МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное агентство по образованию
Московский физико-технический институт
(государственный университет)
Учреждение Российской академии наук
Институт проблем передачи информации им. А.А. Харкевича
РАН

В.В.Вьюгин

КОЛМОГОРОВСКАЯ СЛОЖНОСТЬ И ТЕОРИЯ ИНФОРМАЦИИ

Допущено
Учебно-методическим объединением
высших учебных заведений Российской Федерации
по образованию в области прикладных математики и физики
в качестве учебного пособия для студентов

по направлению «Прикладные математика и физика»

МОСКВА
МФТИ
2012

УДК 005.519.8(075.8)
ББК 65.290-2в6я73

Рецензенты:

Кафедра математической логики и теории алгоритмов механико-математического факультета Московского государственного университета им. М.В. Ломоносова

Зав. кафедрой, доктор физико-математических наук, профессор В.А. Успенский

Доктор физико-математических наук А.В. Бернштейн

Вьюгин В.В.

Колмогоровская сложность и теория информации: учеб. пособие

– М.: МФТИ : ИППИ РАН, 2012. – 140 с.

Предназначено 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.

Для студентов и аспирантов математических и прикладных математических специальностей.

Библ. 28.

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 [7], 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.

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 [22].

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 [20], Р.Е. Кричевского [8], Ю.М. Штарькова [23], [24], Б.Я.Рябко [14], [15], Д. Риссанена [34] 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 [37], [38] 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 [7], 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, вызы-

вает 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 [7], 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 сложности было также введено Г. Чейтиным [26], [27].

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 колмого-

ровским 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. Левин и Шнорр ввели в работах [9], [11], [10], [12] и [36] 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, эквивалентное определению 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 Соломонова были не ясны, он уточнил их позже в статьях [37] и [38]. 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 Л. А. Левин построил полувычислимый универсальный предсказатель. Здесь наиболее ценной является идея Соломонова об универсальности предсказателя, ко-

торая 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. [6]. Это понятие изучается в главе 6.

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. [33] и [5]).

Близкие определения случайности с использованием теории мартингалов позже привели к новому теоретико-игровому обоснованию теории вероятностей и финансовой математики, предложенному Вовком и Шейфером [35].

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 вероятностных 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 приложениям алгоритмического подхода к теории вероятностей. В главе 7 излагаются некоторые результаты алгоритмического анализа теории вероятностей, в частности, в разделе 7.1 приведено «сложностное» доказательство закона повторного логарифма, предложенного Вовком [2]. В разделе 7.2.1 проводится алгоритмический анализ эргодической тео-

ремы Биркгофа, предложенный в статье [4].

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 [6]. Изложение колмогоровской концепции случайности и новые результаты в области алгоритмической теории информации были представлены в 1981 г. в обзоре Вьюгина [3]. Здесь впервые были приведены доказательства новых результатов Левина, опубликованных (без доказательства) в 1970-х годах в статьях [9], [10], [11]. В 1990 г. был опубликован обзор Успенского и др. [18].

За рубежом общепринятым источником в области колмогоровской сложности является монография Ли и Витаньи [32]. В настоящее время наиболее полное изложение теории колмогоровской сложности представлено в монографии Успенского, Верещагина, Шеня [18]. Для более глубокого изучения предмета рекомендуются лекции П. Гача [30].

Данное учебное пособие составлено на основе курса «Колмогоровская сложность и ее приложения», прочитанного автором на факультете прикладной математики и управления Московского физико-технического института в 2011 г.

Часть I книги основана на монографии [28] и была добавлена при обновлении 2020г. Часть II данного учебного пособия (главы 3–6) представляет собой расширенное изложение обзорной работы [3], которая была дополнена рядом результатов и доказательств из работы [39] и монографии [18]. При этом содержание раздела 6.2 основано на материале из статьи [38] и ее изложении в монографии [32]. Часть III (глава 7) основана на материале статей [2] и [4].

Часть I

Вероятностная теория 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 3.

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.

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 $y = y_1 y_2 \ldots y_n$ можно записывать одно после другого: $xy = x_1 x_2 \ldots x_n y_1 y_2 \ldots y_n$ – 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. $\triangle$

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. 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

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

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. $\triangle$

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

Теорема 1.2. *Код является однозначно декодируемым тогда и только тогда, когда* $\sum\limits_{i=1}^{k} 2^{-l_i} \leqslant 1$, *где* $l_1, l_2, \ldots, l_k$ *– длины кодовых слов.*

Доказательство. В одну сторону утверждение следует из теоремы 1.1.

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 $X(L) = \{x_1, \ldots, x_M : \sum\limits_{i=1}^{M} l_{x_i} = L\}$. 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

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

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

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 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.

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

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

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

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

$\triangle$

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. $\triangle$

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

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

$\triangle$

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. $\triangle$

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'. $\triangle$

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

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

где сходимость – по вероятности, т.е.

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

при $n \to \infty$, где $H = H(X_1)$.

Для произвольного $\epsilon > 0$ рассмотрим множество

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

Свойства множества A_ϵ^n представлены в следующей теореме.

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

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

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

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. $\triangle$

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

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

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. Кодирование стационарных процессов

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* $a_1, a_2, \ldots$, *если* $a_n \to a$ *при* $n \to \infty$, *то* $\frac{1}{n}\sum_{i=1}^{n} a_i \to a$.

По правилу цепи

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

По лемме Чезаро предел

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

существует. $\triangle$

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 [22] 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 $H(X, X) = H(X)$, $H(Y|X) \leqslant H(Y)$, $I(X : X) = H(X)$.

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 $H(X|Y, Z) \leqslant H(X|Y)$.

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

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 [40], [41], [28].

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, $L = k$ – 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 $E[R_1(u)|U_0 = u] = \frac{1}{p(u)}$. $\triangle$

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. $\triangle$

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

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

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

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

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

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

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

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 [13].

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 [17].

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 $g(n) = g(n - 1)$. $\triangle$

[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. $\triangle$

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 [7] 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. $\triangle$

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:*

$$\mathrm{K}(x|y) - O(1) \leqslant \mathrm{K}(x) \leqslant \mathrm{K}(x|y) + \mathrm{K}(y) + 2\log \mathrm{K}(y) + O(1). \quad (3.6)$$

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 правой части неравенства (3.6). $\triangle$

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

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

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 (3.7).

Верно также соотношение

$$\mathrm{K}(x|y) \leqslant \mathrm{K}(x|\psi(y)) + O(1). \tag{3.8}$$

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 (3.8).

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

$$\mu(x) = \min\{y : \psi(y) > x\}. \tag{3.9}$$

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

Так как $\psi(x)$ неограничена, функция $\mu(x)$ является всюду определенной и вычислимой. По свойству (3.7)

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

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 (3.9) может не существовать алгоритма. Однако совсем не обязательно искать минимум в (3.9), 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. $\triangle$

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 [26].

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. $\triangle$

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

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

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:

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

где

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

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

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

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

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

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

$$\mathrm{K}(C_n^k) \leqslant \log n + \log k + 2\log\log k + O(1) \leqslant$$
$$\leqslant 3\log n + O(1). \qquad (3.13)$$

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:

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

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 [7] 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:

$$d(x|D) = \log |D| - K(x|D),$$

где $K(x|D)$ – условная сложность конечной последовательности x относительно конечного объекта D.

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 5.4

3.4. Сложность пары

В этом разделе докажем теорему Колмогорова–Левина о декомпозиции сложности пары. Впервые это доказательство было опубликовано в обзоре [6].

