Отдельный оттиск

УСПЕХИ МАТЕМАТИЧЕСКИХ НАУК

ТОМ
XXV
ВЫПУСК
6(156)

1970

1970 г. ноябрь — декабрь т. XXV вып. 6 (1956)
УСПЕХИ МАТЕМАТИЧЕСКИХ НАУК

УДК 519.24+517.11+519.9

СЛОЖНОСТЬ КОНЕЧНЫХ ОБЪЕКТОВ И ОБОСНОВАНИЕ ПОНЯТИЙ ИНФОРМАЦИИ И СЛУЧАЙНОСТИ С ПОМОЩЬЮ ТЕОРИИ АЛГОРИТМОВ

А. К. Звонкин и Л. А. Левин

В 1964 г. А. Н. Колмогоров ввел понятие сложности конечного объекта (например, слова в некотором алфавите). Сложность он определял как минимальное число двоичных знаков, содержащих всю информацию о задаваемом объекте, достаточную для его восстановления (декодирования). Это определение существенно зависит от метода декодирования, однако с помощью общей теории алгоритмов А. Н. Колмогорову удалось дать инвариантное (универсальное) определение сложности. Близкие понятия рассматривались Р. Дж. Соломоновым (США) и А. А. Марковым. На базе понятия сложности А. Н. Колмогоров дал определение количества информации в конечных объектах и понятия случайной последовательности (уточненное потом в работах П. Мартин-Лёфа). Впоследствии этот круг вопросов быстро развивался. В частности, интересное развитие получили идеи А. А. Маркова о применении понятия сложности для изучения количественных вопросов теории алгоритмов. Настоящая статья представляет собой обзор основных результатов, связанных со всем изложенным.

СОДЕРЖАНИЕ

Предварительные замечания

При написании статьи мы, кроме цитированной литературы, существенно использовали материалы лекций А. Н. Колмогорова, спецкурса Н. В. Петри и М. И. Кановича, а также семинара В. А. Душского и Л. А. Левина. Мы горячо благодарны Андрею Николаевичу Колмогорову, который оказал нам большую помощь, редактируя все промежуточные варианты текста статьи; без его постоянной поддержки статья вообще не могла бы быть написана. Весьма ценным для нас был постоянный контакт и обсуждение результатов с М. И. Кановичем и Н. В. Петри, за что мы им крайне признательны. Мы

очень благодарны А. Б. Сосинскому, прочитавшему всю рукопись и сделавшему много ценных замечаний. Мы хотим также поблагодарить В. Н. Агафонова, Я. М. Барздиня, А. Н. Колодия, П. Мартин-Лёфа, Л. Б. Медведовского, В. А. Успенского. Дж. Т. Шварца и всех участников семинара А. А. Маркова за ценное обсуждение.

1. Некоторые определения и обозначения. Мы будем рассматривать *слова* в алфавите $\{0, 1\}$, т. е. конечные последовательности нулей и единиц. Установим взаимно однозначное соответствие между словами и *натуральными числами*:

$$\begin{aligned}
\Lambda &\leftrightarrow 0\\
0 &\leftrightarrow 1\\
1 &\leftrightarrow 2\\
00 &\leftrightarrow 3\\
01 &\leftrightarrow 4\\
10 &\leftrightarrow 5\\
11 &\leftrightarrow 6\\
000 &\leftrightarrow 7\\
001 &\leftrightarrow 8\\
&\ldots\ldots
\end{aligned}$$

(Λ — пустое слово), и в дальнейшем не будем различать эти объекты, употребляя произвольно любой из терминов «слово» или «число». Обозначать их мы будем, как правило, малыми латинскими буквами, множество всех слов-чисел будем обозначать S.

Если к слову x справа приписать слово y, получится слово, которое мы будем обозначать xy. Нам потребуется также уметь записывать одним словом упорядоченную пару слов (x, y). Для того чтобы не вводить специальных разделительных знаков (вроде запятой), условимся, что если $x = x_1x_2 \ldots x_n$ (x_i — 0 или 1), то

$$\bar{x} = x_1x_1x_2x_2 \ldots x_nx_n01. \tag{0.1}$$

Тогда по слову $\bar{x}y$ можно однозначно восстановить и x, и y. Обозначим $\pi_1(z)$ и $\pi_2(z)$ функции такие, что $\pi_1(\bar{x}y) = x$, $\pi_2(\bar{x}y) = y$; если слово z не представимо в виде $\bar{x}y$, то $\pi_1(z) = \Lambda$, $\pi_2(z) = \Lambda$ [1]).

Длиной $l(x)$ слова x будем называть количество знаков в нем; $l(\Lambda) = 0$. Очевидно,

$$l(xy) = l(x) + l(y), \tag{0.2}$$

$$l(\bar{x}) = 2l(x) + 2. \tag{0.3}$$

Обозначим $d(A)$ *количество элементов в множестве* A. Очевидно,

$$d\{x:\ l(x) = n\} = 2^n, \tag{0.4}$$

$$d\{x:\ l(x) \leqslant n\} = 2^{n+1} - 1. \tag{0.5}$$

---

[1]) Можно было бы устроить более стандартную нумерацию пар (x, y), однако для нас важно, чтобы выполнялось свойство (0.11) (см. ниже)

Объектом нашего рассмотрения будет также пространство Ω *бесконечных двоичных последовательностей* (их мы будем обозначать малыми греческими буквами). $\Omega^* = \Omega \cup S$ — множество всех *конечных и бесконечных последовательностей*. Пусть $\omega \in \Omega^*$; тогда будем называть *n-фрагментом* ω и обозначать $(\omega)_n$ слово, состоящее из первых n знаков ω (при этом если ω — слово, и $l(\omega) \leqslant n$, то, по определению, $(\omega)_n = \omega$). Последовательность $\omega \in \Omega$ будем называть *характеристической* для множества натуральных чисел $A = \{n_1, n_2, \ldots\}$, не содержащего 0, если в этой последовательности n_1-я, n_2-я, ... цифра — единицы, а все остальные цифры — нули. Множество A, для которого ω — характеристическая последовательность, будем обозначать также S_ω.

Обозначим Γ_x множество всех последовательностей (конечных и бесконечных или только бесконечных, в зависимости от того, рассматриваем мы пространство Ω^* или Ω; в каждом конкретном случае это будет ясно из контекста), начинающихся со слова x, т. е.

(0.6) $\quad \Gamma_x = \{\omega\colon (\omega)_{l(x)} = x\}$.

Будем обозначать $x \subset y$, если $\Gamma_x \supseteq \Gamma_y$ (т. е. слово x есть начало слова y). Отношение $\subset$ частично упорядочивает множество S (рис. 1).

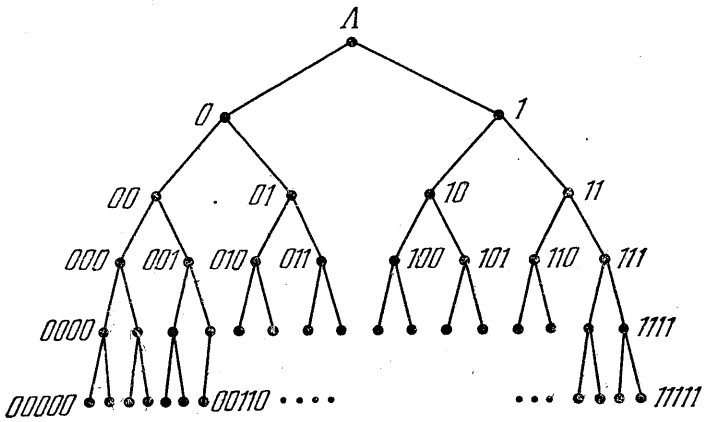

Рис. 1.

Функции, определенные на декартовом произведении $S^n = S \times S \times \ldots \times S$ (n раз), будем (за исключением, может быть, стандартных функций) обозначать большими латинскими буквами, иногда ставя вверху индекс n (обозначающий число переменных): $F^n = F^n(x_1, \ldots, x_n)$. Будем всегда стандартным образом заменять фразу: для любых допустимых значений переменных $y_1, \ldots, y_m$ найдется константа C такая, что для всех допустимых значений $x_1, \ldots, x_n$

(0.7) $\quad F^{n+m}(x_1, \ldots, x_n; y_1, \ldots, y_m) \leqslant G^{n+m}(x_1, \ldots, x_n; y_1, \ldots, y_m) + C$,

на более короткую фразу (использующую новое обозначение):

(0.8) $\quad F^{n+m}(x_1, \ldots, x_n; y_1, \ldots, y_m) \lesssim G^{n+m}(x_1, \ldots, x_n; y_1, \ldots, y_m)$

($y_1, \ldots, y_m$ входят как параметры).

Аналогично определяется отношение $\gtrsim$; $F \asymp G$ тогда и только тогда, когда $F \lesssim G$ и $G \lesssim F$. Очевидно, отношения $\lesssim$, $\gtrsim$ и $\asymp$ транзитивны. Очевидно также,

(0.9) $\quad l(x) \gtrsim \log_2 x$ для $x > 0$,

(0.10) $\quad l(\bar{x}) \gtrsim 2l(x)$,

(0.11) $\quad l(\bar{x}y) \gtrsim l(y)$ (x входит как параметр),

и т. д.

2. Необходимые сведения из теории алгоритмов. Приведем некоторые необходимые нам определения и теоремы из теории алгоритмов. Большинство из приведенных фактов доказывается в любом руководстве по теории алгоритмов (см., например, [1]—[4]), доказательство остальных не представит труда для читателя, знакомого с одним из таких руководств.

Пусть функции C^1, O^n, I_m^n принимают, по определению, следующие значения: $C^1(x)=x+1$, $O^n(x_1, \ldots, x_n)=0$, $I_m^n(x_1, \ldots, x_n)=x_m$. Говорят, что $(n+1)$-местная функция F возникает из n-местной функции G и $(n+2)$-местной функции H *примитивной рекурсией*, если для всех натуральных значений $x_1, \ldots, x_n, y$ имеем

$$F(x_1, \ldots, x_n, 0)=G(x_1, \ldots, x_n),$$

$$F(x_1, \ldots, x_n, y+1)=H(x_1, \ldots, x_n, y, F(x_1, \ldots, x_n, y)).$$

Обозначим через

$$\mu_y(F(x_1, \ldots, x_{n-1}, y)=x_n) \tag{0.12}$$

наименьшее значение a, для которого

$$F(x_1, \ldots, x_{n-1}, a)=x_n. \tag{0.13}$$

При этом будем считать, что значение (0.12) не определено в следующих случаях:

а) значения $F(x_1, \ldots, x_{n-1}, y)$ определены для всех $y<a$, но отличны от x_n, а значение $F(x_1, \ldots, x_{n-1}, a)$ не определено $(a=0, 1, 2, \ldots)$;

б) значения $F(x_1, \ldots, x_{n-1}, y)$ определены для всех $y=0, 1, 2, \ldots$ и отличны от x_n.

Значение выражения (0.12) при заданной функции F зависит от значений $x_1, \ldots, x_{n-1}, x_n$, т. е. является функцией от этих переменных. Говорят, что эта функция получена из функции F при помощи *операции минимизации*.

Определение 0.1. Функция F называется *частично рекурсивной*, если она может быть получена из функций C^1, O^n, I_m^n конечным числом операций *подстановки* (т. е. суперпозиции), примитивной рекурсии и минимизации. Всюду определенная частично рекурсивная функция называется *общерекурсивной*. Свойство n-ок чисел $\Pi^n(a_1, \ldots, a_n)$ называется *частично рекурсивным* (*общерекурсивным*) *предикатом*, если существует частично рекурсивная (общерекурсивная) функция, равная 0 на всех n-ках, удовлетворяющих этому свойству, и только на них.

Легко проверить, что функции $l(x)$, $\pi_1(z)$, $\pi_2(z)$, $F(x)=\bar{x}$, $G(x, y)=\overline{xy}$ общерекурсивны.

В настоящее время общепринятой является следующая естественнонаучная гипотеза:

Тезис Черча. *Класс алгоритмически вычислимых (в инстуитивно ясном смысле) числовых функций совпадает с классом всех частично рекурсивных фунций.*

В дальнейшем изложении мы неоднократно будем, приводя алгоритм, вычисляющий некоторую функцию, предполагать ее частичную рекурсив-

ность без доказательства, т. е. не выписывая (из-за громоздкости) построение, требуемое определением 0.1. Трудолюбивый читатель, не желающий в каждом таком случае принимать на веру тезис Черча, всегда сможет выписать такое построение самостоятельно.

Замечание 0.1. Легко видеть, что частично рекурсивные функции, строящиеся без употребления операции минимизации (такие функции называются *примитивно рекурсивными*), являются всюду определенными. Только операция минимизации может приводить к не всюду определенным функциям, так как процесс вычисления результата минимизации (состоящий в последовательной проверке справедливости равенства (0.13) для $a = 0, 1, 2, \ldots$) может никогда не кончиться. Будем говорить, что значение частично рекурсивной функции F^n на данном наборе $(x_1, \ldots, x_n)$ вычислено *не более чем за t шагов* (*операций*), если все процессы вычисления результатов минимизаций, входящих в построение функции F^n, закончились на значениях соответствующих параметров a, не превышающих t. Мы будем часто употреблять понятие количества шагов, совершенных алгоритмом, вычисляющим функцию F^n, в выше приведенном смысле [1]).

Теорема 0.1. *Какова бы ни была частично рекурсивная функция F^n, свойство набора $(t; x_1, \ldots, x_n)$, состоящее в том, что значение $F^n(x_1, \ldots, x_n)$ вычисляется не более чем за t шагов, является общерекурсивным предикатом.*

Определение 0.2. Частично рекурсивная функция $U^{n+1}(i; x_1, \ldots, x_n)$ называется *универсальной* для n-местных частично рекурсивных функций, если для любой частично рекурсивной функции $F^n(x_1, \ldots, x_n)$ найдется i такое, что

$$F^n(x_1, \ldots, x_n) \equiv U^{n+1}(i; x_1, \ldots, x_n). \tag{0.14}$$

Число i будем называть *номером* функции F^n относительно U^{n+1} (функция может иметь много номеров).

Теорема 0.2. *Для любого натурального n существует частично рекурсивная функция, универсальная для всех n-местных частично рекурсивных функций.*

Будем называть *нумерацией множества S^n* любую n-ку общерекурсивных функций F_i $(i = 1, 2, \ldots, n)$, отображающую S на S^n. Натуральное число k называется *номером* n-ки $(x_1, \ldots, x_n)$ в этой нумерации, если $F_i(k) = x_i$ для всех $i = 1, 2, \ldots, n$. Очевидно, пара функций $\pi_1(z)$, $\pi_2(z)$ является нумерацией S^2.

Следующее определение не зависит от нумерации.

Определение 0.3. Множество $X \subseteq S^n$ называется *перечислимым*, если множество номеров его элементов (в выбранной нумерации) является областью значений какой-либо частично рекурсивной функции (при этом говорят, что эта функция *перечисляет* множество X).

Замечание 0.2. Любое перечислимое множество перечисляется также и общерекурсивной функцией.

---

[1]) Определенное таким образом количество шагов является *сигнализирующей* функцией в смысле Трахтенброта [42].

Теорема 0.3. *Пусть есть частично рекурсивный предикат* Π^{n+k}. *Тогда множество* $\{(x_1, \ldots, x_n)$: $\exists a_1, \ldots, a_k\, \Pi^{n+k}(x_1, \ldots, x_n; a_1, \ldots, a_k)$ *истинно*$\}$ *перечислимо.*

Следующая теорема показывает, что семейство перечислимых множеств, зависящих от параметров $p_1, \ldots, p_k$, перечислимо *без повторений.*

Теорема 0.4. *Пусть дано перечислимое множество* $A \subseteq S^{n+k}$. *Тогда существует частично рекурсивная функция* $F(t; p_1, \ldots, p_k)$ *такая, что*

а) *при любых фиксированных* $p_1, \ldots, p_k$ *множество значений функции* $F(t; p_1, \ldots, p_k)$ *будет совпадать с множеством номеров наборов* $(x_1, \ldots, x_n)$ *таких, что* $(x_1, \ldots, x_n; p_1, \ldots, p_k) \in A$ *(номера берутся в некоторой фиксированной нумерации* S^n*);*

б) *если* $t_1 < t_2$ *и* $F(t_2; p_1, \ldots, p_k)$ *определено, то* $F(t_1; p_1, \ldots, p_k)$ *тоже определено и отлично от* $F(t_2; p_1, \ldots, p_k)$.

Определение 0.4. Множество $X \subseteq S^n$ называется *разрешимым*, если существует общерекурсивная функция, равная 0 на X и 1 на $S^n \setminus X$. Последовательность, характеристическую для разрешимого множества, будем называть *вычислимой.*

Очевидно, что всякое разрешимое множество перечислимо.

Теорема 0.5. *Всякое бесконечное перечислимое множество включает в себя бесконечное разрешимое подмножество.*

§ 1. Сложность

В этом параграфе вводится понятие сложности. Выводятся простейшие оценки величины сложности и изучаются алгоритмические свойства этой функции.

1. Определения. Теорема оптимальности. Одним из центральных понятий в этой статье будет понятие сложности некоторого текста (сообщения). Сложностью текста мы будем называть длину самого короткого двоичного слова, содержащего всю информацию, необходимую для восстановления рассматриваемого текста при помощи какого-нибудь фиксированного способа декодирования. Точнее:

Определение 1.1. (А. Н. Колмогоров). Пусть F^1 — произвольная частично рекурсивная функция. Тогда *сложность слова* x *по* F^1 *есть.*

$$K_{F^1}(x) = \begin{cases} \min l(p)\colon F^1(p) = x, \\ \infty, \text{ если } \forall p \in S\ F^1(p) \neq x. \end{cases} \tag{1.1}$$

Слово p такое, что $F^1(p) = x$, будем называть *кодом* или *программой*, по которой F^1 восстанавливает слово x.

