребра $(u',v') \in G$, проведенного выше уровня установки задания $p(l(x))$, но ранее чем ребро (x',z') (т.е. для которого $w \leqslant l(v') \leqslant l(z)$, где $w = w(p(l(x)), q^{l(x)})$), его задание имеет большее значение и установлено оно позже, т.е. $p(l(v')) \geqslant p(l(x))$ и $w(p(l(v')), q^{l(v')}) \geqslant w$. В частности, $w(p(n), q^n) \geqslant w$. Поэтому из $z' \sim_w z$ следует, что $(z')^n \sim z^n$. Отсюда, если $s_m(z^n) = 1$ определено по 1), то $s_m((z')^n) = 1$ также будет определено по 1). Пусть $s_m(z^n) = 1$ было определено по 2). Так как $i = p(n) \geqslant p(l(x))$ и $[i]_3 = m$ (т.е. m – мишень задания i), то $[p(l(x))]_2 = m$ (т.е. m – база задания $p(l(x))$) и $i > p(l(x))$. Поэтому из $(z')^n \sim_w z^n$ и из конструкции следует, что $s_m((z')^n) = 1$ также будет определено по 2).

Ребро (x',z') – самое короткое среди всех ребер из G таких, что $(x',z') \sim (x,z)$ для некоторого $(x,z) \in G_m$, для которого $x \subseteq y^n$ и $y^{n+1} \subseteq z$. Поэтому в множестве дополнительных ре-

бер сети $q^{l(z')-1}$ нет ребра с таким свойством. Отсюда и из того, что $q_m((z')^n,(z')^{n+1})=0$, следует, что $Q_m(z')=0$, где Q_m есть $\Psi_m(q^{l(z')-1})$-поток. Это равенство противоречит тому, что $(x',z')\in G$, так как для того, чтобы это ребро попало в G необходимо, чтобы $Q_m(z')\neq 0$. Полученное противоречие доказывает лемму. $\square$

Носителем полумеры P было названо множество

$$E=\{\omega\in\Omega:\forall n(P(\omega^n)\neq 0)\}.$$

Как было замечено, E замкнуто в Ω и $\bar{P}(E)=\bar{P}(\Omega)$. Из леммы 12.11 следует, что отношение $P_m(y)=0$ разрешимо и носителем P_m является множество $E_m=\Omega\setminus\cup_{s_m(x)=1}\Gamma_x$.

Лемма 12.12. *$\bar{P}_m(\mathbf{1})>0$ для любого m.*

Доказательство. Оценим снизу величину $\bar{P}_m(\Omega)$. Пусть

$$S_n=\sum_{u:l(u)=n}R_m(u)-\sum_{\sigma:\sigma\in G_m,l(\sigma_2)=n}q_m(\sigma)R_m(\sigma_1).$$

Из определения функции задержки имеем

$$\sum_{l(u)=n+1}R_m(u)=\sum_{l(u)=n}(1-s_m(u))R_m(u)+ \tag{12.15}$$

$$\sum_{\sigma:\sigma\in G_m,l(\sigma_2)=n+1}q_m(\sigma)R_m(\sigma_1). \tag{12.16}$$

Если m не является ни базой ни мишенью на шаге n задания $p(n)$ (т.е. $[p(n)]_2\neq m$ и $[p(n)]_3\neq m$), то $s_m(u)=0$ для любого u длины n и не существует ни одного пути $\sigma\in G_m$, для которого $l(\sigma_2)=n$. Поэтому $S_{n+1}\geqslant S_n$.

Пусть m – база задания $p(n)$. Предварительно рассмотрим случай $w(p(n),q^{n-1})<n$. Обозначим

$$P(\sigma,u)\Longleftrightarrow l(\sigma_2)=l(u)\&(\sigma_1\sim u^{l(\sigma_1)})\&u\not\sim\sigma_2.$$

Из определения функций задержки имеем

$$\sum_{l(u)=n} s_m(u)R_m(u) \leqslant \sum_{l(u)=n} \left(\sum_{z\sim u} s(z) \right) R_m(u) =$$

$$\sum_{l(u)=n} \left(\sum_{\sigma:P(\sigma,u)} \frac{s(\sigma_1)}{1-s(\sigma_1)} \right) R_m(u) =$$

$$\sum_{\sigma:\sigma\in G, l(\sigma_2)=n} \left(\frac{s(\sigma_1)}{(1-s(\sigma_1)} \sum_{u:P(\sigma,u)} R_m(u) \right) \leqslant$$

$$\sum_{\sigma:\sigma\in G, l(\sigma_2)=n} \left(s(\sigma_1) \sum_{z:z\sim\sigma_1} R_m(z) \right) =$$

$$\sum_{\sigma:\sigma\in G_m, l(\sigma_2)=n} \left(R_m(\sigma_1') \sum_{\sigma':\sigma'\sim\sigma,\sigma'\in G} s(\sigma_1') \right) =$$

$$\sum_{\sigma:\sigma\in G_m, l(\sigma_2)=n} q_m(\sigma)R_m(\sigma).$$

Здесь мы использовали лемму 12.10: (для любого ребра $\sigma \in G(i)$ не существует $\sigma' \in G(i)$ такого, что $\sigma' \sim \sigma$ и $\sigma' \neq \sigma$), а также, неравенство

$$\sum_{z\subset u, l(u)=n, u\neq\sigma_2} R_m(u) \leqslant R_m(z) - s(\sigma_1)R_m(z)$$

для любого $\sigma \in G$ такого, что $l(\sigma_2) = n$, и любого $z \sim \sigma_1$, а также равенство $q_m(\sigma) = \sum_{\sigma':\sigma'\in G,\sigma'\sim\sigma} s(\sigma_1')$, которое следует из того, что $q(\sigma) = s(\sigma_1)$ при $\sigma \in G$. Объединяя полученную оценку с (12.15), получим $S_{n+1} \geqslant S_n$.

Рассмотрим теперь случай $w(p(n), q^{n-1}) = n$. Тогда

$$\sum_{l(u)=n} s_m(u)R_m(u) \leqslant 2^n/\rho(n),$$

так как $s_m(u) = 2^n\rho(n)$ при $l(u) = n$. Объединяя это неравенство с (12.15), получим $S_{n+1} \geqslant S_n - 2^n/\rho(n)$.