Теорема 3.3.

$$\mathrm{K}(x, y) = \mathrm{K}(x) + \mathrm{K}(y|x) + O(\log \mathrm{K}(x, y)). \tag{3.16}$$

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 пару (x, y). Длина такого кода равна правой части неравенства (3.16). 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

$$\mathrm{K}(y|x) \leqslant \log |A_x| + 2l(\mathrm{str}(a)) + O(1) \leqslant$$
$$\leqslant m + 2 \log \mathrm{K}(x, y) + O(1) \leqslant m + 3 \log \mathrm{K}(x, y) + O(1). \quad (3.17)$$

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

$$\mathrm{K}(x) \leqslant a - m + O(\log \mathrm{K}(x, y)). \tag{3.18}$$

Сложим неравенства (3.17) и (3.18) и получим необходимое неравенство

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

Теорема доказана. $\triangle$

Задача 10 из раздела (3.6) утверждает, что оценка (3.16) не улучшаемая.

3.5. Количество информации

На основе понятия сложности А. Н. Колмогоров предложил в [7] 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 5.1.5

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

Отсюда получаем

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

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 вероятности 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. 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 $\epsilon > 0$.

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

- $U_m = \cup\{\Gamma_x : (m, x) \in T\}$,

- $L(U_m) \leqslant 2^{-m}$ 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.

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

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

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

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

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 $\{U^i_n\}$ одно множество $U^i_{n_i}$ такое, что $L(U^i_{n_i}) < 2^{-m-i}$, и определим семейство эффективно открытых множеств $U_m = \cup_i U^i_{n_i}$. 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 $\{U^i_m, m = 1, 2, \dots\}$, $i = 1, 2, \dots$, 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

$$L\left(\cup_{x'}\{\Gamma_{x'} : (m, x') \in T_i^{s-1}\} \cup \Gamma_x\right) \leqslant 2^{-m}. \tag{4.2}$$

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. $\triangle$

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

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

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 $\{U^i_m\}$ будет выполнено

$$U^i_{m+i} \subseteq U_m$$

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

$$\cap_n U^i_n \subseteq \cap_m U_m.$$

Теорема 4.1 доказана. $\triangle$

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.

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 [1] . 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.

4.3. Задачи и упражнения

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

[1] 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».

(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.

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 *без префиксным*, если для любых $p, q \in C$ будет $p \not\sqsubseteq q$ и $q \not\sqsubseteq p$.

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 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*

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

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 $\bar{U}(q,p,y)$ такая, что для любой вычислимой префиксно-корректной по p функции $B(p,y)$ существует q такое, что $B(p,y) = \bar{U}(q,p,y)$ для всех p и y.*

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

Развернем процесс вычисления значений функции $U(q,p,y)$: на каждом шаге s этого процесса делается один шаг вычисления значения $U(q,p,x)$ для одного из наборов (q,p,y). Для этого

мы каким-либо образом просматриваем каждый набор (q, p, y) на бесконечном числе шагов процесса.

FOR $s = 1, 2, \ldots$

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

Если значение $\bar{U}^{s-1}(q, p, y)$ определено, то полагаем $\bar{U}^s(q, p, y) = \bar{U}^{s-1}(q, p, y)$. В противном случае делаем следующее. Определим $\bar{U}^s(q, p, y) = U(q, p, y)$, если значение $U(q, p, y)$ определено за $\leqslant s$ шагов вычисления и множество, состоящее из p и всех p' таких, что значения $\bar{U}^{s-1}(q, p', y)$ определены, является безпрефиксным. В противном случае значение $\bar{U}^s(q, p, y)$ оставляем неопределенным.

Полагаем $\bar{U}^s(q', p', y') = \bar{U}^{s-1}(q', p', y')$ для всех остальных троек (q', p', y') [1] .

ENDFOR

По построению функция $\bar{U}^s(q, p, y)$ обладает следующими свойствами:

- для любого s функция $\bar{U}^s(q, p, y)$ является префиксно-корректной по p;
- если $\bar{U}^s(q, p, y)$ определено, то $\bar{U}^s(q, p, y) = \bar{U}^{s+1}(q, p, y)$;
- если функция $U(q, p, y)$ является префиксно-корректной по p, то для любых p и y имеет место $\bar{U}^s(q, p, y) = U(q, p, y)$ для всех достаточно больших s.

Определим $\bar{U}(q, p, y) = \bar{U}^s(q, p, y)$ для наименьшего s такого, что правая часть определена, в противном случае значение $\bar{U}(q, p, y)$ считаем неопределенным.

Легко видеть, что функция $\bar{U}(q, p, y)$ удовлетворяет условиям:

- $\bar{U}(q, p, y)$ префиксно-корректная по p;

[1] В частности, если правая часть этого равенства неопределена, то неопределена и левая часть.

- если функция $B(p, y)$ является префиксно-корректной по p, то найдется такое q, что $B(p, y) = \bar{U}(q, p, y)$ выполнено для всех p и y;
- если $U(q, p, y)$ – префиксно-корректная по p, то выполнено $\bar{U}(q, p, y) = U(q, p, y)$ для всех значений своих аргументов.

Лемма доказана. $\triangle$

Переходим к доказательству теоремы 5.1.

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

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

для всех $p, q, y \in \{0, 1\}^*$.

Для всех остальных входов, не имеющих вида $(\overline{q}01p, y)$, значение этой функции не определено.

Легко видеть, что эта функция является префиксно-корректной.

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

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

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

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

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

Безпрефиксный метод декодирования, удовлетворяющий заключению теоремы 5.1, называется универсальным. Фиксируем один из универсальных 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

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

Прежде всего приведем соотношения, связывающие префиксную и простую 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. $\triangle$

Более нетривиальные соотношения для префиксной сложности будут доказаны в следующем разделе.

5.1.2. 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.

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.

Мы введем структуру эффективной вычислимости для функций с вещественными значениями. Область определения функции типа $f : \mathcal{N}^+ \to \mathcal{R}$ состоит из конструктивных объектов [2] . 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. Функция f перечислима снизу, если 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.

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 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.

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

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 по-

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

ложительному 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 $f(x)$.

Функция $f : \mathcal{N}^+ \to \mathcal{R}^+$ называется перечислимой снизу полумерой, если она перечислима снизу и

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

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);*
- *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 $W_i^0 = 0$. Таким образом, W_i^s обозначает конечное подмножество пар, перечисленных в W_i за s шагов процесса вычисления универсальной функции. Определим

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

Функция $P_i^s(x) > 0$ для не более чем конечного числа x. Определим

$$P_i(x) = \sup_s \{P_i^s(x) : \sum_y P_i^s(y) \leqslant 1\}.$$

Легко видеть, что P_i является перечислимой снизу полумерой для каждого i и, более того, 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. Легко видеть, что $Q(x) = P_i(x)$ для всех x. $\triangle$

Доказательство теоремы. Определим

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

Функция P перечислима снизу [3] . Она является полумерой, так как

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

Кроме этого, 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. $\triangle$

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 $c > 0$

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

[3] Как двойной предел неубывающей последовательности функций с рациональными значениями.

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.*

5.1.3. 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. 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 [27].

Мы сформулируем соответствующее утверждение.

Предложение 5.2. *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.

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

Если на некотором входе x машина $\mathcal{M}_1$ остановилась и напечатала выходное слово y, то проверяем, будет ли $p' \subseteq x$.

Если $p' \not\subseteq x$, то продолжаем параллельное моделирование работы $\mathcal{M}_1$, а с этим x больше ничего не делаем.

(a) Если $p' \subset x$, то машина $\mathcal{M}_2$ читает следующий бит b входа и в дальнейшем обрабатывает слово $p'' = p'b$. В случае $p' \subset x$ 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 $\mathcal{M}_2$.

После этого машина $\mathcal{M}_2$ проверяет, будет ли $p'' \subseteq x$. Если $p'' \not\subseteq x$, то продолжаем параллельное моделирование работы $\mathcal{M}_1$, а с этим x больше ничего не делаем.