Такое определение сложности очень сильно зависит от вида F^1. Однако следующая замечательная теорема позволяет дать инвариантное определение этого понятия (благодаря чему на базе понятия сложности смогла возникнуть теория, излагаемая в статье).

Теорема 1.1. (А. Н. Колмогоров, Р. Соломонов). *Существует частично рекурсивная функция* F^1_0 *(называемая оптимальной) такая, что для любой другой частично рекурсивной функции* G^1

$$K_{F^1_0}(x) \preccurlyeq K_{G^1}(x). \tag{1.2}$$

Доказательство. См. следствие 1.3.

Следствие 1.1. *Для любых двух оптимальных частично рекурсивных функций* F^1 *и* G^1

$$K_{F1}(x) \asymp K_{G1}(x). \tag{1.3}$$

Определение 1.2. *Слжностью* $K(x)$ слова x назовем сложность $K_{F_0^1}(x)$ по некоторой раз и навсегда фиксированной оптимальной частично рекурсивной функции F_0^1 (например, по той, которая будет определена в следствии 1.3).

Определение 1.3 (А. Н. Колмогоров). (*Условная*) *сложность слова* x *при известном* y *по частично рекурсивной функции* F^2 *есть*

$$K_{F^2}(x \mid y) = \begin{cases} \min l(p)\colon F^2(p, y) = x, \\ \infty, \text{ если } \forall p \in S\ F^2(p, y) \neq x. \end{cases} \tag{1.4}$$

Теорема 1.2 (А. Н. Колмогоров, Р. Соломонов). *Существует частично рекурсивная функция* F_0^2 (*называемая оптимальной*) *такая, что для любой частично рекурсивной функции* G^2

$$K_{F_0^2}(x \mid y) \preccurlyeq K_{G^2}(x \mid y). \tag{1.5}$$

Доказательство. Пусть $U^3(n;\ p,\ y)$ — частично рекурсивная функция, универсальная для всех двуместных частично рекурсивных функций (см. определение 0.2, теорему 0.2). Определим функцию

$$F_0^2(z, y) = U^3(\pi_1(z),\ \pi_2(z),\ y), \tag{1.6}$$

и докажем, что эта функция оптимальна. Действительно, пусть G^2 — частично рекурсивная функция, n_{G^2} — какой-нибудь ее номер (см. определение 0.2), и пусть

$$K_{G^2}(x \mid y) = l_0, \tag{1.7}$$

т. е. существует программа p_0 такая, что $G^2(p_0, y) = x$, $l(p_0) = l_0$, причем среди всех слов p таких, что $G^2(p, y) = x$, слово p_0 имеет минимальную длину. Тогда, если вместо z подставить $z = \bar{n}_{G^2} p_0$, согласно (1.6) мы получим

$$F_0^2(z, y) = F_0^2(\bar{n}_{G^2} p_0, y) = U^3(\pi_1(\bar{n}_{G^2} p_0),\ \pi_2(\bar{n}_{G^2} p_0),\ y) =$$
$$= U^3(n_{G^2};\ p_0, y) = G^2 p_0, y) = x,$$

поэтому из (1.4), (1.7) и (0.2) следует

$$K_{F_0^2}(x \mid y) \leqslant l(z) = l(\bar{n}_{G^2} p_0) = l(\bar{n}_{G^2}) + l(p_0) =$$
$$= l_0 + l(n_{G^2}) = K_{G^2}(x \mid y) + l(\bar{n}_{G^2}) \asymp K_{G^2}(x \mid y),$$

так как $l(\bar{n}_{G^2})$ не зависит от x и y, а зависит только от функции G^2.

Следствие 1.2. *Для любых двух оптимальных частично рекурсивных функций* F^2 *и* G^2

$$K_{F^2}(x \mid y) \asymp K_{G^2}(x \mid y). \tag{1.8}$$

Определение 1.4. (*Условной*) *сложностью слова* x *при известном* y $K(x \mid y)$ назовем сложность $K_{F_0^2}(x \mid y)$ по некоторой раз и навсегда фиксированной оптимальной частично рекурсивной функции F_0^2 (например, по функции, определяемой равенством (1.6)).

Следствие 1.3. *Частично рекурсивная функция*

$$F_0^1(p) = F_0^2(p, \Lambda) \tag{1.9}$$

будет оптимальной в смысле теоремы 1.1.

Доказательство. Покажем, что $K_{F_0^1}(x) \preccurlyeq K_{G^1}(x)$, где G^1 — произвольная частично рекурсивная функция. Действительно, определим $G^2(p, y) = G^1(p)$. Тогда из (1.5) и (1.9) $K_{G^1}(x) = K_{G^2}(x \mid \Lambda) \succcurlyeq K_{F_0^2}(x \mid \Lambda) = K_{F_0^1}(x)$, что и требовалось доказать.

В дальнейшем F_0^1 и F_0^2 будут обозначать раз и навсегда выбранные оптимальные функции.

2. Оценки величины сложности. В этом пункте мы докажем наиболее важные для дальнейшего оценки величин $K(x)$ и $K(x \mid y)$.

Теорема 1.3 (А. Н. Колмогоров). *Пусть A — перечислимое множество пар (x, a), и пусть $M_a = \{x: (x, a) \in A\}$. Тогда*

$$K(x \mid a) \preccurlyeq l(d(M_a)). \tag{1.10}$$

Доказательство. Пусть частично рекурсивная функция $F^2(p, a)$ вычисляется следующим алгоритмом: мы выбираем p-ю в порядке перечисления без повторений (см. теорему 0.4) пару вида (x, a) и выдаем в качестве значения функции F^2 первый элемент этой пары (т. е. слово x). Очевидно, если $x \in M_a$, то найдется $p \leqslant d(M_a)$ такое, что $F^2(p, a) = x$; отсюда согласно (1.5) $K(x \mid a) \preccurlyeq K_{F^2}(x \mid a) \leqslant l(d(M_a))$, что и требовалось доказать.

Замечание 1.1. Для произвольного слова y и конечного множества M доля тех $x \in M$, для которых

$$K(x \mid y) \leqslant l(d(M)) - m, \tag{1.11}$$

не превосходит 2^{-m+1}. Действительно, если $K(x \mid y) \leqslant n$, то найдется слово p длины, не превосходящей n, такое, что $F_0^2(p, y) = x$. Значит, количество таких слов x заведомо не превосходит количества всех программ p длины, не превосходящей n; количество таких программ p равно $2^{n+1} - 1$ (см. (0.5)). В свою очередь $d(M) \geqslant 2^{l(d(M))} - 1$. В итоге доля слов $x \in M$, удовлетворяющих условию (1.11), не превосходит $\dfrac{2^{l(d(M))-m+1} - 1}{2^{l(d(M))} - 1} < 2^{-m+1}$. Таким образом, оценка теоремы 1.3 для большинства слов точная; эта теорема часто дает возможность получать наилучшие (т. е. вообще говоря, неулучшаемые) оценки сложности многих типов слов и будет неоднократно использоваться нами в дальнейшем.

Докажем несколько свойств абсолютной (т. е. не условной) сложности.

Теорема 1.4 (А. Н. Колмогоров). *Справедливы утверждения:*

(1.12) а) *величина* $K(x) \preccurlyeq l(x)$

(*следовательно*, $K(x) < \infty$ *для всех* $x \in S$);

б) *доля слов* x, *для которых* $K(x) < l_0 - m$, *среди всех слов* x, $l(x) = l_0$, *не превосходит* 2^{-m+1} (*т. е. оценка* (1.12) *для большинства слов точная*);

в) *предел*

$$\lim_{x \to \infty} K(x) = \infty \tag{1.13}$$

(*следовательно, и* $\lim\limits_{x\to\infty} m(x)=\infty$, *где*

$$m(x)=\min_{y\geqslant x} K(y), \tag{1.14}$$

т. е. $m(x)$ *— наибольшая монотонно неубывающая функция, ограничивающая* $K(x)$ *снизу*);

г) *для любой монотонно стремящейся к бесконечности частично рекурсивной функции* $\Phi(x)$ *начиная с некоторого* x_0 $m(x)<\Phi(x)$ (*т. е.* $m(x)$ *хотя и стремится к бесконечности, но медленнее любой частично рекурсивной монотонно стремящейся к бесконечности функции*);

(1.15) д) *справедливо* $|K(x+h)-K(x)|\leqslant 2l(h)$

(*т. е. функция* $K(x)$ *хотя и колеблется все время между* $l(x)$ *и* $m(x)$, *но делает это довольно плавно*).

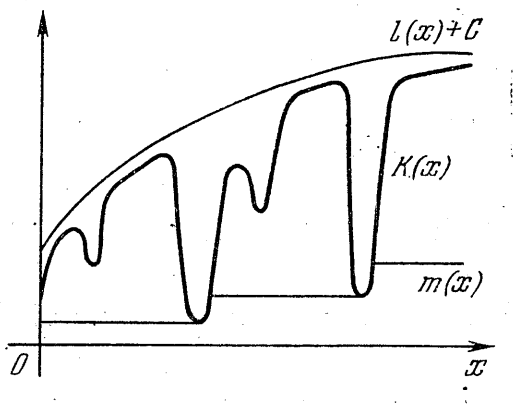

Рис. 2.

Доказательство (рис. 2). а) Пусть $G^1(x)=x$; тогда $K_{G1}(x)=l(x)$ и по теореме 1.1 $K(x)\leqslant K_{G1}(x)=l(x)$, что и требовалось доказать.

б) Это утверждение является тривиальным следствием замечания 1.1 (для $y=\Lambda$). Добавим к этому, что для любого l_0 найдется слово x длины l_0 такое, что $K(x)\geqslant l_0$ (так как количество текстов, имеющих длину l_0, равно 2^{l_0}, а количество программ, имеющих длину меньше l_0, равно $2^{l_0}-1$).

в) По аналогии с замечанием 1.1 количество слов x таких, что $K(x)\leqslant a$, не превосходит 2^{a+1}, т. е. конечно, значит, для любого a найдется x_0 ($x_0=\max\limits_{K(x)\leqslant a} x$) такое, что $K(x)>a$ для всех $x>x_0$, что и требовалось доказать.

г) Пусть утверждение теоремы неверно, т. е. существует частично рекурсивная монотонно стремящаяся к бесконечности функция $\Phi(x)$ такая, что $\Phi(x)\leqslant m(x)$ в бесконечном множестве точек x. Функция $\Phi(x)$ определена на бесконечном перечислимом множестве U. По теореме 0.5 U содержит бесконечное разрешимое подмножество V. Положим

$$\Psi(x)=\begin{cases}\Phi(x)\dot{-}1^{1}), & x\in V,\\ \Phi(\max\limits_{y\leqslant x,\ y\in V} y)\dot{-}1, & x\overline{\in} V.\end{cases}$$

Построенная функция $\Psi(x)$ общерекурсивна, монотонно стремится к бесконечности, и $\Psi(x)\leqslant m(x)$ на бесконечном множестве точек x. Обозначим $M(a)=\max\limits_{K(x)\leqslant a} x$. Легко проверить, что $M(a)+1=\min\limits_{m(x)>a} x$. Нетрудно показать, что $\max\limits_{\Psi(x)\leqslant a} x\geqslant \min\limits_{m(x)>a} x>M(a)$ на бесконечном множестве точек a, причем функция $F(a)=\max\limits_{\Psi(x)\leqslant a} x$, очевидно, общерекурсивна. Таким образом,

---

[1]) $a\dot{-}b=\max\{a-b;\ 0\}$; эта операция вводится для того, чтобы не выходить за пределы множества натуральных чисел.

$F(a) > M(a) = \max\limits_{K(x)\leqslant a} x$ на бесконечном множестве точек a, т. е. $K(F(a)) > a$. Но по теореме 1.1 $K(F(a)) \lesssim K_F(F(a)) \leqslant l(a)$. Отсюда найдется константа C такая, что $l(a) + C > a$ для бесконечно большого количества чисел a, что невозможно.

д) Пусть p_x — программа минимальной длины для слова x, т. е. $F_0^1(p_x) = x$ и $K(x) = l(p_x)$. Тогда слово $x+h$ можно получить из программы $\overline{h}p_x$, если применить к ней функцию $G^1(z) = F_0^1(\pi_2(z)) + \pi_1(z)$; поэтому из (0.2) и (0.10)

$$K_{G1}(x+h) \leqslant l(\overline{h}p_x) = l(\overline{h}) + l(p_x) \asymp 2l(h) + l(p_x) = 2l(h) + K(x).$$

Но $K(x+h) \lesssim K_{G1}(x+h)$, откуда $K(x+h) \lesssim K(x) + 2l(h)$, или $K(x+h) - K(x) \lesssim 2l(h)$. Аналогично, применяя функцию $H^1(z) = F_0^1(\pi_2(z)) \dot{-} \pi_1(z)$ [1]) к слову $\overline{h}p_{x+h}$, где p_{x+h} — программа слова $x+h$, получим неравенство

$$K(x) - K(x+h) \lesssim 2l(h).$$

3. Алгоритмические свойства сложности. Теорема 1.5 (А. Н. Колмогоров). а) *Функция $K(x)$ не частично рекурсивна, и, более того, никакая частично рекурсивная функция $\Phi(x)$, определенная на бесконечном множестве точек, не может во всей своей области определения совпадать с $K(x)$.*

б) *Существует общерекурсивная функция $H(t, x)$, монотонно не возрастающая по t, такая, что*

$$\lim_{t\to\infty} H(t, x) = K(x) \tag{1.16}$$

(т. е. хотя и нет способа вычислять $K(x)$, все же есть возможность получать сколь угодно хорошие оценки сверху этой величины).

Доказательство. а) Выделим в области определения U функции $\Phi(x)$ бесконечное разрешимое подмножество V (см. теорему 0.5). Функция $F(m) = \min\limits_{K(x)\geqslant m,\ x\in V} x$ общерекурсивная (так как $K(x) = \Phi(x)$ на V) и принимает сколь угодно большие значения, причем $K(F(m)) \geqslant m$ (по построению). Но, с другой стороны, $K(F(m)) \lesssim K_F(F(m)) \leqslant l(m)$, откуда $m \lesssim l(m)$, что неверно.

б) Пусть C — достаточно большая константа (такая, что $K(x) < l(x) + C$). Возьмем алгоритм, вычисляющий функцию F_0^1, и заставим его совершить по t шагов (см. замечание 0.1) на всех словах p длины, меньшей $l(x) + C$. Если слово x еще не получилось в качестве результата, положим $H(t, x) = l(x) + C$; если оно уже получилось (и, возможно, не один раз) в качестве результата, положим $H(t, x)$ равным минимальной длине программ p, из которых получено слово x. Ясно, что $H(t, x)$ общерекурсивна и монотонно не убывает по t. Если мы будем совершать все больше и больше шагов алгоритма, вычисляющего $F_0^1(p)$ (т. е. когда $t \to \infty$), мы, наконец, получим x из его «настоящей» программы p_0 минимальной длины, т. е. найдем сложность x ($K(x) = l(p_0)$) (правда, мы ни на каком шагу не сможем узнать, произошло это уже или нет).

---

[1]) См. сноску на стр. 93.

Теорема 1.6 (Я. М. Барздинь). *Пусть* $f(x)$ — *общерекурсивная функция и* $\lim\limits_{x\to\infty} f(x) = \infty$. *Тогда множество* $A = \{x : K(x) \leqslant f(x)\}$ *перечислимо* (*и вообще, предикат* $\Pi(x, a) \sim [K(x) \leqslant a]$ *частично рекурсивен*). *Дополнение к* A *бесконечно, но не содержит никакого бесконечного перечислимого подмножества* (*такие множества* A *называются простыми*).

Доказательство. Утверждение $[K(x) \leqslant a]$ эквивалентно утверждению $[\exists t : H(t, x) \leqslant a]$ (см. теорему 1.5б), что и доказывает первую часть теоремы.

Пусть D — бесконечное перечислимое множество, лежащее в дополнении к A, и пусть функция G^1 действует следующим образом: она берет первое в порядке перечисления без повторений (см. теорему 0.4) число $x \in D$ такое, что $f(x) \geqslant n$, и полагает $G^1(n) = x$. Ясно, что $K(x) \preccurlyeq K_{G^1}(x) \leqslant l(n)$. Но число x лежит в дополнении к A, т. е. по определению $K(x) \geqslant f(x)$, откуда $K(x) \geqslant n$ и $l(n) \geqslant n$, что неверно.

4. Мажоранты сложности. Очевидно, если мы знаем само слово x и его сложность, то можно эффективно (например, перебором) найти одну из программ наименьшей длины, кодирующих слово x. Более того, если мы знаем слово x и какое-нибудь число $s \geqslant K(x)$, то можно эффективно найти одну из программ слова x, которая хотя, возможно, и не будет самой короткой, но все же будет иметь длину, не превосходящую s. Поскольку, как следует из теоремы 1.5, эффективно найти сложность нельзя, на практике приходится довольствоваться эффективно вычислимыми (точнее, частично рекурсивными) функциями, которые во всей своей области определения не меньше сложности, т. е. дают оценку длины кода, хотя и не самого короткого, но зато эффективно вычислимого.