Пусть m – мишень задания $p(n)$. Тогда

$$\sum_{l(u)=n} s_m(u)R_m(u) = \sum_{u\in D} R_m(u) \leqslant \sum_{\sigma\in G, l(\sigma_2)=n} 2^{-(\sigma_1+3)}, \quad (12.17)$$

где D – множество всех u длины $l(\sigma_2)$, которые являются $p(n)$-закрытыми последовательностями σ_2 для произвольных $\sigma \in G$. Напомним, что в показателе степени мы отождествляем конечную последовательность σ_2 и ее номер. Объединяя это неравенство с (12.15), получим

$$S_{n+1} \geqslant S_n - \sum_{\sigma\in G, l(\sigma_2)=n} 2^{-(\sigma_2+3)}.$$

Из полученных оценок и $S_0 = 1$ получим

$$S_n \geqslant 1 - \sum_{i=1}^{\infty} 2^i/\rho(i) - \sum_{\sigma\in G} 2^{-(\sigma_2+3)} \geqslant \frac{1}{2}$$

для всех n. Так как $P_m \geqslant R_m$, получим

$$\bar{P}_m(\Omega) = \inf_n \sum_{l(u)=n} P_m(u) \geqslant \inf_n S_n \geqslant \frac{1}{2}.$$

Лемма доказана. □

Лемма 12.13. *При $k \neq m$ для любого t выполнено $F_t(E_k)\cap E_m = \emptyset$.*

Доказательство. Допустим, что существует $\omega \in E_k$, для которой $F_t(\omega) \in E_m$. Рассмотрим задание $i = \langle t, k, m\rangle$. Для любого n пусть D_n – множество всех z длины n, которые являются i-закрытыми последовательностью ω^n. Легко видеть, что из непрерывности P_m следует, что для любого $j < i$

$$\lim_{n\to\infty} \sum_{z\in D_n} P_m(z) \leqslant \lim_{n\to\infty} 2^{w(j,q)} P_m(F_t(\omega^n)) = 0.$$

Кроме этого, $P_k(\omega^n) \neq 0$ для всех n. Отсюда легко видеть, что для любого n конечная последовательность ω^n имеет i-продолжение (в качестве такого продолжения подходит сама ω). По лемме 12.4 существует ребро $\sigma \in G(i)$ такое, что $\sigma_2 \subset \omega$. Тогда $(F_t(\omega))^n$ будет i-закрыто последовательностью σ_2. Поэтому $q_m((F_t(\omega))^n, (F_t(\omega))^{n+1}) = 0$, где $n = l(\sigma_2)$. Отсюда следует, что $F_t(\omega) \notin E_m$. Полученное противоречие доказывает лемму. $\square$

Следующая лемма даст нам возможность применить закон 0 или 1 А.Н.Колмогорова к мере $\bar{P}_m$. Заметим, что эта мера не является нормированной. Это не ведет к потере общности, так как последующие утверждения не зависят от мультипликативного множителя.

Пусть $w(i) = w(i, q)$ и $f_i(\omega) = \omega_i$ для любого i.

Лемма 12.14. *Пусть m произвольное и $[i]_2 = m$. Тогда при $n > w(i)$ случайная величина f_n не зависит от совокупности случайных величин $\{f_j : j \leqslant w(i)\}$ в вероятностном пространстве $(\Omega, \bar{P}_m)$.*

Доказательство. Пусть $y \sim_{w(i)} z$. Из леммы 12.3 следует, что при $l(x) \geqslant w(i)$ в определении значения $s_m(x)$ могут лишь рассматриваться задания $p(l(x)) \geqslant i$, установленные на шагах $w(p(l(x)), q^{l(x)}) \geqslant w(i)$. В том случае, когда $m = [p(l(x))]_3$, т.е. m – мишень задания $p(l(x))$, будет даже $p(l(x)) > i$. Отсюда и по определению множества дополнительных ребер G_m для любого ребра σ такого, что $y^{w(i)} \subseteq \sigma_1 \subseteq y \subseteq \sigma_2$, найдется эквивалентное ему ребро σ' с тем же свойством для z (т.е. $\sigma' \sim_{w(i)} \sigma$ и $\sigma'_1 \subseteq z \subseteq \sigma'_2$) и такое, что $q_m(\sigma) = q_m(\sigma')$.

По свойству $w(i)$ при $l(v) \geqslant w(i)$ формулу (12.2) можно переписать в виде

$$R_m(v) = \sum_{l(\sigma_1) > w(i), \sigma_2 = v} q_m(\sigma) R_m(\sigma).$$

Отсюда и из предыдущего рассуждения легко получить соотношение

$$R_m(v)/R_m(y^{w(i)}) = R_m(v')/R_m(z^{w(i)}) \qquad (12.18)$$

для любых v и v' таких, что $y^{w(i)} \subseteq v \subseteq y$, $z^{w(i)} \subseteq v' \subseteq z$ и $l(v) = l(v')$. Из соотношений (12.4) и (12.18) получаем

$$P_m(v)/P_m(y^{w(i)}) = P_m(v')/P_m(z^{w(i)}) \tag{12.19}$$

для любых y и z таких, что $y \sim_{w(i)} z$. Отсюда и из соотношения (11.7), связывающего P_m и $\bar{P}_m$, получаем

$$\bar{P}_m(v)/\bar{P}_m(y^{w(i)}) = \bar{P}_m(v')/\bar{P}_m(z^{w(i)})$$

для любых y и z таких, что $y \sim_{w(i)} z$. Поэтому условная вероятность $\bar{P}_m(y|x)$ не зависит от выбора начального фрагмента x последовательности y при $l(x) = w(i)$ и $l(y) > w(i)$. Иными словами,

$$\bar{P}_m(xz|x) = \bar{P}_m(x'z|x')$$

для любых x, x', z таких, что $l(x) = l(x') = w(i)$ и $l(z) > 0$.