(a1) Если $p'' = x$, то машина $\mathcal{M}_2$ выдает y в качестве результата работы над входным словом p и прекращает работу. Вычисление проведено корректно, так как из свойства префиксной корректности функции B машина $\mathcal{M}_1$ может быть определена на не более чем одном начальном фрагменте входного слова p (включая его самого).

(a2) Если $p'' \subset x$, то $\mathcal{M}_2$ возвращается к пункту (a).

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

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

$$\sum_n 2^{-\mathrm{KP}(n)} = \sum_{n=0}^{\infty} L(\Gamma_{p_n}) \leqslant 1,$$

где все двоичные строки $p_0, p_1, \ldots$ попарно несравнимы, поэтому интервалы Γ_{p_n} попарно не пересекаются. Кроме этого, функция $Q(n) = 2^{-\mathrm{KP}(n)}$ перечислима снизу.

Следовательно, функция $Q(n)$ есть перечислимая снизу полумера на натуральных числах. Отсюда следует, что

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

для всех n, 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 $x_1, \ldots, x_n$ такие, что $l(x_i) = k_i$ при $i = 1, \ldots, n$. Данное неравенство используется в теории информации для эффективного построения

префиксных кодов [4] . 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 $x_1, x_2, \ldots$ такую, что $l(x_i) = k_i$ при $i = 1, 2, \ldots$

Доказательство. Построим нужную последовательность по индукции. Пусть $x_1, \ldots, x_n$ уже определены так, что выполнено $l(x_i) = k_i$ при $i = 1, \ldots, n$. Найдем x_{n+1}. Используем предположение индукции:

$$\Omega \setminus \cup_{i=1}^{n} \Gamma_{x_i} = \cup_{i=1}^{m} \Gamma_{t_i},$$

где $l(t_i) \neq l(t_j)$ при $i \neq j$.

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

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

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

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

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

Если $l(t_1) = k_{n+1}$, то определим $x_{n+1} = t_1$. В этом случае

$$\Omega \setminus \cup_{i=1}^{n+1} \Gamma_{x_i} = \cup_{i=2}^{m} \Gamma_{t_i},$$

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 $x_{n+1} = a_s$. Предположение индукции выполнено, так как

$$\Omega \setminus \cup_{i=1}^{n+1} \Gamma_{x_i} = \cup_{i=2}^{m} \Gamma_{t_i} \cup_{i=2}^{s} \Gamma_{a_i}.$$

Лемма доказана. $\triangle$

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 неопределено.

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

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

Приведем некоторые неравенства, которые имеют место для префиксной сложности.

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 5.2 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).

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

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

Действительно, сходится ряд

$$\sum_x 2^{-l(x)-\mathrm{KP}(l(x))} = \sum_m \sum_{l(x)=m} 2^{-l(x)-\mathrm{KP}(l(x))} =$$
$$= \sum_m 2^m 2^{-m-\mathrm{KP}(m)} = \sum_m 2^{-\mathrm{KP}(m)} < \infty. \qquad (5.2)$$

В частности, имеет место неравенство

$$\mathrm{KP}(n) \leqslant \log n + 2\log\log n + O(1).$$

Можно его усилить:

$$\mathrm{KP}(n) \leqslant \log n + \log\log n + 2\log\log\log n + O(1). \qquad (5.3)$$

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*

$$\sum_{n=1}^{\infty} a_n = \infty,$$

имеет место неравенство

$$\mathrm{KP}(n) \geqslant -\log a_n$$

для бесконечно многих n.

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 (5.3) 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 5.5.

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. $\triangle$

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

$$cP(y|x, \mathrm{KP}(x)) \geqslant \frac{P(x, y)}{P(x)} = 2^{\mathrm{KP}(x)} P(x, y),$$

где c – 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

$$c_1 2^{-\mathrm{KP}(x)} \geqslant P(x) \geqslant \sum_y P(x, y),$$

для некоторой константы c_1 или

$$\sum_y c_1^{-1} 2^{\mathrm{KP}(x)} P(x,y) \leqslant 1. \tag{5.4}$$

Если бы функция $Q(y|x,m) = c_1^{-1} 2^m P(x,y)$, 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 (5.4) 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

$$c_1^{-1} 2^m \sum_{z \leqslant t} P^t(x,z) \leqslant 1,$$

и полагаем

$$Q^s(y|x,m) = \begin{cases} c_1^{-1} 2^m P^t(x,y), \text{ если } y \leqslant t, \\ 0, \text{ если } y > t. \end{cases}$$

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 (5.4),

$$c_1^{-1} 2^{\mathrm{KP}(x)} \sum_{z \leqslant t} P^t(x,z) \leqslant$$
$$\leqslant c_1^{-1} 2^{\mathrm{KP}(x)} \sum_z P(x,z) \leqslant 1$$

для каждого t, выполнено

$$Q(y|x.\mathrm{KP}(x)) = c_1^{-1} 2^{\mathrm{KP}(x)} P(x,z).$$

Отсюда получаем

$$P(y|x,\mathrm{KP}(x))O(1) \geqslant 2^{\mathrm{KP}(x)} P(x,y).$$

Теорема доказана. $\triangle$

5.2. Монотонные способы декодирования

В современной практике и теории информации широко используются алгоритмы, кодирующие информационные массивы, получаемые потоком в режиме онлайн. Процессы 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 $f(x) \subseteq f(y)$.

Для определение монотонных способов декодирования мы будем использовать определение более общего характера.

Заданием вычислимой операции называется перечислимое множество пар конечных последовательностей $\hat{F}$, которое обладает свойствами:

1) для любых пар $(x,y),(x'y') \in \hat{F}$, если $x \subseteq x'$, то $y \subseteq y'$ или $y' \subseteq y$;

2) если $(x,y) \in \hat{F}$, то $(x,y') \in \hat{F}$, то для любых y' таких, что $y' \subseteq y$.

При таком определении для любой конечной или бесконечной последовательности α, если $(x,y),(x',y') \in \hat{F}$ и $x \subseteq \alpha$, $x' \subseteq \alpha$, то одна из последовательностей y или y' продолжает другую. Это свойство обеспечивает корректность следующего определения вычислимой операции.

Для произвольной конечной или бесконечной последовательности α определим значение вычислимой операции:

$$F(\alpha) = \sup\{y : (x \subseteq \alpha \& (x,y) \in \hat{F}\}. \tag{5.5}$$

Здесь под sup понимается объединение множества попарно согласованных слов в одно слово. Эквивалентная запись (5.5):

$$F(\alpha) = \sup_n\{y : (\alpha^n, y) \in \hat{F}\}.$$

Для любой вычислимой операции F определим меру монотонной сложности

$$\mathrm{KM}_F(x) = \min\{l(p) : x \subseteq F(p)\}.$$

Учитывая свойство 2), это определение эквивалентно следующему равенству:

$$\mathrm{KM}_F(x) = \min\{l(p) : (p,x) \in \hat{F}\}.$$

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*

$$\hat{F}_i = \{(p, y) : (i, x, y) \in \mathcal{F}\} \tag{5.6}$$

  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* (5.6)*).*

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. $\triangle$

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. $\triangle$

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.

Нетрудно доказать, что $\mathrm{KM}(x) \leqslant l(x) + O(1)$.

Монотонная сложность связана со сложностями других видов следующим образом:

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

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 $\omega^n = \omega_1 \ldots \omega_n$ 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

$$\{\mathrm{KM}(\omega^n) \ngeqslant n + O(1)\} = \{\omega : \forall m \exists n (\mathrm{KM}(\omega^n) < n - m)\} \quad (5.7)$$

является эффективно нулевым.

Предварительно докажем лемму.