Определение 1.5. Будем называть *мажорантой сложности* любую частично рекурсивную функцию $\Phi(x)$, для которой

$$K(x) \preccurlyeq \Phi(x). \tag{1.17}$$

Теорема 1.7 (Л. А. Левин). *Частично рекурсивная функция* $\Phi(x)$ *тогда и только тогда будет мажорантой сложности, когда*

$$l(d\{x : \Phi(x) = a\}) \preccurlyeq a. \tag{1.18}$$

Доказательство. Пусть Φ — мажоранта сложности, и x принадлежит области ее определения; $\Phi(x) = a$. Согласно (1.17) найдется константа C такая, что $K(x) \leqslant \Phi(x) + C$, откуда $d\{x : \Phi(x) = a\}$ не превосходит количества слов x таких, что $K(x) \leqslant a + C$, следовательно (аналогично замечанию 1.1),

$$d\{x : \Phi(x) = a\} \leqslant 2^{a+C+1} \quad \text{и} \quad l(d\{x : \Phi(x) = a\}) \leqslant a + C + 1,$$

что и доказывает теорему в одну сторону.

Пусть теперь для частично рекурсивной функции Φ выполнено условие (1.18), т. е. существует константа C такая, что $d\{x : \Phi(x) = a\} \leqslant 2^{a+C}$ для всех a. Если $\Phi(x) = a$, то слово x можно закодировать следующим образом: пусть $F(i, a)$ перечисляет без повторений все слова y такие, что $\Phi(y) = a$ (предикат $[\Phi(x) = a]$ частично рекурсивен, поэтому такая функция $F(i, a)$ существует; см. теоремы 0.4 и 0.3 и определение 0.1). Запишем слово i, для

которого $F(i, a) = x$ (легко видеть, что $i \leqslant 2^{a+C}$), поставим перед ним цифру 1 и припишем слева столько нулей, чтобы длина слова стала $a + C + 1$. По этому слову легко восстановить x (сначала найти a, для чего вычесть из длины кода $C + 1$, затем найти i, для чего отбросить слева все нули и первую единицу, и выдать слово $F(i, a)$), поэтому $K(x) \preccurlyeq a + C + 1 = \Phi(x) + C + 1$, что и доказывает теорему в другую сторону.

Замечание 1.2. Из любой частично рекурсивной функции $F(x)$ можно сделать мажоранту сложности, сузив ее область определения до множества тех x, для которых $F(x) \geqslant K(x)$ (априори не очевидно, что полученная функция будет частично рекурсивной, однако это легко следует из теоремы 1.5б). Отсюда, в частности, легко следует перечислимость множества мажорант сложностей.

Для практики особый интерес представляют общерекурсивные мажоранты сложности [1]), так как при поиске короткого кода слова важно быть уверенным, что мы рано или поздно хоть какой-нибудь код найдем. Примерами таких общерекурсивных мажорант могут служить сложности по какой-нибудь общерекурсивной функции (см. определение 1.1 [2])). В теореме 5.1 приведен еще один важный пример мажоранты сложности — «подправленная» энтропия Шеннона.

Сложность интересно рассматривать постольку, поскольку она является (с точностью до аддитивной константы) точной нижней гранью мажорант сложностей (см. теорему 1.5б), поэтому для широкого класса утверждений их формулировка для сложности является обобщением их формулировок для всех мажорант сложностей. Замечателен факт, что даже в таком сильном виде эти утверждения остаются справедливыми.

Замечание 1.3. Все результаты пп. 3 и 4, а также определение мажоранты сложности без труда переносятся на случай условной сложности $K(x|y)$; при этом во всех формулировках и доказательствах слово y будет фигурировать как параметр.

§. 2. Алгоритмические проблемы и сложность разрешения

Изучается поведение сложности фрагментов различных бесконечных двоичных последовательностей. С этой целью вводится понятие сложности разрешения, более приспособленное для изучения бесконечных последовательностей, чем $K(x)$.

1. Определение и простейшие свойства. В предыдущем параграфе был развит аппарат сложностей таких слов, интерпретацией которых являются законченные тексты. Однако на практике часто приходится рассматривать слова, представляющие последовательности, оборванные в более или менее произвольном месте. Примерами таких слов являются приближенные значения физических констант, куски потока телеграмм, таблицы случайных чисел, подшивки газет до данного числа и т. п. Измерять сложность алгоритма, восстанавливающего такое слово, неинтересно, так как, даже обладая

---

[1]) О таких функциях см., например, [19], а также теорему 2.5 настоящей статьи.

[2]) Общерекурсивная функция, конечно, не может быть оптимальной.

полной информацией о всей последовательности, мы не будем знать, на каком знаке она оборвана. Измерять же сложность слова при известной длине (т. е. предполагая уже заданным место обрыва) тоже неестественно, так как случайно может оказаться, что длина слова содержит дополнительную информацию о нем (например, двоичная запись длины может совпадать с началом слова). Гораздо естественнее измерять сложность алгоритма (или кода), который по каждому числу $i \leqslant l(x)$ выдает i-й знак рассматриваемого слова, т. е. моделирует деятельность источника последовательности до i-го знака.

Определение 2.1 (Д. Ловеланд)[1]. *Сложностью разрешения слова x по частично рекурсивной функции F^2 называется*

$$KR_{F^2}(x) = \begin{cases} \min l(p) : \forall i \leqslant l(x)\ F^2(p, i) = x_i, \\ \infty, \text{ если такого } p \text{ не существует} \end{cases} \tag{2.1}$$

(*здесь x_i — i-й знак слова x*).

Теорема 2.1. (Д. Ловеланд). *Существует (оптимальная) частично рекурсивная функция G_0^2 такая, что для любой частично рекурсивной функции F^2*

$$KR_{G_0^2}(x) \preccurlyeq KR_{F^2}(x). \tag{2.2}$$

Доказательство аналогично доказательству теоремы 1.2.

Определение 2.2. *Сложностью разрешения $KR(x)$* слова x называется сложность его разрешения по некоторой раз и навсегда фиксированной оптимальной частично рекурсивной функции[2]).

Свойства функции $KR(x)$ аналогичны свойствам $K(x)$, и их без труда установит читатель. Мы отметим только некоторые из них.

Теорема 2.2 (Д. Ловеланд). а) *Если $x \subset y$, то*

$$KR(x) \leqslant KR(y). \tag{2.3}$$

б) *Бесконечная последовательность ω вычислима тогда и только тогда, когда сложность разрешения ее фрагментов ограничена.*

в) *Справедливо*

$$K(x) \gtrsim KR(x) \gtrsim K(x \mid l(x)). \tag{2.4}$$

Доказательство очевидно.

2. Вычислимые последовательности. Существует и более нетривиальная связь между величинами $KR(x)$ и $K(x \mid l(x))$.

Теорема 2.3 (А. Н. Колодий, Л. А. Левин, Д. Ловеланд, В. А. Мишин). *Для $\omega \in \Omega$ величина $K((\omega)_n \mid n)$ ограничена тогда и только тогда, когда ограничена $KR((\omega)_n)$*[3]).

Доказательство. В одну сторону утверждение очевидно: если последовательность вычислима, то существует общерекурсивная функция $F^1(n) = (\omega)_n$. Положим $F^2(p, n) = F^1(n)$; тогда $K_{F^2}((\omega)_n \mid n) = l(\Lambda) = 0$, так

---

[1]) Аналогичные понятия были рассмотрены А. А. Марковым (см. [15]).

[2]) Эту функцию в дальнейшем всегда будем обозначать G_0^2.

[3]) Однако, как показал Н. В. Петри, не существует эффективного способа оценки $KR((\omega)_n)$ по константе, ограничивающей $K((\omega)_n \mid n)$, т. е. первая может быть очень велика.

как $F^2(\Lambda, n) = (\omega)_n$; следовательно, и $K((\omega)_n|n) \preccurlyeq 0$, т. е.

$$K((\omega_n)\,|\,n) \leqslant C. \tag{2.5}$$

Докажем обратное утверждение. Пусть выполняется условие (2.5). Мы хотим доказать существование процедуры, которая бы по номеру n выдавала ω_n — n-й знак последовательности ω. Выпишем в столбик все слова p длины, не превосходящей C, и построим следующую таблицу:

	0	1	2	...	n	...
Λ						
0		$(\omega)_1$				
1					$(\omega)_n$	
0 0		$(\omega)_1$				
p					$F_0^2(p, n)$	
	$(\omega)_0$		$(\omega)_2$			
$\underbrace{11\ldots1}_{C}$						

в n-м столбце против слова p стоит $F_0^2(p, n)$ (см. (1.6)), если функция F_0^2 определена на паре (p, n). Множество слов $F_0^2(p, n)$, стоящих в n-м столбце, обозначим A_n. В каждом A_n не более 2^{C+1} слов, причем обязательно $(\omega)_n \in A_n$. Пусть

$$l = \overline{\lim_{n\to\infty}}\, d(A_n).$$

Очевидно, множество

$$U = \{n : d(A_n) \geqslant l\}$$

перечислимо и бесконечно. При этом в силу определения l только для конечного количества чисел n $d(A_n) > l$; наибольшее из таких чисел n обозначим m_1.

Пусть количество последовательностей ω, удовлетворяющих условию (2.5), равно k. Обозначим через m_2 минимальное число, такое, что все m_2-фрагменты этих k последовательностей различны [кстати, во всех столбцах, начиная с m_2-го, должно стоять по меньшей мере k слов — фрагментов этих последовательностей (эти фрагменты будут различны); поэтому $k \leqslant l$]. Пусть $m = \max(m_1, m_2)$ [1]).

---

[1]) Приводимое построение алгоритма использует числа l, k и m. Это построение не эффективно, так как не приводится эффективной процедуры построения чисел l, k и m (см. сноску [3]) на стр. 97). Мы лишь доказываем, что искомый алгоритм существует (интуиционист выразился бы: «Не может не существовать»), поэтому для нас достаточно лишь самого факта *существования* чисел l, k и m.

Выберем из U бесконечное разрешимое подмножество U' (см. теорему 0.5). Пусть $V = U' \cap \{n\colon n > m\}$; очевидно, V тоже разрешимо. Элементы множества V перенумеруем в порядке возрастания: $V = \{n_1, n_2, \ldots\}$. Алгоритм, разрешающий i-ю (в лексикографическом порядке) из наших k последовательностей, действует так: пусть мы хотим определить j-й знак i-й последовательности. Выберем минимальное $n_r \in V$ такое, что $n_r > j$, и начнем заполнять n_r-й столбец (т. е. строить слова $F_0^2(p, n_r)$, $l(p) \leqslant C$). Как только окажется, что уже построено l слов, мы останавливаемся: мы получили все слова из A_{n_r}. Следующий шаг: выберем из A_{n_r} слова длины n_r; множество этих слов обозначим B_{n_r}. Дальше, аналогично построим множество $B_{n_{r+1}}$ и выберем из $B_{n_{r+1}}$ слова, являющиеся продолжениями слов из B_{n_r}; множество этих слов обозначим $C_{n_{r+1}}$. Далее, из $B_{n_{r+2}}$ выберем слова, являющиеся продолжениями слов из $C_{n_{r+1}}$ — они образуют множество $C_{n_{r+2}}$; $C_{n_{r+3}}$ — множество слов из $B_{n_{r+3}}$, являющихся продолжениями слов из $C_{n_{r+2}}$, и так далее. Остановимся на том шаге, когда в очередном множестве C_{n_s} окажется ровно k слов: теперь мы уверены, что все слова из C_{n_s} есть n_s-фрагменты последовательностей, удовлетворяющих условию (2.5). Выберем из слов в C_{n_s} i-е по величине слово, и найдем его j-й знак; он и будет искомым.

3. Характеристические последовательности перечислимых множеств. Сложность разрешения вычислимых последовательностей ограничена. Интересно посмотреть, как растет сложность разрешения у последовательностей, имеющих более сложную алгоритмическую структуру (например, характеристических последовательностей перечислимых множеств).

Теорема 2.4 (Я. М. Барздинь). а) *Для любой последовательности* ω *с перечислимым* S_ω

$$KR((\omega)_n) \preccurlyeq l(n). \tag{2.6}$$

б) *Существует последовательность с перечислимым* S_ω *такая, что*

$$KR((\omega_n)) > l(n). \tag{2.7}$$

Доказательство. а) Пусть $F(x)$ — функция, перечисляющая множество S_ω без повторений (см. теорему 0.4). Для того чтобы мы могли полностью восстановить слово $(\omega)_n$, достаточно задать число s — последнее (в порядке получения) значение функции F, не превосходящее n. Действительно, пусть $F^2(k, i)$ получается следующим образом: вычисляем значения $F(x)$ до тех пор, пока не получим число k (если $\forall x \in S\ F(x) \neq k$, то $F^2(k, i)$ не определено), после чего полагаем $F^2(k, i) = 1$, если i уже появилось среди значений $F(x)$, и $F^2(k, i) = 0$ в противном случае. Тогда $\omega_i = F^2(s, i)$ для всех $i \leqslant n$, откуда из (2.1) $KR_{F^2}((\omega)_n) = l(s) \leqslant l(n)$. Но согласно (2.2) $KR((\omega)_n) \preccurlyeq KR_{F^2}((\omega)_n)$, следовательно, верно (2.6).

б) Положим

$$\omega_i = \begin{cases} 1, & \text{если } G_0^2(i, i) = 0, \\ 0, & \text{если } G_0^2(i, i) \neq 0 \text{ или не определено} \end{cases}$$

(здесь G_0^2 из теоремы 2.1). Утверждается, что для такой последовательности ω (S_ω, очевидно, перечислимо) выполняется неравенство (2.7). Действительно, пусть $KR((\omega)_n) \leqslant l(n)$ для какого-то n; тогда существует $p \leqslant n$ такое, что $G_0^2(p, i) = \omega_i$ для всех $i \leqslant n$. В частности, так как $p \leqslant n$, то и $G_0^2(p, p) = \omega_p$, что противоречит определению ω_p.

Приведем без доказательства один результат, связывающий строение последовательностей с перечислимым S_ω с их сложностью (получен М. И. Кановичем).

Определение *процесса* и понятий, связанных с ним, дается на стр. 102. Назовем последовательность α с перечислимым S_α *универсальной*, если для любой последовательности β с перечислимым S_β существует быстрорастущий (слаботабличный) процесс F такой, что $\beta = F(\alpha)$. Последовательность α назовем *достаточно сложной*, если существует неограниченная общерекурсивная функция $F(n)$ такая, что $KR((\omega)_n) \geqslant F(n)$.

Предложение 2.1. *Понятия универсальности и достаточной сложности последовательности α с перечислимым S_α эквивалентны.*

Следствие 2.1. *Всякая достаточно сложная последовательность α с перечислимым S_α универсальна относительно сводимости по Тьюрингу.*

Примечательно, что в случае последовательностей с перечислимым S_ω общерекурсивные мажоранты сложности (которые, собственно, и представляют практический интерес) ведут себя совсем иначе, чем сама сложность [1]).

Теорема 2.5 (Я. М. Барздинь, Н. В. Петри). *Существует последовательность ω с перечислимым S_ω такая, что для любой общерекурсивной мажоранты сложности Φ найдется константа C такая, что*

$$\Phi((\omega)_n) \geqslant \frac{n}{C}. \tag{2.8}$$

Доказательство. Построим нужную нам последовательность. Она будет состоять из записанных друг за другом кусков удваивающейся длины (длина i-го куска равна 2^i). Кусок с номером i заполняется следующим образом: рассмотрим частично рекурсивную функцию F с номером k (см. определение 0.2), где k есть максимальная степень, с которой двойка входит сомножителем в i (номера i, для которых k одно и то же, образуют арифметическую прогрессию с разностью 2^{k+1}), и выдадим в качестве i-го куска последовательности ω первое (в порядке получения при переборе) слово x длины 2^i, для которого $F(x) \geqslant l(x) = 2^i$, а если такого слова x нет (в чем алгоритмически убедиться, вообще говоря, нельзя), то i-й кусок пусть состоит из одних нулей. Легко видеть, что S_ω перечислимо.

Будем говорить, что i-й кусок ω «определяется» k-й функцией.

Пусть $G(x)$ — общерекурсивная мажоранта сложности.

Без ограничения общности можно считать, что для функции $G(x)$ в теореме 1.7 выполнено строгое неравенство $<$ (вместо $\preccurlyeq$) [2]). Тогда из этой теоремы вытекает, что для любого i найдется слово x длины 2^i такое, что $G(yxz) \geqslant l(x) = 2^i$ для всех y, z, а следовательно, все куски, определяемые функцией G, не тривиальны.

Оценим $G((\omega)_n)$. Для этого рассмотрим последний целиком лежащий в $(\omega)_n$ кусок x, «определяемый» функцией G. Номер i этого куска удовлет-

[1]) Подробнее об этом см. [19].

[2]) Для этого достаточно увеличить $G(x)$ на константу, что не изменит ее асимптотического поведения.

воряет неравенству $i \geqslant l(n) - 2^{k+1} - 1$, где k — номер функции G (это неравенство следует из неравенства $2^{i+2^{k+1}} \geqslant \frac{n}{2}$).

Пусть y и z — слова, дополняющие x до $(\omega)_n$ (т. е. $yxz = (\omega)_n$; очевидно, $l(y) = 2^i - 1$, $l(z) \leqslant 2^{2^i}$). Тогда $G((\omega)_n) = G(yxz) \geqslant l(x) = 2^i \geqslant 2^{l(n)-2^{k+2}} = n/2^{2^{k+2}}$, что и доказывает теорему, если в качестве C выбрать $2^{2^{k+2}}$; константа C зависит только от G, так как только от нее зависит k (k есть номер G).