В частности, случайные величины $f_j(\omega) = \omega_j$ не зависят от совокупности случайных величин $f_s(\omega) = \omega_s$ при $s \leqslant w(i) < j$. □

Следствие 12.7. *Для произвольного m для любого $A \subseteq \Omega$, содержащего вместе с каждой последовательностью все последовательности, отличающиеся от нее к конечном числе битов, будет $\bar{P}_m(A) = 0$ или $\bar{P}_m(A) = \bar{P}_m(\Omega)$, в частности, для любого $\mathbf{a} \in \Upsilon$ будет $\bar{P}_m(\mathbf{a}) = 0$ или $\bar{P}_m(\mathbf{a}) = \bar{P}_m(\mathbf{1})$.*

Доказательство. Пусть $n_1, n_2, \ldots$ – все i такие, что $[i]_2 = m$, взятые в возрастающем порядке. Для применения закона 0 или 1 рассмотрим независимые случайные величины $\tilde{f}_{n_1}, \tilde{f}_{n_2}, \ldots$, где

$$\tilde{f}_{n_i}(\omega) = f_{n_i+1}(\omega) \ldots f_{n_{i+1}}(\omega) = \omega_{n_i+1} \ldots \omega_{n_{i+1}}.$$

Множество A, удовлетворяющее условию следствия, лежит в σ-алгебре, порожденной множеством независимых случайных величин $\tilde{f}_{n_k}, \tilde{f}_{n_{k+1}}, \ldots$ для любого k, а поэтому лежит в остаточной σ-алгебре всей последовательности $\tilde{f}_{n_1}, \tilde{f}_{n_2}, \ldots$. По закону 0 или 1 будет $\bar{P}_m(\mathbf{a}) = 0$ или $\bar{P}_m(\mathbf{a}) = \bar{P}_m(\mathbf{1})$. □

Пусть $\mathbf{a} \in \Upsilon$ и P – р.п. полумера. По определению элемент $i_P(\mathbf{a})$ порождается множеством

$$\{\omega \in A : \frac{d\bar{P}}{d\bar{M}}(\omega) \neq 0\},$$

где $\frac{d\bar{P}}{d\bar{M}}$ – производная Радона – Никодима меры $\bar{P}$ по мере $\bar{M}$, A – произвольное множество, порождающее $\mathbf{a}$. Легко видеть, что $i_P(\mathbf{a})$ не зависит от выбора множества A. Для любого A обозначаем $\bar{A} = \{\omega : \exists\alpha(\alpha \in A \& \alpha \equiv \omega)\}$. Пусть $\mathbf{s}_m = [\bar{E}_m]$. По лемме 12.12 $P_m(\mathbf{s}_m) > 0$, поэтому $\mathbf{s}_m \neq \mathbf{0}$. По лемме 12.13 при $k \neq m$ любые $\alpha \in E_k$ и $\beta \in E_m$ не сводятся друг к другу, поэтому $\mathbf{s}_k \cap \mathbf{s}_m = \mathbf{0}$. Определим

$$\mathbf{d}_m = i_{P_m}(\mathbf{s}_m) = i_{P_m}(\mathbf{1}).$$

По определению $\bar{P}(i_P(\mathbf{a})) = \bar{P}(\mathbf{a})$ для любой р.п. полумеры P. Поэтому $\mathbf{d}_m \neq \mathbf{0}$. Из $\mathbf{d}_m \subseteq \mathbf{s}_m$ следует $\mathbf{d}_k \cap \mathbf{d}_m = \mathbf{0}$ при $k \neq m$. Допустим, что $\mathbf{d}_m = \mathbf{a} \cup \mathbf{b}$, где $\mathbf{a} \neq \mathbf{0}$, $\mathbf{b} \neq \mathbf{0}$ и $\mathbf{a} \cap \mathbf{b} = \mathbf{0}$. Тогда по следствию 12.1 (которое верно также для невычислимых P) $\bar{P}_m(\mathbf{a}) > 0$ и $\bar{P}_m(\mathbf{b}) > 0$, что противоречит следствию 12.7. Полученное противоречие доказывает, что $\mathbf{d}_m$ является атомом для любого m. Теорема доказана. □

12.6. Доказательство теоремы 12.4

Формулировка теоремы и необходимые определения имеются в разделе 12.1.

Пусть $\mathbf{a}_1, \mathbf{a}_2, \ldots$ – все атомы алгебры Υ. В этом разделе будет показано, что $\mathbf{d} = \mathbf{1} \setminus \cup_{i=1}^{\infty}\mathbf{a}_i \neq \mathbf{0}$. Для этого мы построим ненулевой элемент $\mathbf{e}$, в котором не содержится ни одного атома.

Операция Φ будет по произвольной сети q и множеству дополнительных ребер G определять новую сеть q' с тем же множеством дополнительных ребер.

Пусть x – произвольная последовательность длины n, $i = p(n)$ – соответствующее задание. Определим $q'(x, x0) = q'(x, x1) = 0$, если найдется такое ребро σ, для которого $l(\sigma_2) = n$,

$\sigma \in G(i)$ и $F_i(\sigma_2) \subseteq x$. Определим $q'(\sigma) = q(\sigma)$ для любого ребра σ единичной длины, в противном случае.

Для применения общей схемы раздела определим

$$B(i,q,\sigma) \Longleftrightarrow Q(\sigma_2) \neq 0 \& F_i(\sigma_2) \not\subseteq \sigma_2 \& \sum_{F_i(\sigma_2) \subseteq z, l(z)=l(\sigma_2)} Q(z) \leqslant 2^{-(\sigma_1+3)},$$

где буквой Q обозначен $\Phi(q)$-поток.

Пусть сеть q определена по общей схеме раздела 12.2, G – множество дополнительных ребер, s – функция задержки. Определим $q' = \Phi(q)$, s' – функция задержки для сети q', P обозначает q'-поток.

Аналогично тому, как это делалось в разделе 12.5, нетрудно показать, что P – непрерывная р.п. полумера, $\bar{P}(\mathbf{1}) > 0$ при $\rho(n) = (n+2)^2$, и носителем P является множество $E = \Omega \setminus \cup_{s'(z)=1} \Gamma_z$. Покажем, что любые две различные (бесконечные) последовательности ω и ω' из E алгоритмически не сводятся друг к другу. Допустим, что $\omega = F_i(\omega')$ для некоторого i. По лемме 12.4 существует ребро $\sigma \in G(i)$ такое, что $\sigma_2 \subseteq \omega'$. По конструкции $F_i(\sigma_2) \not\subseteq \sigma_2$. Поэтому по определению $q'(\omega^n, \omega^{n+1}) = 0$, где $n = l(\sigma_2)$. Отсюда $\omega \notin E$. Полученное противоречие доказывает, что ω не сводится к ω'.