Лемма 5.7. *Для любой последовательности попарно несогласованных слов $x_1, x_2, \ldots$ 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 $p_1, p_2, \ldots$ также попарно не согласованы и поэтому $\sum_i 2^{-l(p_i)} \leqslant 1$. $\triangle$

Переходим к доказательству теоремы. Определим

$$U_{m,n} = \cup\{\Gamma_x : l(x) \leqslant n \& l(x) - \mathrm{KM}(x) > m\}.$$

Представим это множество в виде объединения 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. Тогда

$$L(U_{m,n}) \leqslant \sum_{j=1}^{k} 2^{-l(x_j)} < 2^{-m} \sum_{j=1}^{k} 2^{-\mathrm{KM}(x_j)} < 2^{-m}$$

для всех m и n. Пусть

$$U_m = \cup_n U_{m,n} = \cup\{\Gamma_x : l(x) - \mathrm{KM}(x) > m\}.$$

$U_{m,n} \subseteq U_{m,n+1}$ для всех m и n. Поэтому $L(U_m) \leqslant 2^{-m}$ для всех m.

Если ω принадлежит множеству (5.7), 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 – произвольный тест Мартин-Лефа. Мы докажем более сильное утверждение: если $\omega \in U$, то величина $n - \mathrm{KP}(\omega^n)$ неограничена [5] .

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. 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 14 из раздела 5.5.

[5] Выполнено $n - \mathrm{KM}(\omega^n) \geqslant n - \mathrm{KP}(\omega^n) - O(1)$.

Определим семейство равномерно перечислимых снизу полумер на множестве всех натуральных чисел:

$$P_m(x) = \begin{cases} 2^m 2^{-l(x)}, \text{ если } (m,x) \in T, \\ 0 \text{ в противном случае.} \end{cases}$$

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. $\triangle$

На самом деле теорема 5.7 дает характеризацию случайных последовательностей не только в терминах монотонной сложности, но и в терминах префиксной сложности. Вторая часть этой теоремы и была доказана для префиксной сложности.

Ввиду важности этого результата мы 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.

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:

$$\begin{aligned} P(\lambda) &= 1; \\ P(x) &= P(x0) + P(x1). \end{aligned} \tag{5.8}$$

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 (5.8). 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 5.1.2.

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* [6] .

Доказательство. Введем 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 $x = x_1 \dots x_n$

[6] Впервые формулировка этой леммы опубликована в [9]. Наиболее точное изложение доказательства имеется в [18].

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

$$\hat{F} = \{(x, y) : I_x \subset \pi_y\}. \tag{5.9}$$

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) если $(x, y) \in \hat{F}$ и $x \subseteq x'$, то $(x', y) \in \hat{F}$;

2) 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 первое 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.

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

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. Случайные по Колмогорову конечные последовательности

А.Н. Колмогоров определил 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.

Пример разбиения – представление $\Xi_n = \cup_{k=0}^{n} C_n^k$, где $C_n^k = \{x \in \Xi_n : \sum_{i=1}^{n} x_i = k\}$ и $x = x_1 \ldots x_n$.

Мы будем рассматривать префиксную модификацию колмогоровской сложности. Колмогоров называл последовательность

$x \in \Xi_n$ m-бернуллиевской, если $\mathrm{KP}(x|k,n) \geqslant \log \binom{n}{k} - m$.[7] Здесь m – произвольное натуральное число.

Функция $k(x|n,\mathcal{D}_n) = \log |D(x)| - \mathrm{KP}(x|n,\mathcal{D}_n)$ 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 $\mathcal{D} = \{D_1, \ldots, D_q\}$, если $k(x|n,\mathcal{D}_n) \leqslant m$, где $D(x)$ – элемент разбиения $\mathcal{D}_n$, содержащий x.

Мы 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 $\sum_{x:l(x)=n} P(x) = 1$.[8] 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 $\mathcal{D}_n$: $\sum_{i=1}^{q} r_n(D_i)$ и $r_n(D_i) \geqslant 0$ при $1 \leqslant i \leqslant q$. В этом случае, инвариантная мера 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 $\mathcal{C}_n = \{C_n^k : k = 0, \ldots, n\}$, если $P(x)$ зависит только от числа единиц k в последовательности x. В этом случае, инвариантная мера представляется в виде $P(x) = \frac{r_n(k)}{\binom{n}{k}}$, где $k = \sum_{i=1}^{n} x_i$,

[7] Напомним, что $|C_n^k| = \binom{n}{k}$. В разделе 3.3 было показано, что $\mathrm{KP}(x|k,n) \leqslant \log \binom{n}{k} + O(1)$.

[8] $\mathcal{R}_+$ – множество всех неотрицательных вещественных чисел.

$r_n(k) = r_n(C_n^k)$ 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 обектом). Рассмотрим условную префиксную сложность $\mathrm{KP}(x|n,P)$, где P – вычислимая вероятностная мера на Ξ_n. Понимаем условие P в $\mathrm{KP}(x|n,P)$ как условие относительно программы для вычисления P.

Дефект случайности последовательности $x \in \Xi_n$ относительно распределения P определяется

$$d(x|n,P) = -\log P(x) - \mathrm{KP}(x|n,P).$$

Пусть $\mathcal{P}_n$ – класс всех вычислимых мер инвариантных относительно разбиения $\mathcal{D}_n$. Определим дефект инвариантности (дефект случайности относительно класса $\mathcal{P}_n$)

$$d(x|\mathcal{P}_n) = \inf_{P\in\mathcal{P}_n} d(x|n,P).$$

Теорема 5.11. *Дефект инвариантности с точностью до аддитивной константы совпадает с дефектом случайности по Колмогорову*

$$d(x|\mathcal{P}_n) = \log|D(x)| - \mathrm{KP}(x|n,D(x)) + O(1),$$

где $D(x)$ – элемент разбиения $\mathcal{D}_n$ содержащий x.

Доказательство. Легко видеть, что дефект инвариантности можно также представить в виде

$$d(x|\mathcal{P}_n) = \log|D(x)| + \inf_{R\in\mathcal{R}_n} (-\log r_n(D(x)) - \mathrm{KP}(x|n,R)). \quad (5.10)$$

Для произвольного $x \in \Xi_n$ рассмотрим вычислимую инвариантную меру $P(z) = \frac{1}{|D(x)|}$ при $z \in D(x)$ и $P(z) = 0$ ддя всех остальных z, т.е. элементу разбиения $D(x)$ приписан вес 1, остальным элементам разбиения приписаны нулевые веса. Обозначим R' соответствующее распределение весов. Тогда $\mathrm{KP}(x|n,R') =$

$\mathrm{KP}(x|n, D(x)) + O(1)$. Данная тривиальная инвариантная мера определяет верхнюю оценку (5.10)

$$d(x|\mathcal{P}_n) \leqslant \log |D(x)| - \mathrm{KP}(x|n, D(x))) + O(1).$$

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

$$Q(x|n, R) = \sum_{D \in \mathcal{D}_n} r_n(D) 2^{-\mathrm{KP}(x|n,D)}.$$

Эта функция является полумерой, так как для любых n и D $\sum_{x \in \Xi_n} 2^{-\mathrm{KP}(x|n,D)} \leqslant 1$, будет

$$\sum_{x \in \Xi_n} Q(x|n, R) = \sum_{D \in \mathcal{D}_n} \sum_{x \in \Xi_n} r_n(D) 2^{-\mathrm{KP}(x|n,D)} \leqslant \sum_{D \in \mathcal{D}_n} r_n(D) \leqslant 1.$$

Функция $Q(x|n, R)$ перечислима снизу при заданных конструктивных объектах n и P (или R).