4. Максимально сложные последовательности. Разрешимые и перечислимые множества — это множества соответственно нулевого и первого ранга проективной классификации Клини. Если рассмотреть последовательности с более сложным множеством S_ω, например, второго ранга, т. е. задаваемые двухкванторным предикатом, то среди них уже существуют максимально сложные последовательности (сложность разрешения их фрагментов асимптотически равна длине этих фрагментов). Более точно этот факт будет сформулирован в теореме 4.5 и следствии 4.1. Там будет доказано, что существует двухкванторная последовательность, сложность n-фрагментов которой отличается от n не более чем на $4l(n)$. Здесь мы покажем, что величину $4l(n)$ уменьшить по порядку нельзя. Хотя для любого n есть слово x длины n такое, что $K(x) > n$ (см. доказательство теоремы 1.4б), однако не существует последовательности, для которой выполнялось бы неравенство $K((\omega)_n) \gtrsim n$. Более того, имеет место

Теорема 2.6 (П. Мартин-Лёф). *Для любой последовательности* $\omega \in \Omega$ *существует бесконечно много номеров* n *таких, что*

$$K((\omega)_n) \lesssim n - l(n)\ ^{1}).$$

Доказательство. Определим среди всех слов длины n совокупность «выделенных» слов A_n следующим образом (по индукции): пусть мы определили все выделенные слова в $(n-1)$-й строке, и пусть наибольшее из них равно y; тогда выделяем в n-й строке $2^{n-l(n)}$ слов, начиная со слова, следующего за $y1$ (см. рис. 1) [если они все в этой строке не уместились, то оставшееся количество выделяем в начале следующей строки, и дальше начинаем выделять уже слова из $(n+2)$-й строки]. Ясно, что любая последовательность имеет бесконечно много выделенных фрагментов. (Этот факт легче понять самому, чем кому-нибудь объяснить. Он следует из того, что доля выделенных слов в n-й строке (как правило) равна $2^{-l(n)} \approx 1/n$, а ряд $\Sigma 1/n$ расходится.)

---

1) На самом деле П. Мартин-Лёф установил более точный факт, который мы приведем без доказательства. Пусть $F(n)$ — общекурсивная функция. Будем говорить, что ω F-сложна, если $K((\omega)_n) \gtrsim n - F(n)$. Тогда: а) если ряд $\sum_{n=1}^{\infty} 2^{-F(n)} = \infty$, то не существует F-сложных последовательностей; б) если ряд $\sum_{n=1}^{\infty} 2^{-F(n)} < \infty$, то существуют двухкванторные F-сложные последовательности, причем F-сложные последовательности образуют множество полной меры (по мере L — см. стр. 103).

Пусть x — выделенное слово длины n. Очевидно, что

$$K(x) \lesssim l\left(d\left\{\bigcup_{k=0}^{n} A_k\right\}\right) \leqslant l\left(\sum_{k=0}^{n} 2^{k-l(k)}\right) \lesssim n - l(n)\ ^{1}).$$

§ 3. Эффективные случайные процессы

В этом параграфе рассматриваются эффективные детерминированные и недетерминированные процессы (алгоритмы со случайными входом), производящие последовательности. Центральным результатом является построение универсальной полувычислимой меры и выяснение ее связи со сложностью.

1. Определения. Эквивалентность мер. Определение 3.1. *Алгоритмическим процессом* или просто *процессом* назовем частично рекурсивную функцию F, отображающую слова в слова, такую, что если для слова x определена $F(x)$ и $y \subset x$, то $F(y)$ также определена, и $F(y) \subset F(x)$.

Пусть ω — некоторая бесконечная последовательность. Будем применять процесс F последовательно ко всем фрагментам ω до тех пор, пока это возможно (т. е. пока F определена). В результате мы будем получать фрагменты некоторой новой последовательности ρ (возможно, конечной, или даже пустой [2])) — *результата* применения процесса F к ω (т. е. процесс F отображает Ω в Ω^*). В этом случае будет использоваться также обозначение $\rho = F(\omega)$.

Замечание 3.1. Существует *универсальный* процесс, т. е. частично рекурсивная функция $H(i, x)$ такая, что для любого i $H(i, x)$ есть процесс и для любого процесса $F(x)$ существует i такое, что

$$H(i, x) \equiv F(x). \tag{3.1}$$

Функцию $H(i, x)$ легко построить из универсальной частично рекурсивной функции $U^2(i, x)$ (см. определение 0.2). Без ограничения общности можно считать, что

$$H(\Lambda, \Lambda) = \Lambda \tag{3.2}$$

(это понадобится нам в дальнейшем). Процессы F и G будем называть *эквивалентными*, если $F(\omega) = G(\omega)$ для любой $\omega \in \Omega$.

Замечание 3.2. Для любого процесса существует эквивалентный ему примитивно рекурсивный процесс.

Определение 3.2. Будем говорить, что процесс *применим* к последовательности, если результатом его применения к ней является бесконечная последовательность.

Замечание 3.3. Любой процесс на множестве тех последовательностей, к которым он применим, является непрерывной функцией (относи-

---

[1]) Последнее неравенство следует из оценки $\sum_{k=0}^{n} 2^{k-l(k)} \leqslant C \cdot 2^{n-l(n)}$.

[2]) Если для некоторого n $F((\omega)_n)$ определена и все $F((\omega)_m)$, $m > n$, совпадают с $F((\omega)_n)$ или не определены, то *результатом* будет $F((\omega)_n)$. Пустое слово получится в случае, когда $F((\omega)_n)$ при всех n не определено или пусто.

тельно естественной топологии пространства бесконечных двоичных последовательностей [1]).

Определение 3.3. Будем называть процесс F слаботабличным или *быстрорастущим* (*быстроприменимым* к последовательности ω), если существует монотонная неограниченная общерекурсивная функция $\Phi(n)$ такая, что для любых x (для любых x, являющихся фрагментами ω) и n таких, что $l(x) = n$ и $F(x)$ определена, длина слова $F(x)$ будет не меньше $\Phi(n)$. Будем в этом случае говорить, что *скорость роста* (*применимости* к ω) процесса F не меньше $\Phi(n)$.

Замечание 3.4. Легко показать, что процесс, применимый ко всем $\omega \in \Omega$, является общерекурсивным и быстрорастущим. Очевидно, верно и обратное.

Определение. 3.4. Пусть P — вероятностная мера на Ω. Будем говорить, что процесс *P-регулярен*, если множество последовательностей, к которым он применим, имеет P-меру 1.

Для того чтобы задать произвольную меру на борелевской σ-алгебре подмножеств Ω, достаточно задать ее значения на множествах Γ_x.

Определение 3.5. Назовем меру P на Ω *вычислимой*, если существуют общерекурсивные функции $F(x, n)$ и $G(x, n)$ такие, что рациональное число

$$\alpha_P(x, n) = \frac{F(x, n)}{G(x, n)} \tag{3.3}$$

приближает число $P\{\Gamma_x\}$ с точностью до 2^{-n}.

Замечание 3.5. Очевидно, если мера P вычислима, то число $\alpha_P(x, n+1) + 2^{-(n+1)}$ приближает меру $P\{\Gamma_x\}$ с точностью до 2^{-n} с избытком. Поэтому мы в дальнейшем без ограничения общности всегда будем считать, что $\alpha_P(x, n)$ уже является приближением с избытком, а в качестве приближения с недостатком с точностью 2^{-n} будем брать $\alpha_P(x, n) - 2^{-n}$.

Будем обозначать L и называть *равномерной* меру

$$L\{\Gamma_x\} = 2^{-l(x)}. \tag{3.4}$$

Эта мера соответствует испытаниям Бернулли с вероятностью $p = 1/2$, она же — мера Лебега на отрезке $[0, 1]$. Мера L, очевидно, вычислима.

Теорема 3.1 (Л. А. Левин). а) *Для любой вычислимой меры P и любого P-регулярного процесса F мера*

$$Q\{\Gamma_y\} = P\{\bigcup \Gamma_x : F(x) = y\} \tag{3.5}$$

(т. е. мера, по которой будут распределены результаты процесса F) будет вычислимой.

б) *Для любой вычислимой меры Q существует L-регулярный процесс F такой, что результаты его применения к последовательностям, распределенным по мере L, распределены по мере Q, причем такой, что для него существует процесс G, применимый ко всем последовательностям, кроме, может быть, разрешимых или лежащих на отрезках Q-меры 0, и обратный к F (в области определения процесса $F \circ G$).*

---

[1]) В этой топологии Ω гомеоморфию канторовскому совершенному множеству.

Доказательство. а) Нам нужно уметь вычислять $Q\{\Gamma_y\}$ с точностью до 2^{-n}, т. е. находить $\alpha_Q(y, n)$ [1]). Выберем m такое, чтобы

$$P\{\omega: l(F((\omega)_m)) > l(y)\} > 1 - 2^{-(n+1)}$$

(такое m существует, так как процесс F P-регулярен, причем m легко найти эффективно). Возьмем все слова x длины m такие, что $y \subset F(x)$, и просуммируем для них меры $P\{\Gamma_x\}$, вычисленные с точностью до $2^{-(m+n+1)}$, т. е. положим

$$\alpha_Q(y, n) = \sum_{x:\, l(x)=m,\ y\subset F(x)} \alpha_P(x,\ m+n+1). \tag{3.6}$$

Тогда наша ошибка (т. е. $|\alpha_Q(y, n) - Q\{\Gamma_y\}|$) не превзойдет $2^{-(n+1)} + 2^m \cdot 2^{-(m+n+1)} = 2^{-n}$ (так как слов x, по которым ведется суммирование, не больше, чем 2^m), что и требовалось.

б) Будем рассматривать двоичные последовательности как действительные числа на отрезке [0, 1] (последовательность является двоичным разложением соответствующего ей числа). Все случаи, когда это может привести к недоразумению (из-за неоднозначности разложения в такую последовательность двоично-рациональных чисел), будут особо оговорены. На рис. 3 (где абсциссы распределены по мере Q, ординаты — по мере L) показана функция распределения g, соответствующая мере Q. Как известно, если случайная величина ξ распределена равномерно на отрезке [0, 1], то случайная величина $g^{-1}(\xi)$ распределена по мере Q. На этой идее и будет основано наше построение.

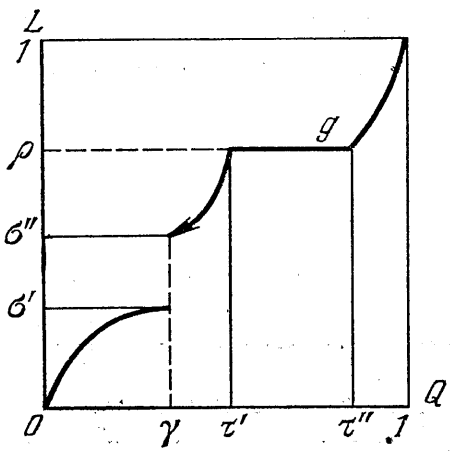

Рис. 3.

I. Построим процесс F, индуцирующий меру Q из меры L (фактически это будет процесс вычисления функции g^{-1}; для возможности такого вычисления существенно требование вычислимости меры Q). Пусть есть последовательность α, и нам дан ее n-фрагмент $(\alpha)_n$. Найдем по нему приближение (с точностью 2^{-n}) с недостатком α'_n и с избытком α''_n числа α. Рассмотрим все слова y длины n; вычислим для каждого из них меру $Q\{\Gamma_y\}$ с избытком с точностью 2^{-2n} (т. е. $\alpha_Q(y, 2n)$). Выделим те слова z длины n, для которых

$$\sum_{y\geqslant z} (\alpha_Q(y, 2n) - 2^{-2n}) \geqslant 1 - \alpha''_n \tag{3.7}$$

(сумма слева есть приближение для $Q\{\bigcup_{y\geqslant z} \Gamma_y\}$ с недостатком с точностью 2^{-n}) и

$$\sum_{y\leqslant z} \alpha_Q(y, 2n) \geqslant \alpha'_n \tag{3.8}$$

(сумма слева есть приближение для $Q\{\bigcup_{y\leqslant z} \Gamma_y\}$ с избытком с точностью 2^{-n}). Выберем наиболее длинный общий фрагмент всех выделенных слов z и выдадим его в качестве значения F на $(\alpha)_n$.

[1]) Мы не будем строить приближение с избытком, а построим произвольное приближение; сделать из него приближение с избытком легко (см. замечание 3.5).

II. Множества $\bigcup \Gamma_z$ согласно (3.7) и (3.8) являются отрезками, содержащими (при каждом n) g-прообраз точки α, поэтому, если процесс F применим к α, его результатом будет $g^{-1}(\alpha)$ (прообразом точек $\alpha \in [\sigma', \sigma'']$ будем считать точку γ; см. рис. 3). Для того чтобы доказать, что процесс F искомый, достаточно, тем самым, доказать его L-регулярность.

1) Пусть α лежит на отрезке типа $[\sigma', \sigma'']$, соответствующей одной-единственной последовательности γ, имеющей положительную меру. Тогда, если α лежит *внутри* отрезка $[\sigma', \sigma'']$, то с того момента, как 2^{-n}-окрестность отрезка $[\alpha'_n, \alpha''_n]$ будет целиком лежать внутри отрезка $[\sigma', \sigma'']$, множество выделенных слов z будет состоять из одного-единственного слова, являющегося n-фрагментом искомой последовательности γ, и следовательно, процесс F будет применим к α. К концам отрезка $[\sigma', \sigma'']$ процесс F, вообще говоря, может быть неприменим.

2) Пусть теперь α не лежит на отрезке типа $[\sigma', \sigma'']$. Тогда из (3.7) и (3.8) следует, что $Q\{\bigcup \Gamma_z\} \to 0$ при $n \to \infty$, откуда, если α не является точкой типа ρ, соответствующей отрезку меры 0, то и сами отрезки $\bigcup \Gamma_z$ стягиваются к одной точке β — g-прообразу α. Поэтому длина наибольшего общего фрагмента выделенных слов z стремится к бесконечности (за исключением, может быть, тех случаев, когда точка β двоично-рациональная, так как если $\beta = m/2^k$, то отрезки $\bigcup \Gamma_z$ могут всегда содержать как последовательности, лежащие слева от $m/2^k$, и, следовательно, начинающиеся на слово $m - 1$, так и последовательности, лежащие справа от $m/2^k$, и, следовательно, начинающиеся на слово m. В этом случае наиболее длинный общий фрагмент всех выделенных слов z будет иметь длину меньше k).

Итак, процесс F может быть неприменим только к последовательностям типа ρ, σ' и σ'' (см. рис. 3), а также к последовательностям, имеющим двоично-рациональные прообразы. Очевидно, что множество таких последовательностей не более чем счетно, следовательно, процесс F L-регулярен.

III. Построить обратный процесс не представляет труда: это будет процесс вычисления функции g. При этом процесс G будет неприменим, во-первых, к последовательностям типа γ, имеющим положительную меру (такие последовательности, как легко показать, вычислимы; мы здесь этого не доказываем, так как в следствии 3.1 будет доказан более общий результат), и, во-вторых, (может быть) к последовательностям β, на которых функция g принимает двоично-рациональные значения α (аналогично II(2)). Если процесс F применим к этим двоично-рациональным значениям α, то наши последовательности β вычислимы (как F-образы двоично-рациональных); если же процесс F неприменим к α, то (см. II) наши последовательности β либо есть точки типа γ (этот случай уже был рассмотрен) либо образуют целый отрезок $[\tau', \tau'']$ Q-меры 0, либо сами двоично-рациональны (следовательно, вычислимы). Теорема доказана.

2. Полувычислимые меры. Определение 3.6 (Л. А. Левин). *Полувычислимой мерой* [1]) называется мера, по которой распределены результаты применения произвольного (не обязательно регулярного) процесса к последовательностям, распределенным по некоторой вычислимой мере.

[1]) Название «полувычислимая» оправдывается теоремой 3.2.

Замечание 3.6. Полувычислимая мера сосредоточена на пространстве Ω^*, так как нерегулярный процесс может выдавать с положительной вероятностью и небесконечные последовательности. Под Γ_x мы будем в дальнейшем (в этом параграфе) понимать множество всех конечных и бесконечных последовательностей, начинающихся со слова x.

Замечание 3.7. Результаты применения любого процесса к последовательностям, распределенным по произвольной полувычислимой мере, распределены также по некоторой полувычислимой мере (так как суперпозиция двух процессов есть процесс), и любую полувычислимую меру можно получить некоторым процессом из равномерной меры (см. теорему 3.1б).

Теорема 3.2 (Л. А. Левин). *Мера P является полувычислимой тогда и только тогда, когда существуют общерекурсивные функции $F(x,t)$ и $G(x, t)$ такие, что функция*

$$\beta_P(x, t) = \frac{F(x, t)}{G(x, t)} \tag{3.9}$$

монотонно не убывает по t, и

$$\lim_{t\to\infty} \beta_P(x, t) = P\{\Gamma_x\}. \tag{3.10}$$

Доказательство. Пусть P — полувычислимая мера. Тогда существует процесс F, получающий эту меру из равномерной. Заставим его совершить по t шагов на всех словах y длины, не превосходящей t; обозначив результат, за $F_t(y)$ (если он еще не получился, то $F_t(y) = \Lambda$), положим

$$\beta_P(x, t) = L\{\bigcup \Gamma_y : x \subset F_t(y)\}. \tag{3.11}$$

Обратно, пусть для меры P существует функция $\beta_P(x, t)$, удовлетворяющая условиям теоремы; мы хотим построить процесс F, получающий меру P из равномерной. Идея этого построения проста: нужно, грубо говоря, разбить отрезок [0, 1] на непересекающиеся множества меры $P\{\Gamma_x\}$, и выдавать слово x в том случае, если наша равномерно распределенная последовательность попала в соответствующее множество. Теперь проведем построение четко. Очевидно, что $P\{\Gamma_x\} \geqslant P\{\Gamma_{x0}\} + P\{\Gamma_{x1}\}$. Более того, без ограничения общности можно считать, что $\beta_P(x, t) \geqslant \beta_P(x0, t) + \beta_P(x1, t)$ при всех t (каждый раз, когда это неравенство не выполняется, можно уменьшить пропорционально $\beta_P(x0, t)$ и $\beta_P(x1, t)$ настолько, чтобы неравенство стало верным; при этом условие (3.10) не нарушится). Легко построить на отрезке [0, 1] множества, удовлетворяющие следующим условиям: каждой паре (x, t) соответствует множество — объединение конечного числа интервалов с рациональными концами, имеющее лебеговскую меру $\beta_P(x, t)$; при этом для слов $x \neq y$ одинаковой длины множества, соответствующие (x, t_1) и (y, t_2), не пересекаются ни при каких t_1 и t_2; для слов $x \subset y$ при каждом t множество, соответствующее (x, t), включает в себя множество, соответствующее (y, t); для $t_1 < t_2$ при каждом x множество, соответствующее (x, t_2), включает в себя множество, соответствующее (x, t_1).