Определим $\mathbf{e} = [\bar{E}]$ и $\mathbf{f} = i_P(\mathbf{e})$. Допустим, что $\mathbf{x} \subseteq \mathbf{f}$ и $\mathbf{x} \not\subseteq \mathbf{0}$. Выберем $X \in \mathbf{x}$ и положим $X' = X \cap E$. Легко видеть, что $\bar{M}(X') > 0$. По лемме 12.1 $\bar{P}(X') > 0$. Так как $X' \subseteq E$, любые две последовательности из X' не сводимы друг к другу. Представим каким-либо образом $X' = X_1 \cup X_2$, где $X_1 \cap X_2 = \emptyset$ и $\bar{P}(X_1) > 0$, и $\bar{P}(X_2) > 0$. Пусть $\mathbf{x}_1 = [\bar{X}_1]$ и $\mathbf{x}_2 = [\bar{X}_2]$. Тогда $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2$, $\mathbf{x}_1 \neq \mathbf{0}$, $\mathbf{x}_2 \neq \mathbf{0}$ и $\mathbf{x}_1 \cap \mathbf{x}_2 = \mathbf{0}$. Следовательно, $\mathbf{f}$ не включает в себя ни одного атома. Поэтому $\mathbf{d} = \mathbf{1} \setminus \cup_{i=1}^{\infty} \mathbf{d}_i \neq \mathbf{0}$. Теорема доказана. $\square$

12.7. Сводимость атомов

Мы введем отношение сводимости атомов Υ, аналогичное сводимости по Ю.Т.Медведеву (см. [21]), которое позволит различать определенные уже атомы и выделить из множества всех атомов естественные атомы $\mathbf{c}$ и $\mathbf{r}$.

Для любых двух атомов $\mathbf{a}$ и $\mathbf{b}$ алгебры Υ будем считать, что $\mathbf{a}$ сводится к $\mathbf{b}$, если для некоторого множества B, порождающего атом $\mathbf{b}$ и некоторого вычислимого оператора F множество $F(B)$ порождает $\mathbf{a}$. Это эквивалентно тому, что для любых двух множеств A и B, порождающих $\mathbf{a}$ и $\mathbf{b}$, соответственно, $\bar{M}$-почти каждая последовательность из A алгоритмически сводится к некоторой последовательности из B, а к M-почти каждой последовательности из B алгоритмически сводится некоторая последовательность из A.

Ввиду того, что при $k \neq m$ никакие две последовательности $\alpha \in E_k$ и $\beta \in E_m$ алгоритмически не сводимы друг к другу, все атомы $\mathbf{d}_1, \mathbf{d}_2, \ldots$, построенные в разделе 12.5, попарно несравнимы относительно введенной сводимости. Легко видеть, что атом $\mathbf{c}$, порожденный всеми вычислимыми последовательностями, является наименьшим, а атом $\mathbf{r}$, порожденный случайными невычислимыми последовательностями, является наибольшим элементом.

Отсюда, в частности, следует, что каждый из атомов $\mathbf{d}_1, \mathbf{d}_2, \ldots$ порождается нестандартными последовательностями.

12.8. Задачи и упражнения

1. Бесконечное множество натуральных чисел A называется гиперимуннным, если не существует всюду определенной вычислимой функции f такой, что $f(i) > z_i$, где $z_1 < z_2 < \ldots$ – все элементы множества A, расположенные в возрастающем порядке.

Бесконечная двоичная последовательность называется разреженной, если она содержит бесконечно много единиц и не су-

ществует всюду определенной вычислимой функции $f(k)$ такой, что для любого k значение $f(k)$ равно длине какого-нибудь начального фрагмента этой последовательности, содержащего не менее k единиц.

a) Доказать, что множество натуральных чисел является гиперимунным тогда и только тогда, когда его характеристическая последовательность является разреженной.[3]

b) Для любого $\delta > 0$ построить вероятностный алгоритм, который с вероятностью не менее $1 - \delta$ выдает бесконечную разреженную последовательность.

2. Бесконечная двоичная последовательность ω называется сильно разреженной, если она содержит бесконечно много единиц и не существует всюду определенной вычислимой функции $f(x, k)$ такой, что для начального фрагмента x этой последовательности и любого k значение $f(x, k)$ равно длине какого-нибудь начального фрагмента последовательности ω, продолжающего x и содержащего не менее k единиц.

Для любого $\delta > 0$ построить вероятностный алгоритм, который с вероятностью не менее $1 - \delta$ выдает бесконечную сильно разреженную последовательность.

3. Вычислимая функция $f(x)$ детектирует редкие события на бесконечной последлвательности ω, если она определена на всех начальных фрагментах ω и для любого $x \subseteq \omega$ значение $f(x)$ равно длине какого-либо продолжения x вдоль ω, на котором встречается хотя бы одна единица.

Для любого $\delta > 0$ построить вероятностный алгоритм, который с вероятностью не менее $1 - \delta$ выдает бесконечную последовательность ω, содержащую бесконечно много единиц и для которой не существует функции, детектирующей редкие собы-

[3]Характеристической последовательностью множества A называется бесконечная двоичная последовательность $\omega_1\omega_2\ldots$, где

$$\omega_i = \begin{cases} 1, \text{ если } i \in A, \\ 0 \text{ в противном случае.} \end{cases}$$

тия.

Часть VI

Приложение

12.9. Меры, порожденные регулярными вычислимыми операциями

Оператор F применим к бесконечной последовательности ω, если $F(\omega) \in \Omega$.[4]

Предположим, что задана вычислимая мера Q на пространстве Ω. Оператор F называется Q-регулярным, если он применим к Q-почти всем бесконечным последовательностям ω.

В следующем утверждении показано, что вычислимая произвольная мера Q и Q-регулярный оператор порождают вычислимую меру – прообраз меры Q.