По свойству априрной перечислимой полумеры $P(x|n, R)$ из раздела 5.1.4 (см. также раздел 5.1.5) будет

$$cP(x|n, R) \geqslant Q(x|n, R) \geqslant r_n(D) 2^{-\mathrm{KP}(x|n,D)}$$

для любого элемента разбиения, 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 $-\mathrm{KP}(x|n, R) \geqslant \log r_n(D(x)) - \mathrm{KP}(x|n, D(x)) - O(1)$ или $-\log r_n(D(x)) - \mathrm{KP}(x|n, R) \geqslant -\mathrm{KP}(x|n, D(x)) - O(1)$. Так как в последнем неравенстве весовое распределение R – произвольное, получаем в (5.10) $d(x|\mathcal{P}_n) \geqslant \log |D(x)| - \mathrm{KP}(x|n, D(x)) - O(1)$. $\triangle$

Для произвольного натурального числа m будем говорить, что конечная последовательность $x \in \Xi_n$ является m-случайной

относительно вероятностного распределения P если $d(x|n,P) \leqslant m$. Аналогичным образом, x является m-случайной относительно класса инвариантных мер $\mathcal{P}_n$, если $d(x|n,\mathcal{P}_n) \leqslant m$. 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 5.11 следует

Следствие 5.4. *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.

5.5. 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 член $\log \mathrm{K}(x)$ нельзя устранить из правой части неравенства.

2. Доказать, что два определения перечислимой снизу (сверху) функции – через перечислимость подграфика (надграфика) и через предел неубывающей (невозрастающей) последовательности функций с рациональными значениями – эквивалентны.

3. Доказать, что если существует алгоритм, который по входу x и произвольному положительному рациональному числу ϵ вычисляет рациональное приближение к $f(x)$ с точностью до ϵ, то функция $f(x)$ перечислима снизу и сверху.

4. Доказать, что имеет место неравенство

$$\mathrm{KP}(x,y) \leqslant \mathrm{KP}(x) + \mathrm{K}(y) + O(1),$$

где $\mathrm{K}(y)$ – простая колмогоровская сложность.

5. Доказать неравенство

$$\mathrm{KP}(\phi(x,y)) \leqslant \mathrm{KP}(x) + \mathrm{KP}(y) + O(1),$$

где $\phi(x,y)$ – произвольная вычислимая функция.

6. Доказать неравенство (5.1), используя результат теоремы 5.3.

7. Используя идею доказательства неравенства (5.2), докажите неравенство

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

8. Доказать неравенство (5.3).

9. Доказать неравенства

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

b) $\mathrm{KP}(x,y,z) \leqslant \mathrm{KP}(x|y) + \mathrm{KP}(y|z) + \mathrm{KP}(z) + O(1)$.

10. Доказать неравенство

$$\begin{aligned}\mathrm{KP}(x) + \mathrm{KP}(y|x) - \log \mathrm{KP}(x) - 2\log\log \mathrm{KP}(x) - O(1) \leqslant \\ \leqslant \mathrm{KP}(x,y) \leqslant \\ \leqslant \mathrm{KP}(x) + \mathrm{KP}(y|x) + O(1).\end{aligned}$$

11. Доказать неравенство

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

12. Пусть F – оптимальная префиксно-корректная функция. Доказать, что число

$$\sum_p \{2^{-l(p)} : F(p) \text{ определено}\}$$

является невычислимым и, более того, случайным относительно равномерной меры. Это число называется числом Чейтина.

В задачах 13–15 рассматриваются величины с аналогичными свойствами.

13. Доказать, что универсальная перечислимая снизу полумера на натуральных числах $P(n)$ не является вычислимой функцией и $\sum_n P(n) < 1$.

14. Доказать, что двоичное разложение действительного числа $\sum_n P(n)$ является случайной по Мартин-Лефу последовательностью.

15. Доказать, что такое же утверждение верно для вещественного числа $\sum_n 2^{-\mathrm{KP}(n)}$.

16. 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.

17. 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:

$$\mathrm{KP}'(x|y) = \mathrm{KP}(x|y) + O(1).$$

18. 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.

19. Доказать, что $\mathrm{KM}(x) \leqslant \mathrm{KM}(y)$ при $x \subseteq y$.

20. Доказать, что $\mathrm{K}(x) \not\leqslant \mathrm{KM}(x) + O(1)$. Привести пример, где нарушается соответствующее неравенство.

21. 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.

22. 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.

23. Привести пример, когда для некоторой вычислимой операции $F(x) = \omega$, где x – конечная последовательность, а ω – бесконечная. Что можно сказать о последовательности ω?

24. Доказать, что вычислимая последовательность случайна по некоторой вычислимой мере.

25. Бернуллиевская мера B_p является вычислимой, если вычислимо число p. Доказать, что верно и обратное утверждение.

26. Пусть I_p – множество всех бесконечных последовательностей, случайных относительно меры B_p. Доказать, что $I_p \cap I_q = \emptyset$, если $p \neq q$.

27. 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).

28. Доказать, что $\sup_n \mathrm{KM}(\omega^n) < \infty$ тогда и только тогда, когда бесконечная последовательность ω является вычислимой.

29. 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:

$$km_f(x) = \min\{l(p) : x \subseteq f(p)\}.$$

а) 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. В частности, $km(x) \neq \mathrm{KM}(x) + O(1)$.

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?

30. Доказать, что

a) $\mathrm{KP}(\omega^n) \leqslant n + \mathrm{KP}(n) + O(1)$.

b) $\mathrm{KP}(\omega^n) = \mathrm{KP}(n) + O(1)$ для любой вычислимой последовательности ω.

31. Проверить утверждение из доказательства леммы 5.8: в любом отрезке π_x строго содержится отрезок с двоичными концами длины не менее $\frac{1}{4}|\pi_x|$.

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: $x \subseteq y$.

Пусть $\Xi = \{0,1\}^*$ – 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 $W_i^0 = 0$.

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. $\triangle$

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. $\triangle$

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

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

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

$$Q(x) = L(\cup_y \{\Gamma_y : x \subseteq A(y)\}). \tag{6.1}$$

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. $\triangle$

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 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 сохраним это свойство на всех последовательностях, 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.

Мы можем утверждать, что функции Q_m равномерно 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. $\triangle$

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

$$P(x) = L\{\omega : x \subseteq F(\omega)\}. \tag{6.3}$$

Легко видеть, что функция $P(x)$ является перечислимой снизу полумерой.

Верно и обратное: для любой перечислимой снизу полумеры P можно построить такую вычислимую операцию F, что для нее выполнено (6.3) (доказательство в виде задачи).

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 (6.3) 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:

$$P(x) = L\{\omega : F(\omega) \in A\}.$$

Соотношение (6.3) устанавливает взаимно однозначное соответствие между вероятностными машинами и перечислимыми снизу полумерами.

В частности, универсальная полумера M также допускает представление (6.3):

$$M(x) = L\{\omega : x \subseteq F_M(\omega)\}$$

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 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 можно 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 $x_1, \ldots, x_m$, для которого m 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. $\triangle$

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 $Q(\cdot|x)$.

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 забирает 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. [16]). Эта теорема позволяет казино с вероятностью единица не разориться.

Легко видеть, что функция $P(x) = \mathcal{P}(x)Q(x)$ удовлетворяет условиям определения полумеры. Верно обратное утверждение: для любой полумеры P функция

$$\mathcal{P}(x) = P(x)/O(x)$$

является Q-супермартингалом.

Супермартингал $\mathcal{P}$ является перечислимым снизу, если множество

$$\{(r, x) : r < \mathcal{P}(x)\}$$

перечислимо.

Как и для перечислимых снизу полумер имеет место теорема о существовании максимального с точностью до мультипликативной константы перечислимого снизу супермартингала.

Теорема 6.4. *Пусть задана вычислимая мера Q. Существует перечислимый снизу Q-супермартингал $\mathcal{M}$ такой, что для любого перечислимого снизу Q-супермартингала $\mathcal{P}$ существует константа c, для которой*

$$c\mathcal{M}(x) \geqslant \mathcal{P}(x)$$

для всех x.

Доказательство этой теоремы аналогично доказательству теоремы 6.1.