Процесс F действует так: по слову z он строит наши множества для всех пар (x, t) таких, что $l(x) \leqslant l(z)$ и $t \leqslant l(z)$, и выдает слово x наибольшей длины такое, что z принадлежит множеству, соответствующему (x, t) для какого-то

t (очевидно, такое x только одно, так как множества, соответствующие разным x, не пересекаются, и $x' \subset x''$ для $z' \subset z''$).

3. Универсальная полувычислимая мера. Теорема 3.3 (Л. А. Левин). *Существует универсальная полувычислимая мера R, т. е. полувычислимая мера, удовлетворяющая следующему условию: для любой полувычислимой меры Q найдется константа C такая, что*

$$C \cdot R\{\Gamma_x\} \geqslant Q\{\Gamma_x\} \tag{3.12}$$

для любого x [1]).

Доказательство. Согласно замечанию 3.1 существует универсальный процесс $H(i, x)$. Положим

$$F(z) = H(\pi_1(z), \pi_2(z)). \tag{3.13}$$

Легко показать, что $F(z)$ — процесс (см. (3.2)). Этот процесс, примененный к последовательностям, распределенным равномерно, индуцирует искомую меру. Действительно, пусть процесс $G(y)$ переводит некоторое множество последовательностей в множество Γ_x. Тогда процесс $F(z)$ переводит в Γ_x те же последовательности с приписанным к ним слева словом $\bar{i}$, где i — номер процесса G [т. е. $H(i, x) = G(x)$ для всех x], и, может быть, некоторые другие последовательности. Поэтому мера не может уменьшиться более чем в C раз, где в качестве C можно взять $C = 2^{l(\bar{i})}$.

Замечание 3.8. Аналогичный результат для вычислимых мер не имеет места: среди всех вычислимых мер не существует универсальной. Этот факт является одним из поводов введения понятия полувычислимой меры.

Мера R оказывается (если пренебречь мультипликативной константой) «больше» любой другой меры и сосредоточена на самом широком подмножестве Ω^*. Математическая статистика ставит задачу: выяснить, по какой мере может «случайно» получиться данная последовательность. При этом, если о свойствах последовательности заранее ничего не известно, то единственное (самое слабое) утверждение, которое мы можем сделать относительно нее,— это то, что она может случайно получиться по мере R. Таким образом, мера R соответствует тому, что мы интуитивно понимаем под словами «априорная вероятность». Однако попытка применить это понятие для обоснования математической статистики наталкивается на трудности, связанные с тем, что мера R невычислима.

Представляет интерес следующий факт:

а) существует константа C такая, что вероятность (по мере R) выпадения единицы после n нулей не меньше $\frac{1}{n} \cdot \frac{1}{C \log_2^2 n}$;

б) для любой константы C доля тех n, для которых вероятность (по мере R) выпадения единицы после n нулей больше $\frac{1}{n} C \log_2^2 n$, не превосходит $1/C$ на любом достаточно большом отрезке от 0 до N.

---

[1]) Иными словами, Q абсолютно непрерывна относительно R, причем производная Радона — Никодима ограничена константой C.

Таким образом, эта вероятность имеет порядок примерно $1/n$ [1]).

Доказательство этого утверждения легко вытекает из (3.14), если учесть, что сложность разрешения слова, состоящего из n нулей и единицы, не превосходит $\log_2 n$, причем для большинства таких слов почти равна $\log_2 n$.

Можно проследить аналогию между построением сложности и универсальной полувычислимой меры. Оказывается, эти величины имеют и численную связь.

Теорема 3.4 (Л. А. Левин).

$$|KR(x)-(-\log_2 R\{\Gamma_x\})| \preccurlyeq 2\log_2 KR(x). \tag{3.14}$$

Доказательство. Пусть $KR(x)=i$, т. е. существует слово p, $l(p)=i$, такое, что $G_0^2(p, n)=x_n$ для всякого $n\leqslant l(x)$ (здесь G_0^2 из теоремы 2.1). Тогда легко построить процесс, который любую последовательность, начинающуюся на слово $\overline{l(p)}\,p$, переводит в последовательность, начинающуюся на слово x: он должен сначала выделить $\overline{l(p)}$; по нему восстановить $l(p)$, зная $l(p)$, «прочитать» само слово p; начать приписывать друг к другу $G_0^2(p, n)$ для $n=1, 2, \ldots$ Если применять этот процесс к последовательностям, распределенным равномерно, то индуцированная мера множества Γ_x будет не меньше, чем $2^{-l(\overline{l(p)}p)}$. Поэтому согласно теореме 3.3

$$R\{\Gamma_x\}\geqslant C\cdot 2^{-l(\overline{l(p)}p)},$$

откуда

$$-\log_2 R\{\Gamma_x\} \preccurlyeq l(\overline{l(p)}\,p)=l(\overline{l(p)})+l(p)\asymp l(p)+2l(l(p))= \tag{3.15}$$
$$=i+2l(i)=KR(x)+2l(KR(x)).$$

Пусть теперь $R\{\Gamma_x\}=q$. Обозначим $l(q)=[-\log_2 q]$ [2]). Оценим сложность разрешения слова x; для этого мы покажем, что любой знак слова x можно восстановить по информации, задаваемой тройкой слов $\overline{l(q)}$, k и i (или, что то же самое, одним словом $\overline{l(q)}\,\overline{k}i$), где $k=0$ или 1, а $i\leqslant 2^{l(q)+1}$. Наш алгоритм будут действовать так; по слову $l(q)$ он начнет выстраивать дерево (см. рис. 1) слов y таких, что $R\{\Gamma_y\}>2^{-l(q)-1}$ (для этого нужно вычислять $\beta_R(y, t)$ для все бо́льших значений t и y и подстраивать слово y к дереву, как только для какого-то t станет $\beta_R(y, t)>2^{-l(q)-1}$). Слово x принадлежит этой совокупности. На каждом шагу алгоритма мы будем выделять в уже выстроенной части дерева совокупность «максимальных» слов, т. е. слов, которые пока не имеют продолжения в уже выстроенной части дерева. Ясно, что количество максимальных слов от шага к шагу будет не убывать, оставаясь меньше $2^{l(q)+1}$. Пусть точка A (см. рис. 4, где изображено разрастание дерева слов, имеющих достаточно большую меру R; сплошной линией изображено дерево в момент, когда количество ветвей

---

[1]) Заметим, что это утверждение относится только к универсальной (априорной) вероятности. Например, если известно, что Солнце всходило 10 000 лет, то это еще не означает, что вероятность того, что оно завтра не взойдет, равна примерно $\frac{1}{3\,650\,000}$. Это было бы верно, если бы наша информация о Солнце исчерпывалась указанным фактом.

[2]) Здесь квадратными скобками обозначается целая часть действительного числа.

впервые стало равно i (в этот момент разветвление происходит в точке A); пунктиром изображено дерево, выстроенное настолько, что все знаки слова x уже разрешены) есть точка, от которой отходит последнее «побочное ответвление» от слова x: дальше слово x идет без ответвлений. Для разрешения слова x нам достаточно, во-первых, задать k, равное 0 или 1, в соответствии с тем, идет слово x «влево» от точки A или «вправо», и, во-вторых, задать какую-нибудь информацию, по которой алгоритм мог бы «найти» точку A. В качестве этой информации мы зададим число i — количество максимальных слов в тот момент, когда впервые от точки A будут отходить (в уже построенной части дерева) оба ростка (как раз, когда мы пристроим второй в порядке получения росток, количество максимальных слов увеличится на 1 и станет равным i), при этом $i \leqslant 2^{l(q)+1}$; т. е. $l(i) \leqslant l(q)+1$. В итоге

$$KR(x) \lesssim l(\overline{l(q)}\,\bar{k}i) \asymp 2l(l(q))+ \\ +l(i) \lesssim 2l(l(q))+l(q) \asymp \\ \asymp -\log_2 R\{\Gamma_x\}+2\log_2(-\log_2 R\{\Gamma_x\}).$$

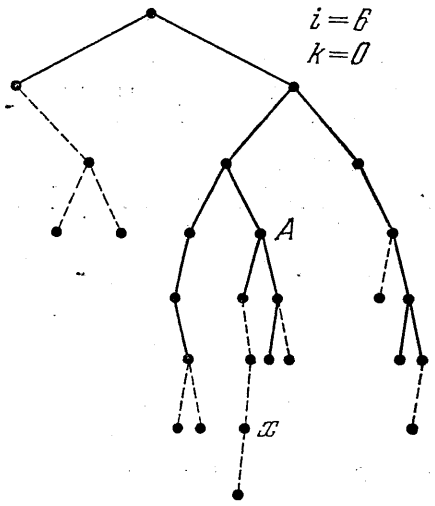

Рис. 4.

Но согласно (3.15)

$$2\log_2(-\log_2 R\{\Gamma_x\}) \lesssim 2\log_2[KR(x)+ \\ +2l(KR(x))] \lesssim 2\log_2 KR(x),$$

откуда

(3.16) $KR(x) \lesssim -\log_2 R\{\Gamma_x\}+2\log_2 KR(x)$;

(3.15) и (3.16) вместе дают (3.14).

Интересно заметить, что из обычных соображений теории меры вытекает, что любая (не обязательно полувычислимая) мера P почти вся сосредоточена на множестве таких ω, что $\exists C \forall n$

$$P\{\Gamma_{(\omega)n}\} \geqslant C \cdot R\{\Gamma_{(\omega)n}\}. \tag{3.17}$$

Точно так же для R-почти всех последовательностей имеет место неравенство в обратную сторону; если P абсолютно непрерывна относительно R, то оно выполняется для P-почти всех последовательностей. Отсюда вытекает, что факт, аналогичный теореме 3.4, имеет место для произвольной полувычислимой меры P на фрагментах P-почти любой последовательности (конечно, константа для каждой последовательности своя).

В качестве следствия теоремы 3.4 получаем известную теорему де Леу — Мура — Шеннона — Шапиро о вероятностных машинах.

Следствие 3.1. *Последовательность* ω *имеет положительную вероятность по некоторой (а следовательно, и по универсальной) полувычислимой мере и только тогда, когда она (т. е.* ω*) вычислима.*

Доказательство. Из (3.14) следует, что мера R всех фрагментов ω больше некоторого положительного числа тогда и только тогда, когда сложность их разрешения ограничена.

4. Вероятностные машины. Предыдущий результат Шеннона иногда интерпретируется как невозможность решения с помощью вероятностных машин задач, недоступ-

ных детерминированным машинам. Однако не всегда задача состоит в том, чтобы построить некий конкретный однозначно определенный объект; иногда у задачи может существовать много решений, и нам требуется построить лишь какое-нибудь из них. В такой постановке, очевидно, существуют задачи, которые недоступны детерминированным машинам, но могут быть решены с помощью машин, использующих датчик случайных чисел (примером может служить задача: построить какую-нибудь невычислимую последовательность).

Будем говорить, что задача построения (какой-нибудь) последовательности, обладающей свойством П, *разрешима с помощью вероятностной машины*, если универсальная мера R таких последовательностей больше нуля. Следующее предложение показывает, что такие задачи можно решать со сколь угодно большой надежностью.

Предложение 3.1 (Л. А. Левин). *Пусть* $A \subseteq \Omega$, $R\{A\} > 0$. *Тогда для любого* $\varepsilon > 0$ *существует слаботабличный (быстрорастущий), со скоростью роста* $l(F(x)) \geqslant l(x)$, *процесс, который, будучи применен к последовательностям, распределенным по мере* L, *выдает последовательности из множества* A *с вероятностью не меньше* $1 - \varepsilon$ [1].

Очевидно, что, например, задачу получения какой-нибудь максимально сложной последовательности решить процессом, растущим быстрее, нельзя, так как при применении процесса сложность слов (а точнее, близкая к ней величина $[-\log_2 R\{\Gamma_x\}]$) не может возрастать. Оказывается, что, когда это соображение не существенно, процесс можно существенно ускорить (т. е. получать результат, используя меньшее число знаков датчика случайных чисел).

Предложение 3.2 (Л. А. Левин). *Пусть* g — *произвольная общерекурсивная функция. Задача получения последовательности из множества* A *тогда и только тогда разрешима с помощью процесса, растущего со скоростью* $g(n)$, *когда существует множество* $B \subseteq A$, $R\{B\} > 0$, *такое, что* $-\log_2 R\{\Gamma_x\} \preccurlyeq n$ *для любой последовательности* $\omega \in B$, *где* $x = (\omega)_{g(n)}$.

Приведем без доказательства некоторые результаты, касающиеся возможностей решения на вероятностных машинах стандартных алгоритмических задач. Первый интересный результат такого характера был получен Я. М. Барздинем. Будем называть бесконечное множество натуральных чисел *иммунным*, если оно не содержит никакого бесконечного перечислимого подмножества.

Предложение 3.3 (Я. М. Барздинь). *Существует иммунное множество (например, дополнение к множеству* A *теоремы* 1.6) *такое, что задача получения последовательности, характеристической для некоторого его бесконечного подмножества, разрешима с помощью вероятностной машины.*

Доказательство этого предложения легко получить из теоремы 1.6 и следствия 4.1.

Интересной разновидностью иммунных множеств являются множества, иммунность которых обусловлена слишком быстрым ростом функции, задающей по i их i-й по величине элемент; такие множества называются *гипериммунными* (точнее, гипериммунным называется такое множество натуральных чисел, что не существует общерекурсивной функции $F(i) > x_i$, где x_i — i-й по величине элемент множества).

Предложение 3.4 (В. Н. Агафонов, Л. А. Левин). *Каково бы ни было (фиксированное) гипериммунное множество* M, *задача получения последовательности, характеристической для какого-нибудь его бесконечного подмножества, неразрешима с помощью вероятностной машины.*

Однако имеет место

Предложение 3.5 (Н. В. Петри). *Задача получения последовательности, обладающей тем свойством, что множество, для которого она характеристическая, гипериммунно, разрешима с помощью вероятностной машины.*

Подробнее о вероятностных машинах см. [23], [25].

---

[1]) Заметим, что, во-первых, построение этого процесса по ε не всегда эффективно, во-вторых, как показал Н. В. Петри, если ограничиться общерекурсивными процессами (а не частично рекурсивными), то не всякую разрешимую задачу можно решить быстрорастущим процессом.

§ 4. Случайные последовательности

1. Определения. Универсальный тест. Аксиоматическое построение теории вероятностей на базе теории меры [26] как чисто математической дисциплины, логически безупречно и ни у кого не вызывает сомнений. Однако для того, чтобы можно было с полным основанием применять эту теорию к практике, нужно четко сформулировать ее физическую интерпретацию. До недавних пор не удавалось найти удовлетворительного решения этой проблемы. Действительно, вероятность обычно истолковывают при помощи таких рассуждений: «Если мы будем проделывать много испытаний, то отношение количества благоприятных исходов к количеству проделанных испытаний *всегда* даст число близкое, а в пределе в точности равное вероятности (или мере) рассматриваемого события». Однако сказать «всегда» здесь было бы неверно: строго говоря, это происходит не всегда, а лишь с вероятностью 1 (а для конечных серий испытаний с вероятностью, близкой к 1). Тем самым понятие вероятности произвольного события определяется через понятие события, имеющего вероятность близкую (а в пределе равную) единице, которое, следовательно, уже нельзя определить таким способом без явного логического круга.

В 1919 г. Р. Мизесом был предложен следующий путь устранения этих трудностей: по Мизесу, последовательности бывают случайными и неслучайными [1]). С математической точки зрения случайные последовательности образуют множество полной меры и *все без исключения* удовлетворяют всем законам теории вероятностей. Физически же можно считать установленным, что в результате эксперимента появляются только случайные последовательности.

Однако предложенное Р. Мизесом [27] и уточненное впоследствии А. Вальдом [28], А. Черчем [29] и А. Н. Колмогоровым [31] определение случайной последовательности оказалось неудовлетворительным. Так, например, было доказано существование случайных, по Мизесу, последовательностей (так называемых *коллективов* Мизеса), не удовлетворяющих закону повторного логарифма [30].

В 1965 г. П. Мартин-Лёфу удалось, основываясь на идеях А. Н. Колмогорова, дать свободное от подобных трудностей определение случайной последовательности. Идея А. Н. Колмогорова состояла в том, чтобы «не случайными» считать те последовательности, в которых наблюдается достаточно много закономерностей, где под закономерностью подразумевается *любое проверяемое свойство последовательности, присущее лишь узкому их классу* (достаточно малому по мере). Если измерять «количество закономерностей» по традиционной логарифмической (с основанием 2) шкале Шеннона, то последняя фраза уточняется следующим образом: мера множества тех последовательностей, в которых может быть обнаружено более чем на m бит закономерностей, не должна превосходить 2^{-m}.

---

[1]) Мы построим теорию в простейшем случае, для пространства Ω бесконечных двоичных последовательностей, однако ее легко обобщить (см. мелкий шрифт на стр. 114—115).