Предложение 12.1. *Для любых вычислимой меры Q и вычислимой Q-регулярной операции F функция*

$$P(x) = Q\{\omega \in \Omega : x \subseteq F(\omega)\} \tag{12.20}$$

является вычислимой мерой.

Доказательство. Неравенство $P(x) \geqslant P(x0) + P(x1)$ непосредственно следует из определения $P(x)$. Если $x \subset F(\omega)$ и $F(\omega) \in \Omega$, то $x0 \subset F(\omega)$ или $x1 \subset F(\omega)$. Отсюда следует равенство $P(x) = P(x0) + P(x1)$.

Для того, чтобы вычислить значение $P(x)$ с точностью до 2^{-n}, достаточно найти такие m и s, чтобы $\sum_{l(y)=m} Q^s(F(y)) > 1 - 2^{-(n+1)}$, где $Q^s(y)$ – рациональное приближение к значению $Q(y)$ с точностью до $2^{-(m+n+1)}$. Тогда сумма $\sum_{x \subseteq y, l(y)=m} Q^s(F(y))$ будет рациональным приближением к числу $P(x)$ с точностью до $2^m \cdot 2^{-(n+m+1)} + 2^{-(n+1)} = 2^{-n}$. $\square$

Рассмотрим теперь обратную задачу. Покажем, что любая вычислимая мера Q является прообразом равномерной меры L относительно некоторого L-регулярного вычислимого оператора F, для которого существует обратный вычислимый оператор G.

Предложение 12.2. *Пусть Q – вычислимая мера, L – равномерная мера.*

[4] Определения и утверждения из этого раздела заимствованы из [10].

- *Существует вычислимый L-регулярный оператор F, порождающий меру Q, т.е. такой, что*

$$Q(x) = L\{\omega \in \Omega : x \subseteq F(\omega)\}$$

для всех x.

- *Кроме этого, существует обратный вычислимый оператор G, применимый ко всем бесконечным последовательностям, кроме может быть, вычислимых или лежащих на отрезках Q-меры 0 и такой, что $G(F(\omega)) = \omega$ для Q-почти всех ω.*

Доказательство. Будем рассматривать бесконечные двоичные последовательности как двоичные представления соответствующих вещественных чисел из отрезка $[0, 1]$.[5]

Рассмотрим функцию распределения меры Q:

$$g(t) = Q\{\alpha : \alpha \leqslant t\},$$

где $\alpha \in [0, 1]$ вещественное число (бесконечная двоичная последовательность). Функция $g(t)$ неубывает и непрерывна слева, могут существать интервалы постоянства $[\tau', \tau'']$, где $g(\tau') = g(\tau'') = \rho$, а также скачки в области значений, где $\limsup_{\alpha \nearrow \gamma} g(\alpha) = \sigma'$ и $\liminf_{\alpha \searrow \gamma} g(\alpha) = \sigma'')$, где γ – значение функции g^{-1} на интервале $[\sigma', \sigma'']$.[6]

Оператор F будет вычислять функцию $g^{-1}(\alpha)$, а оператор G будет вычислять функцию $g(\omega)$.

Построим оператор F, который порождает меру Q из равномерной меры L. Пусть на вход подается бесконечная последовательность α. Пусть также $\alpha'_n = \sum_{i=1}^{n} \alpha_i 2^{-i}$ – двоично-рациональное приближение к числу α снизу, а $\alpha''_n = \sum_{i=1}^{n} \alpha_i 2^{-i} +$

[5] Как известно, такие двоичные представления вещественных чисел неоднозначны, например, $0.011\cdots = 0.100\ldots$. Однако, все такие последовательности (представления двоично рациональных чисел) являются вычислимыми. При таком сопоставлении интервал Γ_x из Ω соответствует интервалу $[\sum_{i=1}^{n} x_i 2^{-i}, \sum_{i=1}^{n} x_i 2^{-i} + 2^{-n}]$ из отрезка $[0, 1]$.

[6] Здесь $\tau' < \tau''$ и $\sigma' < \sigma''$.

2^{-n} – приближение сверху с точностью до 2^{-n}. Вычислим все значения $Q(y)$ при $l(y) = n$ с точностью до 2^{-2n}, обозначим каждое такое приближение $Q^s(y)$. Находим все слова z длины n такие, что [7]

$$\sum_{y \geqslant z} (Q^s(y) - 2^{-2n}) \geqslant 1 - \alpha''_n$$

и [8]

$$\sum_{y \leqslant z} Q^s(y) \geqslant \alpha'_n$$

Выберем наиболее длинный общий фрагмент всех таких z и выдадим его в качестве значения $F(\alpha^n)$. Для каждого n каждое такое z определяют интервал, содержащий $g^{-1}(\alpha)$, поэтому если оператор F применим к α, то его значением будет двоичное представление числа $g^{-1}(\alpha)$.

Пусть отрезок $[\sigma', \sigma'']$ соответствует некоторому скачку функции $g(\alpha)$, то есть имеется число (и соответствующее двоичное представление) γ такое, что предел функции $g(\alpha)$ слева при $\alpha \to \gamma$ равен σ', а предел при $\alpha \to \gamma$ справа равен σ''. В этом случае последовательность γ имеет положительную меру L.

Допустим, что α лежит на таком отрезке типа $[\sigma', \sigma'']$. Тогда, если α лежит внутри отрезка $[\sigma', \sigma'']$, то с того момента, как интервал $[\alpha'_n, \alpha''_n]$ (длины 2^{-n}) будет целиком лежать внутри этого отрезка, множество выделенных слов z будет состоять из одного-единственного слова, являющегося фрагментом последовательности γ, и следовательно, процесс F будет применим к α. К концам отрезка $[\sigma', \sigma'']$ процесс F, вообще говоря, может быть неприменим.

Могут также существовать отрезки постоянства функции g, $[\tau', \tau'']$, где $g(\tau') = g(\tau'') = \rho$. Мера Q такого отрезка равна 0.