Легко видеть, что для любой вычислимой меры Q выполнено

$$\mathcal{M}(x) = \frac{M(x)}{Q(x)}O(1),$$

где $\mathcal{M}$ – максимальный перечислимый снизу Q-супермартингал, M – универсальная полумера.

Универсальная полумера M имеет преимущество перед супермартингалом $\mathcal{M}$, так как она не зависит от какой-либо меры.

Имеет место аналог теоремы 6.2.

Теорема 6.5. *Бесконечная последовательность ω случайна по вычислимой мере Q тогда и только тогда, когда*

$$\sup_n \mathcal{M}(\omega^n) < \infty.$$

6.2. Универсальный предиктор Соломонова

В этом разделе мы рассмотрим теорию Р. Соломонова об универсальных предсказаниях (см. его работы [37], [38]).

Будет доказано, что универсальная полумера в среднем прогнозирует биты бесконечной последовательности не хуже чем произвольная вычислимая мера, которая генерирует эту последовательность.

Рассмотрим задачу прогнозирования следующего бита последовательности, поступающей в режиме онлайн.

Допустим, что уже известны первые $n-1$ битов

$$\omega_1,\ \ldots,\ \omega_{n-1}$$

последовательности, генерируемой некоторой неизвестной нам вычислимой мерой P. Необходимо предсказать вероятность того, что следующий бит ω_n равен 0 или 1.

Пусть некоторый источник генерирует бит $\omega_n = 0$ с неизвестной нам вероятностью [1] :

$$P(0|\omega^{n-1}) = \frac{P(\omega^{n-1}0)}{P(\omega^{n-1})}.$$

[1] Для простоты мы предполагаем, что $P(x) > 0$ для всех x.

Соответственно вероятность события $\omega_n = 1$ равна

$$P(1|\omega^{n-1}) = \frac{P(\omega^{n-1}1)}{P(\omega^{n-1})}.$$

Предсказателю эти вероятности не известны. Наша задача – построить некоторый универсальный метод предсказания, который вычислял бы эти вероятности или некоторые приближения к ним на основе только последовательности $\omega_1 \ldots \omega_{n-1}$ без использования распределения P источника.

Универсальный предсказатель 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:

$$M(b|x) = \frac{M(xb)}{M(x)}. \tag{6.4}$$

Функция (6.4) не является мерой (см. упр. 7 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 [2] .

Следующая теорема показывает силу прогностических возможностей функции (6.4).

Теорема 6.6. *Для любой вычислимой меры P и для любого бита $b \in \{0,1\}$ выполнено*

$$\sum_{n=1}^{\infty} \sum_{l(x)=n} P(x)(M(b|x) - P(b|x))^2 < \infty. \tag{6.5}$$

[2] Правда, можно рассматривать различные вычислимые приближения к ней эвристического типа.

Выражение (6.5) 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 (6.4). 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. Нам будет удобно обозначать посредством P_n и Q_n их ограничения на множестве $\{0,1\}^n$. Расстоянием Кульбака–Лейблера между P_n и Q_n называется величина

$$D(P_n|Q_n) = \sum_{l(x)=n} P_n(x) \ln \frac{P_n(x)}{Q_n(x)}.$$

Из выпуклости логарифма следует, что всегда $D(P_n|Q_n) \geqslant 0$ и $D(P_n|Q_n) = 0$ тогда и только тогда, когда $P_n = Q_n$.

Выражение $D(P_n|Q_n)$ 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) *Для мер P_1 и Q_1 на двухэлементном множестве $\{0,1\}$ выполнено неравенство*

$$D(P_1|Q_1) \geqslant 2(P_1(0) - Q_1(0))^2. \tag{6.6}$$

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*

$$D(P_1|Q_1) \geqslant D(P_1|Q'_1). \tag{6.7}$$

Доказательство этой леммы предлагается в виде задачи 9 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

$$D(P_1|M_1) \geqslant D(P_1|M'_1).$$

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*

$$D(P_{n+1}|Q_{n+1}) = D(P_n|Q_n) + \sum_{l(x)=n} P_n(x) D(P_{n+1}(\cdot|x)|Q_{n+1}(\cdot|x)).$$

Доказательство. Раскроем величину $D(P_{n+1}|Q_{n+1})$:

$$D(P_{n+1}|Q_{n+1}) = \sum_{l(z)=n+1} P_{n+1}(z) \ln \frac{P_{n+1}(z)}{Q_{n+1}(z)} =$$

$$= \sum_{l(x)=n, b\in\{0,1\}} P_{n+1}(xb) \ln \frac{P_{n+1}(xb)}{Q_{n+1}(xb)} =$$

$$= \sum_{l(x)=n, b\in\{0,1\}} P_n(x) P_{n+1}(b|x) \ln \frac{P_n(x) P_{n+1}(b|x)}{Q_n(x) Q_{n+1}(b|x)} =$$

$$= \sum_{l(x)=n} P_n(x) \ln \frac{P_n(x)}{Q_n(x)} \sum_{b\in\{0,1\}} P_{n+1}(b|x) +$$

$$+ \sum_{l(x)=n} P_n(x) \sum_{b\in\{0,1\}} P_{n+1}(b|x) \ln \frac{P_{n+1}(b|x)}{Q_{n+1}(b|x)} =$$

$$= D(P_n|Q_n) + \sum_{l(x)=n} P_n(x) D(P_{n+1}(\cdot|x)|Q_{n+1}(\cdot|x)).$$

Здесь мы использовали равенства $\sum\limits_{b\in\{0,1\}} P_{n+1}(b|x) = 1$ и

$$D(P_n|Q_n) = \sum_{l(x)=n} P_n(x) \ln \frac{P_n(x)}{Q_n(x)}.$$

Лемма доказана.

Раскрывая сумму в лемме 6.2, получаем следующее следствие.

Следствие 6.1. *Для любого n выполнено*

$$D(P_n|Q_n) = \sum_{i=1}^{n} \sum_{l(x)=i} P_i(x) D(P_i(\cdot|x)|Q_i(\cdot|x)). \tag{6.8}$$

Пусть теперь P – вычислимая мера из условия теоремы, M' – дополненная до меры универсальная полумера. Применяем лем-

му 6.3 к каждому слагаемому суммы (6.8), где $Q = M$, получаем

$$D(P_n|M_n) = \sum_{i=1}^{n} \sum_{l(x)=i} P_i(x) D(P_i(\cdot|x)|M_i(\cdot|x)) \geqslant$$
$$\geqslant \sum_{i=1}^{n} \sum_{l(x)=i} P_i(x) D(P_i(\cdot|x)|M_i'(\cdot|x)) \geqslant$$
$$\geqslant 2 \sum_{i=1}^{n} \sum_{l(x)=i} P(x)(M(0|x) - P(0|x))^2. \qquad (6.9)$$

По основному свойству априорной полумеры $cM(x) \geqslant P(x)$ для всех x, где c – константа, зависящая от меры P. Отсюда получаем

$$D(P_n|M_n) = \sum_{l(x)=n} P(x) \ln \frac{P(x)}{M(x)} \leqslant \ln c \sum_{l(x)=n} P(x) \leqslant \ln c$$

для всех n. Соединяем это неравенство с (6.9) и получаем необходимую нам оценку (6.5):

$$\sum_{i=1}^{\infty} \sum_{l(x)=i} P(x)(M(0|x) - P(0|x))^2 \leqslant \frac{1}{2} \sup_n D(P_n|M_n) \leqslant \frac{1}{2} \ln c.$$

Теорема доказана. $\triangle$

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 (6.3).

5. Доказать, что для любой вычислимой последовательности ω существует такая константа $c > 0$, что $M(\omega^n) > c$ для всех n.

6. Доказать теоремы 6.4 и 6.5.

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 (6.4) не является перечислимой сверху или снизу.

9. Доказать утверждения (6.6) и (6.7) леммы 6.2.

10. 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.