Для описания класса случайных последовательностей выбор шкалы не существен, и 2^{-m} можно заменить на $1/f(m)$, где $f(m)$ — произвольная общерекурсивная монотонная неограниченная функция. Однако выбор логарифмической шкалы не случаен. Вопрос о выборе шкалы есть вопрос о точности измерения количества закономерностей. Но при более подробной шкале измерять количество закономерностей нельзя было бы без явного произвола, так как только при логарифмической шкале теорема о существовании универсального теста (теорема 4.1) выполняется с точностью до аддитивной константы, выбор же менее подробной шкалы привел бы к неоправданной потере точности.

Замечание 4.1. Особо подчеркиваем, что под закономерностями мы понимаем не любые редкие свойства последовательностей, а только проверяемые, т. е. мы будем считать случайными те последовательности, которые при любой алгоритмической проверке и в любом алгоритмическом эксперименте ведут себя как случайные.

Все высказанные соображения приводят нас к следующему определению.

Определение 4.1 (П. Мартин-Лёф). Корректным методом обнаружения P-закономерности (где P — некоторая вероятностная мера на Ω), или *P-тестом* назовем функцию $F(x)$, которая удовлетворяет следующим условиям:

а) она общерекурсивна;

б) для $m > 0$

$$P\{\omega: F(\omega) \geqslant m\} \leqslant 2^{-m}, \tag{4.1}$$

где

$$F(\omega) = \sup_n F((\omega)_n). \tag{4.2}$$

Значением теста является «количество» найденных им закономерностей. Будем говорить, что последовательность ω *не выдерживает* P-тест F, или что P-тест F *отбрасывает* ω, если $F(\omega) = \infty$.

Смысл п. а) определения 4.1 обусловлен замечанием 4.1. В некоторых работах рассматриваются тесты, для которых условие вычислимости заменено более слабым условием формулируемости в некоторой теории, т. е. эти тесты констатируют и закономерности, которые нельзя обнаружить, но можно как-то описать. Условие (4.1) гарантирует, что множество последовательностей, отбрасываемых P-тестом, будет иметь P-меру нуль. Верно и обратное: для любого множества P-меры нуль существует функция, обладающая свойством (4.1) (не обязательно вычислимая), которая отбрасывает все последовательности из этого множества.

Тесты могут быть очень разнообразны, однако, как и в случае, когда мы измеряли сложность по разным частично рекурсивным функциям, имеет место теорема о существовании *универсального* теста.

Теорема 4.1 (П. Мартин-Лёф). *Для любой вычислимой меры P существует P-тест F (называемый универсальным) такой, что для любого P-теста G найдется константа C такая, что для всех $\omega \in \Omega$*

$$G(\omega) \leqslant F(\omega) + C. \tag{4.3}$$

Доказательство. Сначала построим общерекурсивную функцию $H^2(i, x)$ такую, что $H^2(i_0, x)$ при любом фиксированном i_0 есть P-тест и для любого P-теста G найдется i_0 такое, что $H^2(i_0, \omega) \geqslant G(\omega) \dot{-} 1$ при всех $\omega \in \Omega$. Для этого возьмем универсальную частично рекурсивную функцию $U^2(i, x)$ (см. определение 0.2) и при каждом i_0 преобразуем ее так, чтобы она стала P-тестом, причем, если $U^2(i_0, x) + 1$ уже была тестом, чтобы супремумы на $\omega \in \Omega$ функции $U^2(i_0, x)$ не изменились.

Зафиксируем i_0. Возьмем все фрагменты y слова x и на каждом фрагменте проделаем $l(x)$ шагов алгоритма, вычисляющего $U^2(i_0, y)$; положим $G_x(i_0, y)$ равным результату действия этого алгоритма, если результат уже получен, и $G_x(i_0, y) = 0$ в противном случае. Пусть $G(i_0, x) = \sup\limits_{y \subset x} G_x(i_0, y)$. Очевидно, что функция $G(i_0, x)$ общерекурсивна, причем для любого $\omega \in \Omega$

$$U^2(i_0, \omega) = G(i_0, \omega). \tag{4.4}$$

Однако $G(i_0, x)$ может не удовлетворять условию (4.1). Для того чтобы добиться его выполнения, заменим $G(i_0, x)$ на

$$H^2(i_0, x) = \min\{G(i_0, x); M(i_0, x)\}, \tag{4.5}$$

где $M(i_0, x)$ — минимальное m такое, что при $m+1$ для функции $G(i_0, x)$ не выполняется условие (4.1) «с запасом» на точность вычисления меры, т. е.

$$M(i_0, x) = \min\Big\{m: \sum_{\substack{y: l(y)=l(x) \\ G(i_0, y) \geqslant m+1}} \alpha_P(y, l(x)+m+2) > 2^{-(m+1)}\Big\}, \tag{4.6}$$

причем проверка ведется до $m = \max\limits_{y: l(y)=l(x)} G(i_0, y)$ (сумма в формуле (4.6) приближает меру $P\{\bigcup \Gamma_y: l(y) = l(x), G(i_0, y) \geqslant m+1\}$ с избытком с точностью $2^{-(m+2)}$). Функция $H^2(i_0, x)$ по построению удовлетворяет условию (4.1). Более того, если функция $G(i_0, x)+1$ удовлетворяла условию (4.1), то $P\{\bigcup \Gamma_y: l(y) = l(x), G(i_0, y) \geqslant m+1\} \leqslant 2^{-(m+2)}$ (так как на тех же y $G(i_0, y)+1 \geqslant m+2$) и, следовательно, неравенство в формуле (4.6) не может выполниться ни для какого m, т. е $G(i_0, x) \equiv H^2(i_0, x)$. Функция $H^2(i, x)$ построена.

Покажем, что функция

$$F(x) = \max_{i \leqslant l(x)} [H^2(i, x) \dot{-} (i+1)] \tag{4.7}$$

является универсальным тестом. Выполнение для нее условия (4.1) следует из включения $\{x: l(x) = n, F(x) \geqslant m\} \subseteq \bigcup\limits_{i=1}^{n} \{x: l(x) = n, H^2(i, x) \dot{-} (i+1) \geqslant m\}$ и выполнения условия (4.1) для функции $H^2(i, x)$. Наконец, если $G(x)$ — P-тест, и i_0 — его номер (см. определение 0.2), то по построению $H^2(i_0, x) \geqslant G(i_0, x) \dot{-} 1$, следовательно, для слов x длины не меньше i_0 выполняется неравенство $F(x) \geqslant G(i_0, x) \dot{-} (i_0+2)$, откуда для всех $\omega \in \Omega$

$$F(\omega) \geqslant G(i_0, \omega) \dot{-} (i_0+2). \tag{4.8}$$

Сопоставляя (4.4) и (4.8), получаем (4.3).

Определение 4.2 (П. Мартин-Лёф). Будем называть последовательность ω *случайной* по мере P, если она выдерживает любой P-тест.

При таком определении *все без исключения* случайные последовательности удовлетворяют всем мыслимым эффективно проверяемым законам теории вероятностей (под законом понимается утверждение, что некоторое событие происходит с вероятностью 1; примерами законов могут служить усиленный закон больших чисел и закон повторного логарифма для последовательностей независимых испытаний, свойство возвратности марковских цепей, и так далее), так как по любому такому закону можно устроить тест, отбрасывающий все те последовательности, для которых этот закон не выполняется (иными словами, невыполнение закона есть закономерность).

Замечание 4.2. Согласно только что доказанной теореме в случае вычислимой меры P случайность последовательности ω эквивалентна тому, что ω выдерживает универсальный P-тест. Таким образом, для любой вычислимой меры неслучайность последовательности можно эффективно установить.

Замечание 4.3. В дальнейшем нам будет удобно пользоваться «монотонным» универсальным тестом, т. е. таким, что для $x \subset y$ $F(x) \leqslant F(y)$. Его легко получить из построенного, положив

$$F'(x) = \max_{y \subset x} F(x) \text{ [1]}. \tag{4.9}$$

В дальнейшем мы всегда будем предполагать, что универсальный тест именно таков.

Выше мы ввели понятие теста случайности, случайной последовательности, универсального теста (и доказали теорему о его существовании в случае вычислимых мер) для объектов простейшего вида — элементов Ω. Однако те же построения П. Мартин-Лёфа можно провести и в более общем случае. Пусть T — топологическое пространство со счетной базой открытых множеств x_i ($i = 1, 2, \ldots$), а P—мера на σ-алгебре борелевских подмножеств T. Нам будет удобно предполагать, что элементы базы занумерованы таким образом, чтобы по любому номеру n некоторого элемента x базы можно было бы эффективно найти последовательность чисел, бо́льших n, которые являются номерами элементов базы, дающих в объединении x (например, найти другой номер элемента x, бо́льший n) [2]. Очевидно, такое условие не ограничивает общности, так как любую нумерацию можно переделать в нумерацию, обладающую этим свойством, присвоив каждому элементу вместо номера i (в старой нумерации) номера $(2i + 1) \cdot 2^k$ ($k = 1, 2, \ldots$) (в новой нумерации). Будем говорить, что задан элемент $\omega \in T$, если задана последовательность, не обязательно монотонная, всех номеров i таких, что $\omega \in x_i$. *P-тестом* будем называть общерекурсивную функцию $F(n)$ такую, что

$$P\{\bigcup_{n:\, F(n) \geqslant m} x_n\} \leqslant 2^{-m}. \tag{4.10}$$

Значением теста F на элементе $\omega \in T$ будем называть

$$F(\omega) = \sup_{n:\, \omega \in x_n} F(n); \tag{4.11}$$

меру P будем называть *вычислимой*, если есть алгоритм, по любому конечному набору чисел i_k и числу n вычисляющий меру $P\{\bigcup_k x_{i_k}\}$ с точностью до 2^{-n}.

Предложение 4.1. *Для любой вычислимой меры P существует универсальный P-тест (определение универсального P-теста то же).*

---

[1]) При этом, очевидно, условие (4.1) не нарушится, так как

$$\sup_n F'((\omega)_n) = \sup_n F((\omega)_n).$$

[2]) Отсутствие этого условия потребовало бы более громоздкого определения теста.

Определение 4.3. Элемент $\omega \in T$ будем называть *P-случайным*, если $F(\omega) < \infty$ для любого P-теста F.

Очевидно, что для вычислимых мер это эквивалентно тому, что значение универсального теста на ω конечно.

Таким образом, понятие случайного объекта носит очень общий характер. Интересными примерами являются понятия случайного вектора, случайного элемента какого-нибудь функционального пространства (случайного процесса) и другие.

Будем говорить, что две базы *эквивалентны*, если по любому номеру i элемента x одной из баз можно эффективно получить последовательность n_k номеров элементов другой базы, дающих в объединении x.

Предложение 4.2. *Свойство элемента $\omega \in T$ быть P-случайным (мера P не обязательна вычислима) инвариантно относительно перехода к эквивалентной базе.*

Замечание 4.4. Если мы перейдем к другой нумерации, невычислимо связанной с исходной, то мы можем получить неэквивалентную базу. При этом класс случайных элементов может изменяться. Пример: пусть $\gamma \in \Omega$ — случайная последовательность; перенумеруем двоичные слова x (они соответствуют элементам Γ_x базы) так, чтобы множество R номеров фрагментов γ стало разрешимым, а их длина осталась вычислимой функцией номера. Очевидно, что тест $F(n)$, равный длине слова x, если его номер $n \in R$, и нулю в противном случае, отбрасывает последовательность γ. Этот пример показывает, что совокупность случайных элементов зависит не только от топологической структуры пространства T (если бы это было так, то, например, элементы топологически однородного пространства Ω были бы либо все случайны, либо все неслучайны), но и от других структур (например, связанных с системой координат).

2. Правильные последовательности. Теорема 4.2 (Л. А. Левин). а) *Какова бы ни была вычислимая мера P, любой P-регулярный процесс применим ко всем P-случайным последовательностям.*

б) *Если P — произвольная мера (не обязательно вычислимая), F — процесс, Q — мера, индуцированная им с меры P, и ω — P-случайная последовательность, к которой F применим, то его результат $F(\omega)$ Q-случаен.*

Доказательство. а) Пусть P-регулярный процесс F неприменим к последовательности γ, т. е. существует число (обозначим его k) такое, что длина $F(\gamma)$ не превосходит k. Это свойство нашей последовательности уникально, так как P-мера таких последовательностей равна 0, поэтому легко построить P-тест, который отбрасывает все последовательности, результат применения к которым процесса F имеет длину не больше k. Этот тест на слове x делает следующее: он выбирает максимально длинный фрагмент $(x)_m$ такой, что $l(F((x)_m)) \leqslant k$, затем вычисляет приближение (с избытком с точностью до $2^{-l(x)}$) меры тех последовательностей ω, для которых $l(F((\omega)_m)) \leqslant k$, и выдает в качестве своего значения на x целую часть от минус логарифма этой меры. Очевидно, что на словах $(\gamma)_n$ значение теста стремится к бесконечности. Выполнение условий а) — б) определения 4.1 легко проверит читатель.

б) Пусть Q-тест $U(x)$ отбрасывает последовательность $\rho = F(\omega)$, и пусть G — общерекурсивный процесс, эквивалентный F (см. замечание 3.2). Тогда P-тест

$$V(x) = U(G(x)) \tag{4.12}$$

(условия а) — б) определения 4.1 легко проверяются) отбрасывает последовательность ω, т. е. ω не P-случайна.

Определение 4.4. Назовем последовательность *правильной*, если она случайна по некоторой вычислимой мере.

Все последовательности с разрешимым S_ω правильны. Примером последовательности, не являющейся правильной, будет, как легко показать, последовательность ω, у которой S_ω — область определения универсальной частично рекурсивной функции.

Т е о р е м а 4.3 (Л. А. Левин). *Любая правильная последовательность либо вычислима, либо алгоритмически эквивалентна*[1]) *некоторой L-случайной последовательности.*

Д о к а з а т е л ь с т в о. Пусть Q — вычислимая мера, по которой случайна наша правильная последовательность ω. Покажем прежде всего, что ω не может принадлежать отрезку $[\tau', \tau'']$ Q-меры 0 (см. рис. 3), точнее, на всем отрезке $[\tau', \tau'']$ нет ни одной Q-случайной последовательности. Для этого построим Q-тест, отбрасывающий все последовательности из этого отрезка. Пусть α — рациональное число внутри отрезка. На слове x длины n наш тест принимает в качестве значения наибольшее число m такое, что

$$\sum \alpha_Q(y,\ 2n) < 2^{-(m+1)}, \tag{4.13}$$

где сумма берется по всем словам y длины n, лежащим между словами x и $(\alpha)_n$ включительно. Условия а) — б) определения 4.1 проверяются тривиально, и этот тест, очевидно, искомый, так как для любой последовательности β из отрезка $[\tau', \tau'']$ Q-мера всех последовательностей, заключенных между α и β, равна 0, а сумма слева в (4.13) есть приближение этой меры (с избытком) с точностью 2^{-n}, т. е. эта сумма стремится к 0 при $n \to \infty$, и, следовательно, значение теста на $(\beta)_n$ стремится к бесконечности при $n \to \infty$.

Если ω невычислима, то, так как она не лежит на отрезке типа $[\tau', \tau'']$, к ней применим процесс G, обратный к процессу F, индуцирующему меру Q из меры L (см. теорему 3.1б); обозначим $G(\omega) = \delta$. Процесс F применим к δ, так как он может быть неприменим только к последовательностям, которые должны отображаться в двоично-рациональные точки (а ω не двоично-рациональна, так как она невычислима) и к последовательностям типа ρ, в которые функция g переводит целый отрезок (см. теорему 3.1б и рис. 3), а $\omega \in [\tau', \tau'']$. Итак, последовательности ω и δ алгоритмически эквивалентны.

Осталось показать, что последовательность δ L-случайна. Пусть U — универсальный L-тест. Тогда Q-тест $V(x) = U(G(x))$ (условия а) — б) определения 4.1 легко проверяются) будет отбрасывать все последовательности, G-результаты которых не L-случайны; в том числе, если δ не L-случайна, он отбросит и ω, т. е. ω не Q-случайна, что противоречит предположению.

3. Универсальный тест и сложность. Как показывает теорема 4.3, изучение последовательностей, случайных по произвольной вычислимой мере, сводится к изучению последовательностей, случайных по равномерной мере. Такие последовательности мы будем называть просто *случайными*.

Универсальный тест, работая над все более длинными фрагментами последовательности, со временем обнаружит все имеющиеся там закономерности. Однако, поскольку универсальность теста проявляется только в пределе, он будет находить некоторые закономерности, сосредоточенные в начальном

---

[1]) Последовательности ω' и ω'' называются *алгоритмически эквивалентными*, если существуют два процесса F и G такие, что $F(\omega') = \omega''$ и $G(\omega'') = \omega'$.

фрагменте последовательности, только рассматривая более длинный фрагмент. Тогда на некотором слове x тест будет принимать маленькое значение k, а на *любых* достаточно длинных продолжениях его — значение $n > k$. Понятно, что в этом случае все эти n бит закономерностей определяются самим словом x и сосредоточены в нем. Обозначим $F(x, n)$ минимум значений универсального теста на словах длины n, начинающихся со слова x. Устремляя n к бесконечности, мы получим количество всех закономерностей, имеющихся в слове x:

(4.14) $$p(x) = \lim_{n\to\infty} F(x, n)\ ^{1})$$

(согласно замечанию 4.3 $\lim\limits_{n\to\infty} F(x, n) = \sup\limits_{n\to\infty} F(x, n)$). Очевидно (из условия (4.1)), что всегда $p(x) \leqslant l(x)$. Величина $l(x) - p(x)$, т. е. количество знаков в слове минус количество закономерностей в нем (из-за наличия которых в записи слова имеются паразитические знаки) по своим свойствам очень похожа на сложность. Она невычислима (так как невычислима $p(x)$), но ее можно сколь угодно точно оценить сверху (см. замечание 4.3) функцией $l(x) - F(x, n)$ (ср. с теоремой 1.5б). Построение универсального теста похоже на построение оптимальной частично рекурсивной функции; доля слов, на которых $l(x) - p(x)$ принимает значения существенно меньшие, чем $l(x)$, мала, и т. п. Оказывается, величины $l(x) - p(x)$ и $K(x)$ и численно очень близки.