Пусть теперь α не лежит на отрезке типа $[\sigma', \sigma'']$. Тогда из определения z следует, что $Q(z) \to 0$ при $n \to \infty$, откуда, если α

[7] Приведенная ниже сумма слева есть приближение к $Q(\cup_{y \geqslant z} \Gamma_y)$ с недостатком с точностью до 2^{-n}.

[8] Приведенная ниже сумма слева есть приближение к $Q(\cup_{y \leqslant z} \Gamma_y)$ с избытком с точностью до 2^{-n}.

не является точкой типа ρ, соответствующей отрезку L-меры 0, то и сами отрезки, Γ_z стягиваются к одной точке ω (g-прообразу α). Поэтому длина наибольшего общего фрагмента выделенных слов z стремится к бесконечности за исключением, может быть, тех случаев, когда точка ω – двоично-рациональная и имеет неоднозначное двоичное представление.

Множество всех последовательностей, к которым может быть неприменим оператор F, не более чем счетно и имеет L-меру 0. Поэтому оператор F является L-регулярным.

Обратный оператор G соответствует процессу вычисления функции g, при этом он может быть неприменим к последовательностям типа γ, имеющим положительную Q-меру и поэтому являющимся вычислимыми, а также к точкам отрезков типа $[\tau', \tau'']$, Q-мера которых равна 0. Оператор G также может быть неприменим к последовательностям, на которых функция g принимает двоично-рациональные значения. Утверждение доказано. □

Предложение 12.3.

- *(i) Для любой вычислимой меры P любой P-регулярный оператор применим ко всем случайным по мере P последовательностям.*
- *(ii) Пусть Q – мера, порождаемая вычислимым оператором F из вычислимой меры P, ω – случайная по P последовательность. Тогда последовательность $F(\omega)$ является случайной по мере Q.*

Доказательство. Сначала докажем утверждение (i) от противного. Допустим, что оператор F неприменим к последовательности ω. Тогда $F(\omega)$ – конечная последовательность, пусть ее длина не превосходит некоторого числа k.

Для любого m перечислением находим такое $n \geqslant m$, чтобы объединение всех x длины n и таких, что $l(F(x)) \leqslant k$, имело P-меру не более $2^{-(m+1)}$. Такое n всегда найдется так как мера всех ω таких, что $l(F(\omega)) \leqslant k$ равна 0. Рассмотрим $V_m = \cup_{l(x)=n}\Gamma_x$ и $U_m = \cup_{m' \geqslant m} V_{m'}$. Множество U_m есть объединение перечислимого семейства интервалов и $P(U_m) \leqslant 2^{-m}$ для всех m. Таким

образом, $\{U_m\}$ – тест Мартин-Лефа. Так как $l(F(\omega^n)) \leqslant k$ для всех n, последовательность ω лежит в нулевом множестве этого теста.

Докажем утверждение (ii). Допустим, что ω – случайная по мере P, но при этом последовательность $F(\omega)$ – неслучайная по мере Q, т.е. лежит в нулевом множестве некоторого Q-теста $\{U_m\}$, где $Q(U_m) \leqslant 2^{-m}$ для любого m.

Для произвольного m представим U_m в виде объединения попарно непересекающихся интервалов $U_m = \cup\Gamma_{x_i}$. По определению $Q(\Gamma_{x_i}) = P(\{\omega : x_i \subset F(\omega)\})$ для каждого i. Поэтому существует равномерно перечислимая последовательность $z_{i,j}$ такая, что для каждого i, если $x_i \subset F(\omega)$, то $z_{i,j} \subset \omega$ для некоторого j, а также будет

$$Q(\Gamma_{x_i}) = P(\cup_j\{\Gamma_{z_{i,j}} : x_i \subseteq F(z_{i,j})\}).$$

Обозначим $V_m = \cup_{i,j}\{\Gamma_{z_{i,j}} : x_i \subseteq F(z_{i,j}\}$. Тогда $P(V_m) = Q(U_m) \leqslant 2^{-m}$ для всех m. Таким образом, $\{V_m\}$ являетя P-тестом Мартин-Лефа.

По предположению ддя любого m найдется i такое, что $x_i \subset F(\omega)$, а тогда найдется j такое, что $z_{i,j} \subset \omega$. Отсюда следует, что ω лежит в нулевом множестве теста $\{V_m\}$. Полученное противоречие доказывает утверждение (ii). $\square$

Бесконечная последовательность называется стандартной, если она является случайной относительно некоторой вычислимой меры.

В частности, каждая вычислимая последовательность является стандартной, так она является случайной относительно вычислимой меры, сосредоточенной на этой последовательности.

Две бесконечные последовательности ω и α алгоритмически эквивалентны, если $\omega = F(\alpha)$ и $\alpha = G(\omega)$ для некоторых вычислимых операторов.

В работе [10] было доказано следующее свойство стандартных последовательностей.

Предложение 12.4. *Всякая стандартная последовательность либо вычислима, либо алгоритмически эквивалентна последовательности, случайной относительно равномерной меры.*

Доказательство. Допустим, что последовательность ω невычислимая и является случайной по вычислимой мере Q. Заметим, что ω не может лежать внутри отрезка типа $[\tau', \tau'']$ (см. доказательство предложения 12.2). Такой отрезок имеет Q-меру 0, поэтому можно построить тест, отбрасывающий все последовательности из этого отрезка.

Рассмотрим вычислимый оператор F, порождающий меру Q из равномерной меры L и обратный к нему оператор G. Оба они были определены в доказательстве предложения 12.2.

Так как ω невычислимая и не лежит внутри отрезка типа $[\tau', \tau'']$, согласно свойству обратного оператора G из доказательства предложения 12.2, он применим к ω. Пусть $G(\omega) = \alpha$.

Оператор F применим к α, так как он может быть неприменим только к последовательностям, которые должны отображаться при g в двоично-рациональные точки (а ω не соответствует двоично-рациональной точке, так как она невычислима) и к последовательностям типа ρ, в которые функция g переводит целый отрезок $[\tau', \tau'']$ (см. предложение 12.2), так как $\omega \notin [\tau', \tau'']$

Следовательно, последовательности ω и α алгоритмически эквивалентны: $G(\omega) = \alpha$ и $F(\alpha) = \omega$.