11. Доказать, что универсальная перечислимая снизу полумера на дереве двоичных последовательностей $M(x)$ не является вычислимой функцией и $0 < \inf_n \sum_{l(x)=n} M(x) < 1$.

12. Доказать, что функция

$$\bar{M}(x) = \inf_{n \geqslant l(x)} \sum_{x \subseteq y, l(y)=n} M(y)$$

является мерой. Для нее $0 < \bar{M}(\lambda) < 1$.

Доказать, что мера $\bar{M}$ не является ни вычислимой, ни перечислимой снизу или сверху.

Кроме этого, мера $\bar{M}$ является максимальной мерой такой, что $\bar{M}(x) \leqslant M(x)$ для всех x.

13. Можно ли построить универсальный предиктор на основе монотонной сложности $\mathrm{KM}(x)$ так, чтобы для него была верна теорема 6.6?

14. Доказать, что для любой меры Q для Q-почти всех бесконечных последовательнстей ω существует предел $\lim\limits_{n\to\infty} \frac{M(\omega^n)}{Q(\omega^n)}$.

Существует ли этот предел для любой бесконечной последовательности ω случайной относительно вычислимй меры Q?

Часть III

Приложения

Глава 7

Алгоритмические вопросы 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 усиленно-

го 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 7.1.

Как мы покажем в разделе 7.2, 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.

7.1. 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. [16]) прямо транслируется в конструктивную форму, полезно рассмотреть новое доказательство этой теоремы в духе идей колмогоровского алгоритмического подхода к теории вероятностей. Это доказательство было предложено В. Вовком [2]. Первая часть этого доказательства (неравенства $\leqslant$) также представлена в монографии [18]. 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 последовательности 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 $r = p_n$

$$\mathrm{KM}(\omega^n) \leqslant -\log B_{p_n}(\omega^n) + \mathrm{KP}(p_n) + O(1).$$

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). $\triangle$

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.

7.2. 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 доказана в разделе 7.2.3.

Тем не менее конструктивный анализ этой теоремы, приведенный в монографии Бишопа [25], позволяет доказать эргодическую теорему для случайных по Мартин-Лефу последовательностей. Мы докажем этот конструктивный вариант эргодической теоремы в разделе 7.2.5. Для этого нам придется ввести новый тип тестов случайности – интегральные тесты, с помощью которых можно дать новую эквивалентную формулировку случайности по Мартин-Лефу.

7.2.1. 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 $T^k\omega = T(T^{k-1}\omega)$, где $T^0\omega = \omega$.

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

$$P\{\omega : \omega_i = x_1, \ldots \ , \omega_{i+k-1} = x_k\} = P\{\omega : \omega_1 = x_1, \ldots \ , \omega_k = x_k\}$$

для любой конечной последовательности $x = x_1 \ldots x_k$ 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.

[1] То есть $P((T^{-1}(A) \setminus A) \cup (A \setminus T^{-1}(A))) = 0$.

Теорема 7.2. *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 $\int \tilde{f}(\omega)dP = P(\Gamma_1) = P\{\omega : \omega_1 = 1\}$.

7.2.2. 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

[2] Функция $\tilde{f}$ инвариантная относительно T, если $\tilde{f}(T\omega) = \tilde{f}(\omega)$.

$T^n\omega \notin E$ для всех $n > 0$, имеет меру 0. Эквивалентно, для почти всех $\omega \in E$ $T^n\omega \in E$ 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 7.3. *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 7.3 будет прямым следствием следующего утверждения, которое было впервые доказано Кучерой [31].

Предложение 7.1. *Пусть U – эффективно открытое множество равномерной меры $L(U) < 1$. Тогда для любой бесконечной последовательности ω, случайной по мере L, и произвольного 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. $\triangle$

Для доказательства теоремы 7.3 надо взять в предложении 7.1 $U = \Omega \setminus E$. Так как $L(E) > 0$, выполнено неравенство $L(U) < 1$. По предложению 7.1 $T^n\omega \notin U$ для бесконечно многих n. $\triangle$

[3] Это утверждение было предметом задачи 3(d) из раздела 4.3.

7.2.3. 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 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 ин-

тервалов 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 7.2. *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. $\triangle$

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

$$P\{\omega : |f_n(\omega) - f(\omega)| > \delta\} < \epsilon \tag{7.8}$$

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 (7.8) 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 (7.8).

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

$$P\{\omega : |f_n(\omega) - f_{n'}(\omega)| > \delta\} < \epsilon \tag{7.9}$$

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. [16]).

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:

$$P\left\{\omega : \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| > \epsilon\right\} \leqslant 2e^{-2n\epsilon^2},$$

где $S_n(\omega) = \sum_{i=1}^{n} \omega_i$. В данном случае можно определить

$$m(\epsilon, \delta) = \lfloor \frac{1}{2\epsilon^2} \ln \frac{2}{\delta} \rfloor + 1.$$

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 7.4. *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. 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 $m(\delta, \epsilon)$. Напомним, что в данном случае мы используем некоторое отождествление всех положительных рациональных чисел и всех натуральных чисел.

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, 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 7.3 легко показать, что вероятностная мера P_i, порожденная заданными начальными и переходными вероятностями, является стационарной. Кроме того, она является вычислимой.

Согласно теории из монографии [16] 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,

$$P_i\{\omega : |S_n(\omega) - \frac{1}{2}| < 0.01\} \to 1 \tag{7.10}$$

при $n \to \infty$.

Пусть число $k(i)$ равно номеру позиции последнего нуля в двоичном представлении числа α_i, после которого в этом представлении стоят единицы. Тогда $\alpha_i = 2^{-k(i)}$.

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 $k(i)$ [4] . 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 (7.10) следует, что найдется достаточно большое $n > m(\frac{1}{4}, 2^{-(i+1)})$, для которого

$$P_i\{\omega : |S_{k(i)}(\omega) - S_n(\omega)| > \frac{1}{4}\} > \frac{1}{2}.$$

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 (7.9) для регулятора сходимости.

Полученное противоречие доказывает теорему. $\triangle$

Поскольку из алгоритмически эффективной сходимости почти всюду следует алгоритмически эффективная сходимости по вероятности, получаем следующее следствие из теоремы 7.4.

[4] Без потери общности можно предположить, что все шаги s, на которых впервые определилось какое-либо значение универсальной функции, больше некоторого фиксированного значения s_0.

Следствие 7.2. *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 $n \to \infty$.

7.2.4. 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 [29].

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 7.5. *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 7.1. *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. $\triangle$

Фиксируем один из интегральных тестов, удовлетворяющих теореме 7.5, 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 7.6. *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. $\triangle$

7.2.5. 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 7.7. *Пусть P – произвольная вычислимая мера на Ω и f – произвольная вычислимая интегрируемая функция типа $\Omega \to \mathcal{R}$ (которая называется наблюдаемой). Тогда для любого сохраняющего меру P вычислимого преобразования T для любой случайной по Мартин-Лефу последовательности ω выполнено*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = \tilde{f}(\omega), \tag{7.11}$$

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 $a(u,\omega) < b(v,\omega)$.