Теорема 4.4 (П. Мартин-Лёф).

(4.15) $$|[l(x) - p(x)] - K(x)| \lesssim 4l(l(x)).$$

Доказательство. Пусть $l(x) - p(x) \leqslant a$, или $p(x) \geqslant l(x) - a$. Это означает, что на всех достаточно длинных продолжениях y слова x $F(y) \geqslant l(x) - a$, где F — универсальный тест. Из условия (4.1) следует, что

$$L\{\bigcup \Gamma_z : l(z) = l(x),\ p(z) \geqslant l(z) - a\} \leqslant L\{\bigcup \Gamma_y : F(y) \geqslant l(x) - a\} \leqslant 2^{-l(x)+a},$$

следовательно, количество таких слов z, что $l(z) = l(x)$ и $p(z) \geqslant l(z) - a$, не превосходит 2^a, причем оно, очевидно, перечислимо без повторений (см. теорему 0.4). Поэтому, чтобы найти x, досаточно задать в качестве информации числа $l(x)$, a и m — номер слова x (в порядке перечисления) среди слов z, $l(z) = l(x)$, $p(z) \geqslant l(z) - a$ (при этом $m \leqslant 2^a$). Ту же информацию можно запиать в одно слово: $\overline{l(x)}\,\bar{a}m$, поэтому

$$K(x) \lesssim l(\overline{l(x)}\,\bar{a}m) \asymp 2l(l(x)) + 2l(a) + l(m) \leqslant a + 4l(l(x))$$

(так как $a \leqslant l(x)$). Это неравенство справедливо для всех $a \geqslant l(x) - p(x)$, поэтому $K(x) \lesssim [l(x) - p(x)] + 4l(l(x))$.

Докажем неравенство в другую сторону. Для этого построим тест, который будет выделять закономерность, состоящую в том, что сложность слова существенно отличается от его длины (это действительно закономерность,

---

[1]) Напомним, что мы рассматриваем только те закономерности, которые могут алгоритмически проявиться. Если $p(x) \leqslant m$, то существует бесконечная последовательность $\omega \in \Gamma_x$, в которой универсальный тест найдет не больше m закономерностей. А так как универсальный тест в пределе находит все закономерности, то других закономерностей в ω, а значит, и в x, нет.

так как таких слов мало — см. теорему 1.4б). Возьмем функцию $H(t, z)$, аппроксимирующую сверху сложность (см. (1.16)). Тогда искомым тестом будет функция

$$G(y) = \max_{i \leqslant l(y)} [i - 2 - 2l(i) + H(l(y), (y)_i)]. \tag{4.16}$$

Эта функция принимает значение большее или равное m только на тех последовательностях ω, для которых есть такое i, что $K((\omega)_i) \leqslant i - 2 - 2l(i) - m$. Из теоремы 1.4б следует, что мера таких последовательностей не превышает

$$\sum_{i=1}^{\infty} 2^{-2l(i)-2-m} \leqslant 2^{-m-1} \sum_{i=1}^{\infty} \frac{1}{i^2} = 2^{-m-1} \frac{\pi^2}{6} \leqslant 2^{-m},$$

т. е. $G(y)$ удовлетворяет условию (4.1). Общерекурсивность G очевидна.

Следствие 4.1. *Для любой случайной последовательности* ω

$$K((\omega)_n) \geqslant n - 4l(n)\ ^{1}). \tag{4.17}$$

Следствие 4.2. *Последовательность с перечислимым* S_ω *не может быть случайной (по мере* L*).*

4. Пример случайной последовательности. Для более сложных множеств дело обстоит иначе.

Теорема 4.5 (П. Мартин-Лёф). *Существует случайная по* L *последовательность с множеством* S_ω *второго ранга по классификации Клини (т. е. задаваемым арифметическим предикатом с двумя кванторами).*

Доказательство. Пусть A — множество слов, имеющих сколь угодно длинное продолжение, на котором универсальный тест F принимает значение, не превосходящее 1. Множество A непусто, так как $L\{\omega\colon F(\omega) \geqslant 2\} = 1/4$. Очевидно, $\Lambda \in A$, и, если $x \in A$, то либо $x0$ либо $x1$ (либо оба вместе) входят в A. Определим последовательность ω по индукции:

$$(\omega)_1 = \begin{cases} 0, & \text{если } 0 \in A, \\ 1, & \text{если } 0 \notin A, \end{cases}$$

$$(\omega)_{n+1} = \begin{cases} (\omega)_n 0, & \text{если } (\omega)_n 0 \in A, \\ (\omega)_n 1, & \text{если } (\omega)_n 0 \notin A \ \text{ (тогда } (\omega)_n 1 \in A). \end{cases}$$

Очевидно, полученная последовательность ω случайна, так как $F(\omega) \leqslant 1$. Покажем, что эта последовательность — второго ранга. Для этого нужно построить разрешимый предикат $R(n, k, z)$ такой, что предикат

$$P(n) \sim \forall k\ \exists z\ R(n, k, z)$$

характеризует множество S_ω. Для построения $R(n, k, z)$ заметим, что $(\omega)_n$ — наименьшее слово среди слов длины n, принадлежащих A. Поэтому искомый предикат $R(n, k, z)$ по определению удовлетворяется, если

---

[1]) Как легко видеть из второй части доказательства теоремы 4.4, функцию $4l(n)$ можно заменить на $2l(n)$, и вообще, на любую функцию $F(n)$ такую, что ряд $\sum_{n=1}^{\infty} 2^{-F(n)}$ вычислимо быстро сходится (например, на $l(n) + 2l(l(n))$. См. также сноску на стр. 101.

1) $z = \bar{x}\bar{u}l$, где x, u, l — слова, удовлетворяющие следующим условиям:

2) $l(x) = n$; последняя цифра x есть 1;

3) $x \subset u$, $F(u) \leqslant 1$, где F — универсальный тест;

4) $l > n$, и для всех пар слов y, v длины соответственно n и l, и таких, что $y \subset v$ и $y < x$, выполняется неравенство $F(v) > 1$.

Теорема 4.5 и следствие 4.1 уточняют утверждение о существовании максимально сложных последовательностей второго ранга, приведенное на стр. 101. Конечно, тот факт, что последовательность характеризуется предикатом с двумя кванторами, можно рассматривать как закономерность. Однако эту закономерность никак нельзя обнаружить, и во всех алгоритмических экспериментах эта последовательность будет неотличима от остальной массы случайных.

§ 5. Понятие количества информации

1. Определение и простейшие свойства. Сложность $K(x)$ интуитивно обозначает количество информации, необходимое для восстановления текста x. Условная сложность $K(x|y)$ интуитивно обозначает количество информации, которое необходимо добавить к информации, содержащейся в тексте y, чтобы можно было восстановить текст x. Разность между этими величинами естественно назвать количеством информации в y об x.

Определение 5.1. (А. Н. Колмогоров). *Количество информации в y об x есть*

$$I(y:x) = K(x) - K(x|y). \tag{5.1}$$

Замечание 5.1.

$$I(x:y) \geqslant 0, \tag{5.2}$$

$$|I(x:x) - K(x)| \leqslant 0. \tag{5.3}$$

Доказательство. Докажем соотношение (5.2). Пусть $F^2(p, x) = F_0^1(p)$ (см. (1.9)). Тогда, если $F_0^1(p_0) = y$ и $K(y) = l(p_0)$, то, так как $F^2(p_0, x) = y$, имеем

$$K(y|x) \leqslant K_{F^2}(y|x) = K(y).$$

Докажем теперь (5.3). Пусть $F^2(p, x) = x$. Тогда и $F^2(\Lambda, x) = x$, откуда $K(x|x) \leqslant K_{F^2}(x|x) = l(\Lambda) = 0$. Замечая, что $I(x:x) = K(x) - K(x|x)$, получаем требуемое.

Следующая теорема устанавливает связь между определениями количества информации по А. Н. Колмогорову и по К. Шеннону (точнее, между сложностью слова по А. Н. Колмогорову и энтропией распределения частот по К. Шеннону). Оказывается, что энтропия К. Шеннона — просто коэффициент при линейной части одной из частных сложностей.

Теорема 5.1 (А. Н. Колмогоров). *Пусть задано число r и пусть слово x длины $i \cdot r$ состоит из i слов длины r, причем k-е слово длины r входит в x с частотой q_k ($k = 1, 2, \ldots, 2^r$). Тогда*

$$K(x) \leqslant i(H(q_k) + \alpha(i)), \tag{5.4}$$

где

$$H(q_k) = -\sum_{k=1}^{2^r} q_k \log_2 q_k \tag{5.5}$$

и

$$\alpha(i) = C_r \frac{\ln i}{i} \to 0 \text{ при } i \to \infty .$$

В общем случае более тесную связь между энтропией и сложностью установить нельзя. Это и естественно, так как энтропия приспособлена для изучения текстов, не имеющих других закономерностей, кроме частотных, т. е. для последовательностей результатов независимых испытаний. В этом специальном случае между рассматриваемыми величинами можно установить полную связь (это делается в теореме 5.3).

Доказательство теоремы 5.1. Пусть x — m-е по величине слово, состоящее из i слов длины r, входящих в него по s_k раз соответственно $\left(q_k = \frac{s_k}{i}, \sum_{k=1}^{2^r} s_k = i\right)$. Чтобы найти слово x, достаточно задать в качестве информации о нем слова m, s_1, ..., s_{2^r}. Всю эту информацию можно записать одним словом: $p = \bar{s}_1 \bar{s}_2 \ldots \bar{s}_{2^r} m$.

Пусть функция $F^1(p)$ получает из этого слова p слово x. Тогда $K(x) \lesssim 2l(s_1) + \ldots + 2l(s_{2^r}) + l(m)$. Заметим, что m не может превышать количества всех слов, удовлетворяющих условиям, наложенным на x, т. е. $m \leqslant \frac{i!}{s_1! \ldots s_{2^r}!}$. Кроме того, $s_k \leqslant i$. Отсюда

$$K(x) \lesssim 2^{r+1} l(i) + l\left(\frac{i!}{s_1! \ldots s_{2^r}!}\right). \tag{5.6}$$

Используя формулу Стирлинга $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{\theta_n}{12n}}$, где $|\theta_n| \leqslant 1$, для оценки m, получаем (5.4).

2. Коммутативность информации. Классическое шенноновское количество информации в одной случайной величине о другой удовлетворяет условию коммутативности, т. е. $J(\xi : \eta) = J(\eta : \xi)$. Для колмогоровского количества информации в одном тексте о другом точного равенства, вообще говоря, не будет.

Пример 5.1. Согласно замечанию 1.1 для любого l_0 существует слово x длины l_0 такое, что $K(x|l(x)) \geqslant l(x) - 1$.

Согласно теореме 1.4б существуют сколь угодно большие l_0 такие, что $K(l_0) \geqslant l(l_0) - 1$. Для так выбираемых пар слов x и l_0 ($l(x) = l_0$) имеем

$$I(x : l_0) = K(l_0) - K(l_0 | x) \gtrsim l(l_0), \tag{5.7}$$

$$I(l_0 : x) = K(x) - K(x | l_0) \lesssim l_0 - l_0 = 0. \tag{5.8}$$

Таким образом, в некоторых случаях разница между $I(x : y)$ и $I(y : x)$ может иметь порядок логарифма сложностей рассматриваемых слов. Однако, как показали независимо Л. А. Левин и А. Н. Колмогоров, этот порядок является для нее предельным, и, следовательно, если пренебречь величина-

ми, бесконечно малыми по сравнению с информацией, содержащейся в обоих словах, величина $I(x:y)$ будет все же коммутативной.

Теорема 5.2 [1]) (А. Н. Колмогоров, Л. А. Левин).

а) $$|I(x:y)-I(y:x)| \lesssim 12l(K(xy)),$$

б) $$|I(x:y)-[K(x)+K(y)-K(\bar{x}y)]| \lesssim 12l(K(xy)).$$

Доказательство. а) Будем доказывать неравенство только в одну сторону:

(5.9) $$I(x:y) \gtrsim I(y:x)-12l(K(xy)).$$

Обратное неравенство следует из него, если поменять местами x и y.

Построим две вспомогательные функции. Пусть частично рекурсивная функция $F^4(n, b, c, x)$ перечисляет без повторений слова y такие, что $K(y) \leqslant b$, $K(x|y) \leqslant c$. Существование такой функции следует из теоремы 0.4 и теоремы 1.6 (с учетом замечания 1.3). Обозначим через j количество таких y (j невычислимо зависит от x, b, c). Функция F^4 определена для всех $n \leqslant j$ и только для них. Следовательно, предикат $\Pi(b, c, d, x)$, утверждающий, что определенное выше число j превышает 2^d, очевидно, эквивалентен утверждению, что $F^4(2^d, b, c, x)$ определена, и, следовательно, частично рекурсивен. Тогда, аналогично F^4, существует функция $G^5(m, a, b, c, d)$, перечисляющая без повторений все слова x такие, что $K(x) \leqslant a$, $\Pi(b, c, d, x)$. Обозначим через i количество таких слов x (i невычислимо зависит от a, b, c, d). Очевидно, $G^5(m, a, b, c, d)$ определена для всех $m \leqslant i$ и только для них.

Приступим к доказательству. Пусть даны слова x и y, $K(x)=a$, $K(y)=b$, $K(x|y)=c$. Тогда $I(y:x)=a-c$. Далее, как было определено выше, j есть количество слов y' таких, что $K(y') \leqslant b$ и $K(x|y') \leqslant c$ (j зависит от x, b, c), а i есть количество слов x' таких, что $K(x') \leqslant a$ и соответствующее число $j' \geqslant 2^{l(j)}$. Легко видеть, что $i \cdot 2^{l(j)}$ не превосходит количества пар (x', y') таких, что $K(y') \leqslant b$, $K(x'|y') \leqslant c$, которых в свою очередь не больше, чем 2^{b+c+2}, откуда

(5.10) $$l(i)+l(j) \lesssim b+c.$$

Так как слово y будет выдано в качестве значения функции $F^4(n, b, c, x)$ при некотором $n \leqslant j$, то

(5.11) $$K(y|x) \lesssim l(\bar{b}\bar{c}n) \lesssim 2l(b)+2l(c)+l(j).$$

Дальше, так как слово x будет выдано в качестве значения функции $G^5(m, a, b, c, d)$ при $d=l(j)$ и некотором $m \leqslant i$, то

(5.12) $$a=K(x) \lesssim l(\bar{a}\bar{b}\bar{c}\bar{d}m) \lesssim 2l(a)+2l(b)+2l(c)+2l(d)+l(i).$$

Из неравенств (5.10)—(5.12), а также из того, что каждая из величин $l(a)$, $l(b)$, $l(c)$, $l(d)=l(l(j))$ не превосходит $l(K(xy))$, легко получаем $K(y|x) \lesssim b+c-a+12l(K(xy))$, откуда и следует (5.9).

б) Очевидно, что $K(x\bar{x}y) \lesssim K(\bar{x}y)$; отсюда согласно пункту а) настоящей теоремы

$$|I(\bar{x}y:x)-I(x:\bar{x}y)| \lesssim 12l(K(\bar{x}y)),$$

[1]) Проведя оценки более аккуратно, можно их несколько улучшить, например, заменить $12l(K(xy))$ на $(5+\varepsilon)l(K(xy))$. Можно ли довести оценку до $l(K(xy))$, неизвестно

т. е.

$$|K(\bar{x}y) - K(\bar{x}y \mid x) - K(x) + K(x \mid \bar{x}y)| \preccurlyeq 12l(K(\bar{x}y)),$$

или

$$|[K(\bar{x}y) - K(x) - K(y)] + K(y) - K(\bar{x}y \mid x) - K(x \mid \bar{x}y)| \preccurlyeq 12l(K(\bar{x}y)),$$

откуда, замечая, что $K(x \mid \bar{x}y) \succcurlyeq 0$ и $|K(\bar{x}y \mid x) - K(y \mid x)| \succcurlyeq 0$, получаем утверждение б) теоремы 5.2.

3. Независимые испытания. Связь с вероятностным определением информации. Теперь мы можем окончательно выяснить связь между вероятностным и алгоритмическим определениями количества информации. Напомним первое из них в удобной для нас форме (см. [39]). Если ξ — случайная величина, принимающая конечное множество значений x_i с вероятностями q_i, то положим

$$H(\xi) = -\sum_i q_i \log_2 q_i. \tag{5.13}$$

Пусть ξ и ψ — две случайные величины с конечным множеством значений, определенные на одном и том же вероятностном пространстве. Тогда количество информации в ξ о ψ равно

$$J(\xi : \psi) = H(\xi) + H(\psi) - H(\xi, \psi), \tag{5.14}$$

где (ξ, ψ) — случайный вектор. Если ξ и ψ — случайные величины со значениями в Ω (см., однако, петит на стр. 114—115), то положим

$$J(\xi : \psi) = \lim_{n\to\infty} J((\xi)_n, (\psi)_n) \tag{5.15}$$

(заметим, что здесь $\lim\limits_{n\to\infty}$ совпадает с $\sup\limits_n$). Пусть имеются две такие случайные величины, совместно распределенные по мере Q (не обязательно вычислимой). Рассмотрим последовательность независимых случайных векторов $(\xi, \psi)^i$ $(i = 1, 2, \ldots)$, каждый из которых распределен по мере Q. Этими условиями (независимость и одинаковая распределенность по мере Q) однозначно определяется совместное распределение P векторов $(\xi, \psi)^i$. Будем называть последовательностью независимых испытаний случайных величин ξ и ψ последовательность пар бесконечных двоичных последовательностей $(\alpha, \beta)^i$, *случайную по мере* P. Через α_n^i и β_n^i будем обозначать соответственно слова $\overline{(\alpha^1)_n}$ $\overline{(\alpha^2)_n} \ldots \overline{(\alpha^i)_n}$ и $\overline{(\beta^1)_n}$ $\overline{(\beta^2)_n} \ldots \overline{(\beta^i)_n}$.