Докажем, что последовательность α является случайной по мере L. Воспользуемся эквивалентным определением теста Мартин-Лефа: это перечислимая снизу функция $U : \Xi \to \mathcal{R}_+$ такая, что $L(\alpha : U(\alpha) \geqslant m\}) \leqslant 2^{-m}$ для всех m, где $U(\alpha) = \sup_n U(\alpha^n)$.

Мера Q задается в виде $Q(A) = L\{\alpha : F(\alpha) \in A\}$. Обозначим $A = \{\omega : U(G(\omega)) \geqslant m\}$.

Покажем, что функция $U(G(\omega))$ является тестом, коррект-

ным относительно меры Q. Действительно,

$$\begin{aligned} Q\{\omega : U(G(\omega)) \geqslant m\} = Q(A) = \\ L\{\alpha : F(\alpha) \in A\} = \\ L\{\alpha : U(G(F(\alpha))) \geqslant m\} = \\ L\{\alpha : U(\alpha) \geqslant m\} \leqslant 2^{-m}. \end{aligned}$$

Здесь мы использовали равенство $U(G(F(\alpha))) = U(\alpha)$. Пусть $\omega = F(\alpha)$ и $\alpha = G(\omega)$. Из равенства $U(G(\omega)) = U(\alpha)$ следует, что если ω – последовательность случайная по мере Q, то α – последовательность случайная по мере L. Утверждение доказано. $\square$

Литература

[1] Бабкин В.Ф. Метод универсального кодирования источника независимыгх сообщений неэкспоненциальной трудоемкости // Проблемы передачи информации. 1971. Т. 7 (4) С. 13–21.

[2] Вовк В.Г. Закон повторного логарифма для случайных по Колмогорову, или хаотических последовательностей // Теория вероятностей и ее применения. 1987. Т. 32 (3). С. 456–468.

[3] Вьюгин В.В. Алгебра инвариантных свойств двоичных последовательностей // *Проблемы передачи информации*, 1982, т.18, N2, с.83–100.

[4] Вьюгин В.В. Алгоритмическая энтропия (сложность) конечных объектов и ее применения к определению случайности и количества информации // Семиотика и информатика: сб. научн. тр. / ВИНИТИ. – М., 1981. – В. 16. С.14–43. Перевод на англ. в Selecta Mathematica formerly Sovietica, 1994. V. 13 (4). P. 357–389.

[5] Вьюгин В.В. О нестохастических объектах // Проблемы передачи информации. 1985. Т.21 (2). С.3–9.

[6] Вьюгин В.В. О дефекте случайности конечного объекта относительно мер с заданными границами их сложности // Теория вероятностей и ее применения. 1987. Т.32 (3). С.558–563.

[7] Вьюгин В.В. Эффективная сходимость по вероятности и эргодическая теорема для индивидуальных случайных последовательностей // Теория вероятностей и ее применения. 1997. Т. 42 (1). С. 35–50.

[8] Вьюгин В.В. Элементы математической теории машинного обучения: учебное пособие. М.: ГОУ ВПО «Московский физико-технический институт» (государственный университет) : ИППИ РАН, 2010. – 231 с.

[9] Данфорд Н., Шварц Дж.Т. Линейные операторы. Общая теория. М.: Изд-во иностр. лит., 1962.

[10] Звонкин А.К., Левин Л.А. Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов // Успехи математических наук. 1970. Т. 25 (6). С. 85–127.

[11] Колмогоров А.Н. Три подхода к определению понятия "количество информации" // Проблемы передачи информациию. 1965. Т. 1 (1). С. 3–11.

[12] Колмогоров А.Н. Комбинаторные основания теории информации и исчисления вероятностей // Успехи математических наук. 1983. Т.38 (4). С.27–36.

[13] Кричевский Р.Е. Связь между избыточностью кодирования и достоверностью сведений об источнике // Проблемы передачи информации. 1968. Т. 4 (3). С. 48–57.

[14] Кричевский Р.Е. Оптимальное кодирование источника на основе наблюдений // Проблемы передачи информации. 1975. Т. 11(4). С.37—42.

[15] Левин Л.А. О понятии случайной последовательности // Доклады АН СССР. 1973. Т. 212 (3), С. 548–550.

[16] Левин Л.А. Законы сохранения (невозрастания) информации и вопросы обоснования теории информации // Проблемы передачи информации. 1974. Т. 10 (3). С. 30–35.

[17] Левин Л.А. О принципе сохранения информации в интуиционистской математике // *Доклады АН СССР*, 1976, т.227, N6, с.1293–1296.

[18] Левин Л.А. О различных мерах сложности конечных объектов (аксиоматическое описание) // Доклады АН СССР. 1976. Т. 227 (4). С. 804–807.

[19] Левин Л.А. Об одном конкретном способе задания сложностных мер // Доклады АН СССР. 1977. Т. 234 (3). С. 536–539.

[20] де Леу К., Э.Ф. Мур Э.Ф., Шеннон К.Е., Шапиро Н., Вычислимость на вероятностных машинах, Автоматы (сб. переводов), М., ИЛ, 1956 (de Leeuw K., Moore E.F., Shannon C.E., Shapiro N. Computability by Probabilistic Machines. – In: *Automata studies*, edited by C.E. Shannon and J. McCarthy, *Annals of Mathematics Studies*, **34**, lithoprinted, Princeton University Press, 183–212).

[21] Роджерс Х. Теория рекурсивных функций и эффективная вычислимость. М.: Мир., 1972.

[22] Рябко Б.Я. Сжатие данных с помощью стопки книг // Проблемы передачи информации. 1980. Т. 16 (4). С. 16–21.

[23] Рябко Б.Я. Дважды универсальное кодирование // Проблемы передачи информации. 1984. Т. 20 (3). С. 24–28.

[24] Шень А.Х. Понятие (α, β)-стохастичности по Колмогорову и его свойства // Доклады АН СССР, 1983, т.271 (6). С.1337–1340.

[25] Ширяев А.Н. Вероятность. – М.: МЦНМО, 2007. 968 с.

[26] Успенский В.А., Семенов А.Л. Теория алгоритмов: основные открытия и приложения. – М.: Наука. Гл. ред. физ.-мат. лит., 1987. (Библиотечка программиста).