Говорим, что 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 7.2. *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

$$d = \{u_1, v_1, s_1, t_1, \ \ldots, \ s_m, t_m\}. \tag{7.12}$$

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

$$d = \{s_1, t_1, \ \ldots, \ s_{i-1}, v_{i-1}, u_i, t_{i-1}, \ \ldots, \ s_m, t_m\}. \tag{7.13}$$

Если $u_i \geqslant t_{i-1}$, то положим при $i \leqslant m$

$$d = \{s_1, t_1, \dots\ , s_{i-1}, t_{i-1}, u_i, v_i, s_i, t_i, \dots\ , s_m, t_m\}, \qquad (7.14)$$

и полагаем при $i = m + 1$

$$d = \{s_1, t_1,\ \dots,\ s_m, t_m, s_{m+1}, t_{m+1}\}. \qquad (7.15)$$

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 (7.12), (7.14) и (7.15)

$$S(\omega, d) = S(\omega, q) + b(v_i, \omega) - a(u_i, \omega).$$

При определении (7.13)

$$S(\omega, d) = S(\omega, q) + b(v_{i-1}, \omega) - a(u_i, \omega).$$

По определению последовательности $\{u_1, v_1,\ \dots,\ u_N, v_N\}$ прибавленные члены положительны. Значит, в обоих случаях кумулятивные суммы увеличиваются:

$$S(\omega, d) > S(\omega, q).$$

Лемма доказана. $\triangle$

Пусть $d = \{s_1, t_1,\ \dots,\ s_m, t_m\}$ – 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:

$$a(s_i, \omega) = a(s_i - 1, T\omega) + f(\omega) - \alpha,$$
$$b(t_i, \omega) = b(t_i - 1, T\omega) + f(\omega) - \beta.$$

Отсюда и из определения кумулятивной суммы получаем

$$S(\omega, d) = S(T\omega, d') + a - (\beta - \alpha)m_d, \qquad (7.16)$$

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 (7.16) $a = 0$, 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 (7.16) следует, что

$$S(\omega, d) \leqslant \lambda_n(T\omega) + (f(\omega) - \alpha)^+ - (\beta - \alpha)m_d, \tag{7.17}$$

где использовано обозначение $h^+ = \max\{h, 0\}$.

По лемме 7.2 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 (7.17) имеем

$$\begin{aligned} S(\omega, q) < S(\omega, d) \leqslant \\ \leqslant \lambda_n(T\omega) + (f(\omega) - \alpha)^+ - (\beta - \alpha)\sigma_n(\omega|\alpha, \beta), \end{aligned} \tag{7.18}$$

Берем в (7.18) максимум по q и получаем

$$\lambda_n(\omega) \leqslant \lambda_n(T\omega) + (f(\omega) - \alpha)^+ - (\beta - \alpha)\sigma_n(\omega|\alpha, \beta).$$

Следовательно,

$$(\beta - \alpha)\sigma_n(\omega|\alpha, \beta) \leqslant (f(\omega) - \alpha)^+ + \lambda_n(T\omega) - \lambda_n(\omega). \tag{7.19}$$

Интегрируя неравенство (7.19), получим

$$\int (\beta - \alpha)\sigma_n(\omega|\alpha, \beta) dP \leqslant \int (f(\omega) - \alpha)^+ dP. \tag{7.20}$$

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 (7.20) 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 (7.11). 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 7.3. *$\tilde{f}(\omega) = \int f(\omega)dP$ 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 (7.11) 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 7.7. $\triangle$

7.3. Задачи и упражнения

1. Привести полное доказательство леммы 7.1.

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.

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. 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.

[4] 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.

[5] Вьюгин В.В. Элементы математической теории машинного обучения: учебное пособие. М.: ГОУ ВПО «Московский физико-технический институт» (государственный университет) : ИППИ РАН, 2010. – 231 с.

[6] 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.

[7] Колмогоров А.Н. Три подхода к определению понятия "количество информации" // Проблемы передачи информациию. 1965. Т. 1 (1). С. 3–11.

[8] Кричевский Р.Е. Связь между избыточностью кодирования и достоверностью сведений об источнике // Проблемы передачи информации. 1968. Т. 4 (3). С. 48–57.

[9] Левин Л.А. О понятии случайной последовательности // Доклады АН СССР. 1973. Т. 212 (3), С. 548–550.

[10] Левин Л.А. Законы сохранения (невозрастания) информации и вопросы обоснования теории информации // Проблемы передачи информации. 1974. Т. 10 (3). С. 30–35.

[11] Левин Л.А. О различных мерах сложности конечных объектов (аксиоматическое описание) // Доклады АН СССР. 1976. Т. 227 (4). С. 804–807.

[12] Левин Л.А. Об одном конкретном способе задания сложностных мер // Доклады АН СССР. 1977. Т. 234 (3). С. 536–539.

[13] Роджерс Х. Теория рекурсивных функций и эффективная вычислимость. М.: Мир., 1972.

[14] Рябко Б.Я. Сжатие данных с помощью стопки книг // Проблемы передачи информации. 1980. Т. 16 (4). С. 16–21.

[15] Рябко Б.Я. Дважды универсальное кодирование // Проблемы передачи информации. 1984. Т. 20 (3). С. 24–28.

[16] Ширяев А.Н. Вероятность. – М.: МЦНМО, 2007. 968 с.

[17] Успенский В.А., Семенов А.Л. Теория алгоритмов: основные открытия и приложения. – М.: Наука. Гл. ред. физ.-мат. лит., 1987. (Библиотечка программиста).

[18] 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

[19] 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.

[20] Фитингоф Б.М. Оптимальное кодирование при неизвестной и меняющейся статистике сообщений // Проблемы передачи информации. 1966. Т. 2 (2). С. 3–11.

[21] Фитингоф Б.М. Сжатие дискретной информации // Проблемы передачи информации. 1967. Т. 3 (3). С. 28–36.

[22] 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.).

[23] Штарьков Ю.М. Универсальное последовательное кодирование отдельных сообщений // Проблемы передачи информации 1987. Т. 23 (3). С. 3–17.

[24] Ю. М. Штарьков Ю.М. Универсальное кодирование. Теория и алгоритмы. 2013 ISBN: 978-5-9221-1517-9

[25] Bishop E. Foundation of Constructive Analysis. New York: McGraw-Hill, 1967.

[26] Chaitin G. Information-theoretical limitations of formal systems // Journal of the ACM. 1974. V. 21. P. 403–424.

[27] Chaitin G. A theory of program size formally identical to information theory // J. Assoc. Comput. Mach., 1975, V.22, P.329–340.

[28] Cover T. M.. Thomas J. A. Elements of Information Theory. New York: Wiley, 1991.

[29] Gács P. Exact expressions for some randomness tests // Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik. 1980. V. 26. P. 385–394.

[30] 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

[31] Kucera A. Measure, Π_1^0 classes, and complete extensions of PA // Lecture Notes in Mathematics. 1985. V. 1141. P. 245–259.

[32] Li M., Vitányi P. An Introduction to Kolmogorov Complexity and Its Applications, 2nd ed. New York: Springer–Verlag, 1997.

[33] Lugosi G., Cesa-Bianchi N. Prediction, Learning and Games. Cambridge University Press, New York, 2006.

[34] Andrew R. Barron, Jorma Rissanen, Bin Yu The Minimum Description Length Principle in Coding and Modeling // IEEE Transactions on Information Theory. 1998. V. 44 (6). P. 2743–2760.

[35] Shafer G., Vovk V. Probability and Finance. It's Only a Game! New York: Wiley, 2001.

[36] Schnorr C.P. Process complexity and effective random tests // Journal of Computer and System Sciences. 1973. V. 3 (4). P. 376–378.

[37] Solomonoff R.J. A formal theory of inductive inference I, II // Information and Control. 1964. V. 7. P. 1–22. P. 224–254.

[38] Solomonoff R.J. Complexity-based induction systems: Comparisons and convergence theorems // IEEE Transactions on Information Theory. 1978. IT-24. P. 422–432.

[39] V'yugin V.V. Algorithmic complexity and stochastic properties of finite binary sequences // The Computer Journal, 1999, v.42, N4, p.294–317.

[40] Ziv, J., Lempel, A. A Universal Algorithm for Sequential Data Compression // IEEE Transactions on Information Theory. 1977. V. 23 (3). P. 337–343.

[41] Ziv, J. Lempel, A. Compression of Individual Sequences via Variable-Rate Coding // IEEE Transactions on Information Theory. 1978. V. 24 (5): P. 530–536.