Теорема 5.3 (А. Н. Колмогоров). *Если* $(\alpha, \beta)^i$ — *последовательность независимых испытаний случайных величин* ξ *и* ψ, *то*

$$\lim_{n\to\infty} \lim_{i\to\infty} \frac{I(\alpha_n^i : \beta_n^i)}{i} = J(\xi : \psi). \tag{5.16}$$

Доказательство. Утверждение теоремы будет вытекать из равенства

$$\lim_{i\to\infty} \frac{I(\alpha_n^i : \beta_n^i)}{i} = J((\xi)_n, (\psi)_n). \tag{5.17}$$

Для доказательства его заметим, что из (5.8) следует

$$\lim_{i\to\infty} \frac{I(\alpha_n^i : \beta_n^i)}{i} = \lim_{i\to\infty} \frac{K(\alpha_n^i)}{i} + \lim_{i\to\infty} \frac{K(\beta_n^i)}{i} - \lim_{i\to\infty} \frac{K(\overline{\alpha_n^i \beta_n^i})}{i}.$$

По определению,

$$J((\xi)_n : (\psi)_n) = H((\xi)_n) + H((\psi)_n) - H((\xi)_n, (\psi)_n).$$

Отсюда ясно, что утверждение теоремы эквивалентно следующему.

Пусть $\theta_1, \theta_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, принимающих в качестве значений двоичные слова длины r с вероятностями q_k, $k \leqslant 2^r$, и пусть γ — двоичная последовательность, разбитая на слова длины r, случайна по мере, соответствующей распределению $\theta_1, \theta_2, \ldots$ Тогда

$$\lim_{i\to\infty} \frac{K((\gamma)_{i\cdot r})}{i} = H(q_k). \tag{5.18}$$

Докажем равенство (5.18). Пусть есть слово x длины $i \cdot r$, состоящее из i слов длины r, входящих в него по s_k раз соответственно $\left(\sum_{k=1}^{2^r} s_k = i\right)$. Набор чисел $s_1, \ldots, s_{2^r}$ будем называть *частотным набором слова* x. Обозначим логарифм количества слов, имеющих тот же частотный набор, что и слово x, через $h(x)$, т. е.

$$h(x) = l\left(\frac{i!}{s_1! \ldots s_{2^r}!}\right).$$

Наша последовательность γ случайна по мере с независимыми испытаниями, причем в каждом испытании результаты получаются с фиксированными вероятностями q_k. Из усиленного закона больших чисел легко по любому $\varepsilon > 0, k \leqslant 2^r$ построить тест, отбрасывающий все последовательности, имеющие бесконечно много фрагментов, у которых s_k/i отклоняется от q_k более чем на ε. А так как γ случайна и, следовательно, выдерживает эти тесты, то у ее фрагментов пределы s_k/i равны в точности q_k. Отсюда и из формулы Стирлинга вытекает, что для фрагментов γ

$$\lim_{i\to\infty} \frac{h((\gamma)_{i\cdot r})}{i} = H(q_k).$$

Покажем, что

$$\lim_{i\to\infty} \frac{h((\gamma)_{i\cdot r}) - K((\gamma)_{i\cdot r})}{i} = 0.$$

Из теоремы 1.3 имеем $K(x \mid \bar{s}_1 \ldots \bar{s}_{2^r}) \lesssim h(x)$, откуда $K(x) \lesssim h(x) + 2^{r+1} \cdot l(i)$ (r фиксировано), следовательно,

$$\varliminf_{i\to\infty} \frac{h((\gamma)_{i\cdot r}) - K((\gamma)_{i\cdot r})}{i} \geqslant 0.$$

Осталось показать, что

$$\varlimsup_{i\to\infty} \frac{h((\gamma)_{i\cdot r}) - K((\gamma)_{i\cdot r})}{i} \leqslant 0. \tag{5.19}$$

Для этого заметим, что поскольку случайные величины θ_j независимы и одинаково распределены, то все слова с одинаковым частотным набором $s_1, \ldots, s_{2^r}$ равновероятны (их вероятность равна $q_1^{s_1} \ldots q_{2^r}^{s_{2^r}}$). Из замечания 1.1 следует, что доля слов x таких, что $K(x) \leqslant h(x) - m$, среди всех

слов с фиксированным частотным набором $s_1, \ldots, s_{2^r}$, не превышает 2^{-m}. Следовательно, мера множества последовательностей, начинающихся на такие слова, не превосходит 2^{-m}. Мера множества последовательностей, начинающихся на слово с частотным набором $s_1, \ldots, s_{2^r}$ и удовлетворяющее условию

$$K(x) \leqslant h(x) - 2^{r+1} l(i) - m \qquad (i = s_1 + \ldots + s_{2^r}), \tag{5.20}$$

не превышает $2^{-(2^{r+1}l(i)+m)}$. Мера множества последовательностей, имеющих *какой-нибудь* фрагмент, удовлетворяющий условию (5.20), не превышает

$$\sum_{(s_1, \ldots, s_{2^r})} 2^{-(2^{r+1}l(s_1+\ldots+s_{2^r})+m)} \leqslant 2^{-m}.$$

Следовательно, тест, выдающий на ω супремум величины $h(x) - 2^{r+1}l(l(x)) - K(x)$ по всем ее фрагментам, удовлетворяет условию (4.1). Построить его алгоритм не представляет труда (это делается аналогично второй части доказательства теоремы 4.4). Очевидно, этот тест отбросит все последовательности, не удовлетворяющие условию (5.19), а поскольку γ случайна, то она этот тест выдерживает, и, следовательно, условию (5.19) удовлетворяет, что и требовалось доказать.

Теорема 5.3 выполняется не только в случае независимых испытаний. Дж. Т. Шварц поставил вопрос, имеет ли место аналогичный факт для произвольного эргодического стационарного процесса. Положительный ответ на этот вопрос дает

Предложение 5.1 (Л. А. Левин). *Пусть* $\{\xi_i\}$ $(i = 1, 2, \ldots)$ — *произвольный эргодический стационарный случайный процесс со значениями* $\xi_i \in \Omega$, P — *мера на его траекториях* $\omega \in \Omega^S$, *задающая этот процесс, а* H — *его энтропия* [1]. *Обозначим через* $\alpha_n^i(\omega)$ *слово* $(\overline{\xi_1})_n (\overline{\xi_2})_n \ldots (\overline{\xi_i})_n$. *Тогда для* P-*почти всех* ω

$$\lim_{n\to\infty} \lim_{i\to\infty} \frac{K(\alpha_n^i(\omega))}{i} = H.$$

Требование эргодичности здесь не существенно. Разница лишь в том, что в случае не эргодического процесса рассматриваемый предел будет не константой H, а функцией $f(\omega)$, измеримой относительно σ-алгебры инвариантных множеств траекторий. Эту функцию легко описать. Каждое инвариантное множество траекторий A, $P(A) > 0$, можно рассматривать как самостоятельный стационарный случайный процесс (распределенный по соответствующим условным вероятностям). Обозначим $h(A)$ энтропию этого процесса. Легко видеть, что функция $P(A) \cdot h(A)$ аддитивна. Тогда у нее существует производная Радона — Никодима, измеримая относительно σ-алгебры инвариантных множеств. Это и есть искомая функция $f(\omega)$.

Указатель терминов и обозначений

[1]) Определение энтропии произвольного стационарного случайного процесса см. в [40].

слов с фиксированным частотным набором $s_1, \ldots, s_{2^r}$, не превышает 2^{-m}. Следовательно, мера множества последовательностей, начинающихся на такие слова, не превосходит 2^{-m}. Мера множества последовательностей, начинающихся на слово с частотным набором $s_1, \ldots, s_{2^r}$ и удовлетворяющее условию

$$K(x) \leqslant h(x) - 2^{r+1} l(i) - m \qquad (i = s_1 + \ldots + s_{2^r}), \tag{5.20}$$

не превышает $2^{-(2^{r+1}l(i)+m)}$. Мера множества последовательностей, имеющих *какой-нибудь* фрагмент, удовлетворяющий условию (5.20), не превышает

$$\sum_{(s_1, \ldots, s_{2^r})} 2^{-(2^{r+1}l(s_1+\ldots+s_{2^r})+m)} \leqslant 2^{-m}.$$

Следовательно, тест, выдающий на ω супремум величины $h(x) - 2^{r+1}l(l(x)) - K(x)$ по всем ее фрагментам, удовлетворяет условию (4.1). Построить его алгоритм не представляет труда (это делается аналогично второй части доказательства теоремы 4.4). Очевидно, этот тест отбросит все последовательности, не удовлетворяющие условию (5.19), а поскольку γ случайна, то она этот тест выдерживает, и, следовательно, условию (5.19) удовлетворяет, что и требовалось доказать.

Теорема 5.3 выполняется не только в случае независимых испытаний. Дж. Т. Шварц поставил вопрос, имеет ли место аналогичный факт для произвольного эргодического стационарного процесса. Положительный ответ на этот вопрос дает

Предложение 5.1 (Л. А. Левин). *Пусть* $\{\xi_i\}$ $(i = 1, 2, \ldots)$ — *произвольный эргодический стационарный случайный процесс со значениями* $\xi_i \in \Omega$, P — *мера на его траекториях* $\omega \in \Omega^S$, *задающая этот процесс, а* H — *его энтропия* [1]. *Обозначим через* $\alpha_n^i(\omega)$ *слово* $(\overline{\xi_1})_n (\overline{\xi_2})_n \ldots (\overline{\xi_i})_n$. *Тогда для* P-*почти всех* ω

$$\lim_{n\to\infty} \lim_{i\to\infty} \frac{K(\alpha_n^i(\omega))}{i} = H.$$

Требование эргодичности здесь не существенно. Разница лишь в том, что в случае не эргодического процесса рассматриваемый предел будет не константой H, а функцией $f(\omega)$, измеримой относительно σ-алгебры инвариантных множеств траекторий. Эту функцию легко описать. Каждое инвариантное множество траекторий A, $P(A) > 0$, можно рассматривать как самостоятельный стационарный случайный процесс (распределенный по соответствующим условным вероятностям). Обозначим $h(A)$ энтропию этого процесса. Легко видеть, что функция $P(A) \cdot h(A)$ аддитивна. Тогда у нее существует производная Радона — Никодима, измеримая относительно σ-алгебры инвариантных множеств. Это и есть искомая функция $f(\omega)$.

Указатель терминов и обозначений

[1]) Определение энтропии произвольного стационарного случайного процесса см. в [40].

Литературные указания

Литература приведена по параграфам. Особенно полезными (к соответствующим параграфам) нам кажутся работы [5], [6], [11], [34], а из руководств — [1] и [37].

К предварительным замечаниями — [1] — [4].

К § 1 — [5] — [10].

К § 2 — [11] — [22], [33].

К § 3 — [8], [23] — [25].

К § 4 — [6], [7], [10], [22], [26] — [36] (работы [27] — [32] посвящены понятию коллектива Мизеса).

К § 5 — [5] — [7], [37] — [41] (работы [37] — [40] посвящены классическому понятию информации).

В нашей статье мы не затрагивали вопросов, связанных с оценкой числа шагов алгоритма, необходимого объема памяти и другими аспектами *сложности вычислений*. Читатель, интересующийся этими вопросами, может обратиться к работам [42], [43] (там же он найдет и литературу).

Наша библиография не претендует на полноту, однако мы постарались включить в нее основные публикации, дополняющие содержание нашей статьи.

ЛИТЕРАТУРА

[1] А. И. Мальцев, Алгоритмы и рекурсивные функции, М., «Наука», 1965.
[2] В. А. Успенский, Лекции о вычислимых функциях, М., Физматгиз, 1960.
[3] С. К. Клини, Введение в метаматематику, М., ИЛ, 1957.

[4] А. А. Марков, Теория алгорифмов, Труды Матем. ин-та им. В. А. Стеклова **42** (1954).
[5] А. Н. Колмогоров, Три подхода к определению понятия «количество информации», Проблемы передачи информации **1**:1 (1965), 3—7.
[6] А. Н. Колмогоров, К логическим основам теории информации и теории вероятностей, Проблемы передачи информации 5:3 (1969), 3—7.
[7] A. Kolmogoroff, Logical basis for information theory and probability theory, IEEE Trans., IT-14 (1968), 662—664.
[8] R. J. Solomonoff, A formal theory of inductive inference, Information and Control **7**:1 (1964), 1—22.
[9] Г. Б. Маранджян, О некоторых свойствах асимптотически оптимальных рекурсивных функций, Изв. АН Арм. ССР 4:1 (1969), 3—22.
[10] G. J. Chaitin, On the length of programs for computing finite binary sequences, I, II, Jorn. Assoc. Comp. Math. **13** (1966), 547—570; **15** (1968).
[11] Я. М. Барздинь, Сложность и частотное решение некоторых алгоритмически неразрешимых массовых проблем, препринт, 1970.
[12] Я. М. Барздинь, Сложность программ, распознающих принадлежность натуральных чисел, не превышающих n, рекурсивно перечислимому множеству, ДАН **182** (1968), 1249—1252.
[13] Я. М. Барздинь, О частотном решении алгоритмически неразрешимых массовых проблем, ДАН **191** (1970), 967—970.
[14] А. А. Марков, О нормальных алгорифмах, связанных с вычислением булевских функций и предикатов, Изв. АН, сер. матем. **31** (1967), 161—208.
[15] А. А. Марков, О нормальных алгорифмах, вычисляющих булевы функции, ДАН **157** (1964), 262—264.
[16] М. И. Канович, О сложности разрешения алгорифмов, ДАН **186** (1969), 1008—1009.
[17] М. И. Канович, О сложности перечисления и разрешения предикатов, ДАН **190**, (1970), 23—26.
[18] М. И. Канович, Н. В. Петри, Некоторые теоремы о сложности нормальных алгорифмов и вычислений, ДАН **184** (1969), 1275—1276.
[19] Н. В. Петри, Сложность алгорифмов и время их работы, ДАН **186** (1969), 30—31.
[20] Н. В. Петри, Об алгорифмах, связанных с предикатами и булевыми функциями, ДАН **185** (1969), 37—39.
[21] D. W. Loveland, A variant of the Kolmogorov notion of complexity, препринт, 1970.
[22] П. Мартин-Лёф, О колебании сложности бесконечных двоичных последовательностей, препринт, 1970.
[23] Я. М. Барздинь, О вычислимости на вероятностных машинах, ДАН **189** (1969), 699—702.
[24] К. де Леу, Э. Ф. Мур, К. Шеннон, Н. Шапиро, Вычислимость на вероятностных машинах, Автоматы (сб. переводов), М., ИЛ, 1956.
[25] В. Н. Агафонов, Об алгоритмах, частоте и случайности, Кандидатская диссертация, Новосибирск, 1970.
[26] А. Н. Колмогоров, Основные понятия теории вероятностей, М., ОНТИ, 1936.
[27] Р. Мизес, Вероятность и статистика, М.—Л., 1930.
[28] A. Wald, Die Wiederschpruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung, Ergebnisse eines mathemathishen Kolloquiums 8 (1937), 38—72.
[29] A. Church, On the concept of random sequence, Bull. Amer. Math. Soc. **46** (1940), 254—260.
[30] J. Ville, Étude critique de la notion de collectif, Paris, Gauthier-Villars, 1939.
[31] A. Kolmogoroff, On the tables of random numbers, Sankhya Indian Jorn. Statist., ser. A **25** (1963), 369—376.
[32] D. W. Loveland, A new interpretation of the von Mises concept of random sequence, Z. Math. Logik und Grundlagen der. Math. **12** (1966), 279—294.

[33] П. Мартин-Лёф, О понятии случайной последовательности, Теория вероятн. и ее примен. **11** (1966), 198—200.
[34] P. Martin-Löf, The definition of random sequences, Information and Control 9 (1966), 602—619.
[35] P. Martin-Löf, Algorithms and random sequences, University of Erlangen, Germany, 1966.
[36] P. K. Schnorr, Eine neue Charakterisierung der Zufälligkeit von Folgen, препринт, 1970.
[37] П. Биллингслей, Эргодическая теория и информация, М., Мир, 1969.
[38] К. Э. Шеннон, Математическая теория связи, 1948.
[39] И. М. Гельфанд, А. Н. Колмогоров, А. М. Яглом, К общему определению количества информации, ДАН **111** (1956), 745—748.
[40] А. Н. Колмогоров, Новый метрический инвариант транзитивных динамических систем и автоморфизмов пространств Лебега, ДАН **119** (1958), 861—864.
[41] А. Н. Колмогоров, Несколько теорем об алгоритмической энтропии и алгоритмическом количестве информации, УМН **23**:2 (1968), 201.
[42] Б. А. Трахтенброт, Сложность алгоритмов и вычислений, Новосибирск, 1967.
[43] M. Blum, A machine-independent theory of the complexity of recursive functions, Jorn. Assoc. Comp. Mach. **14** (1967), 322—337.

Поступило в редакцию 7 августа 1970 г.