[27] Успенский В.А., Верещагин Н.К., Шень А. Колмогоровская сложность и алгоритмическая случайность. – М.: МЦНМО, 2010. 556 с. Доступно онлайн: http://www.lif.univ-mrs.fr/ ashen/nafit/kolmbook.pdf

[28] Успенский В.А., Семенов А.Л., Шень А.Х. Может ли (индивидуальная) последовательность нулей и единиц быть случайной? // Успехи математических наук. 1990. Т. 45 (1). С. 105–162.

[29] Фитингоф Б.М. Оптимальное кодирование при неизвестной и меняющейся статистике сообщений // Проблемы передачи информации. 1966. Т. 2 (2). С. 3–11.

[30] Фитингоф Б.М. Сжатие дискретной информации // Проблемы передачи информации. 1967. Т. 3 (3). С. 28–36.

[31] Шеннон К. Работы по теории информации и кибернетике. – пер. с англ. под. ред. Р. Л. Добрушина и О.Б. Лупанова; предисл. А. Н. Колмогорова. М., 1963. (Shannon C. E. A Mathematical Theory of Communication. Bell Systems Technical Journal. July and Oct. 1948 // Claude Elwood Shannon. Collected Papers. N. Y., 1993. P. 8-111.).

[32] Шенфилд Дж. Степени неразрешимости. М.: Наука, 1977.

[33] Штарьков Ю.М. Универсальное последовательное кодирование отдельных сообщений // Проблемы передачи информации 1987. Т. 23 (3). С. 3–17.

[34] Ю. М. Штарьков Ю.М. Универсальное кодирование. Теория и алгоритмы. 2013 ISBN: 978-5-9221-1517-9

[35] Barron, A.R., Rissanen, J., Bin Yu The Minimum Description Length Principle in Coding and Modeling // IEEE Transactions on Information Theory. 1998. V. 44 (6). P. 2743–2760.

[36] Bishop E. Foundation of Constructive Analysis. New York: McGraw-Hill, 1967.

[37] Chaitin G. Information-theoretical limitations of formal systems // Journal of the ACM. 1974. V. 21. P. 403–424.

[38] Chaitin G. A theory of program size formally identical to information theory // J. Assoc. Comput. Mach., 1975, V.22, P.329–340.

[39] Cover T.M., Gács P., Gray R.M. Kolmogorov's contributions to information theory and algorithmic complexity // Annals of Probability. 1989. V.17 (1). P.840–865.

[40] Cover T. M.. Thomas J. A. Elements of Information Theory. New York: Wiley, 1991.

[41] Dawid A.P. Calibration-based empirical probability [with discussion] // The Annals of Statistics. 1985. V.13. P.1251–1285.

[42] Gács P. Exact expressions for some randomness tests // Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik. 1980. V. 26. P. 385–394.

[43] Gács P. Lecture notes on descriptional complexity and randomness, Boston University, 1997. 191 p. Доступно онлайн: http://www.cs.bu.edu/ gacs/papers/ait-notes.pdf

[44] Krichevsky, R. E. and Trofimov V. K. The Performance of Universal Encoding // IEEE Trans. Inf. Theory, 1981. IT-27(2) P. 199–207.

[45] Kucera A. Measure, Π^0_1 classes, and complete extensions of PA // Lecture Notes in Mathematics. 1985. V. 1141. P. 245–259.

[46] van Lambalgen M. Random sequences. Amsterdam: Academish Proefshri't. 1987.

[47] Levin L.A., V'yugin V.V. Invariant properties of informational bulks // *Springer Lecture Notes on Computer Science* v.53, 1977, p.359–364.

[48] Li M., Vitányi P. An Introduction to Kolmogorov Complexity and Its Applications, 2nd ed. New York: Springer–Verlag, 1997.

[49] Lugosi G., Cesa-Bianchi N. Prediction, Learning and Games. Cambridge University Press, New York, 2006.

[50] Muchnic An.A., Semenov A.L., Uspensky V.A. Mathematical metaphysics of randomness // Theoretical Computer Science. 1998. V.207. P.263–317.

[51] Shafer G., Vovk V. Probability and Finance. It's Only a Game! New York: Wiley, 2001.

[52] Schnorr C.P. Process complexity and effective random tests // Journal of Computer and System Sciences. 1973. V. 3 (4). P. 376–378.

[53] Solomonoff R.J. A formal theory of inductive inference I, II // Information and Control. 1964. V. 7. P. 1–22. P. 224–254.

[54] Solomonoff R.J. Complexity-based induction systems: Comparisons and convergence theorems // IEEE Transactions on Information Theory. 1978. IT-24. P. 422–432.

[55] Ville J. Etude critique de la notion de collectif. Gauthier-Vllars, Paris, 1939.

[56] Vovk V.G. Aggregating strategies. In M. Fulk and J. Case, editors, Proceedings of the 3rd Annual Workshop on Computational Learning Theory, p.371–383, San Mateo, CA, 1990. Morgan Kaufmann.

[57] Vovk V.G. and V'yugin V.V. Prequential Level of impossibility with some applications // Journal of the Royal Statistical Society B. 1994. V.56. P.115–123.

[58] Vovk V.G. A game of prediction with expert advice // Journal of Computer and System Sciences. 1998. V.56, P.153–173.

[59] Vovk V.G., Watkins C.J.H.C., Universal portfolio selection // Proceedings of the 11th Annual Conference on Computational Learning Theory. 1998. P.12–23.

[60] Vovk V.G., Gammerman A., Complexity estimation principle // The Computer Journal. 1999. V.42 (4). P.318–322.

[61] V'yugin V.V. Algorithmic complexity and stochastic properties of finite binary sequences // The Computer Journal, 1999, v.42, N4, p.294–317.

[62] V'yugin V.V. Most Sequences are Stochastic // Information and Computation. 2001. V.168. P.1–12.

[63] Ziv, J., Lempel, A. A Universal Algorithm for Sequential Data Compression // IEEE Transactions on Information Theory. 1977. V. 23 (3). P. 337–343.

[64] Ziv J., Lempel A. Compression of Individual Sequences via Variable-Rate Coding // IEEE Transactions on Information Theory. 1978. V. 24 (5): P. 530–536.