МОСКОВСКИЙ ордена ЛЕНИНА и ордена ТРУДОВОГО
КРАСНОГО ЗНАМЕНИ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА

Л.А.Левин

НЕКОТОРЫЕ ТЕОРЕМЫ ОБ АЛГОРИТМИЧЕСКОМ ПОДХОДЕ К ТЕОРИИ ВЕРОЯТНОСТЕЙ И ТЕОРИИ ИНФОРМАЦИИ

Диссертация
на соискание ученой степени кандидата
физико-математических наук

Научный руководитель –
академик А.Н.КОЛМОГОРОВ

Москва – 1971

О Г Л А В Л Е Н И Е:

В диссертации используются термины и обозначения статьи [6], которая прилагается к тексту. В ней же находятся рисунки, на которые автор ссылается в тексте диссертации и указатель терминов и обозначений.

Автор испытывает глубокую благодарность к своему научному руководителю А.Н.Колмогорову, к А.К.Звонкину, оказавшему большую помощь в изложении результатов, а также к В.Н.Агафонову, Я.М.Барздиню, Р.Л.Добрушину, А.Г.Драгалину, М.И.Кановичу, А.Н.Колодию, П.Мартин-Лефу, Л.Б.Медведовскому, Н.В. Петри, А.Б.Сосинскому, В.А.Успенскому; Дж.Т.Шварцу и всем участникам семинара А.А.Маркова за обсуждение.

НЕКОТОРЫЕ ОПРЕДЕЛЕНИЯ И ОБОЗНАЧЕНИЯ.

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

$$\begin{aligned}
\Lambda &\longleftrightarrow 0\\
0 &\longleftrightarrow I\\
I &\longleftrightarrow 2\\
00 &\longleftrightarrow 3\\
0I &\longleftrightarrow 4\\
I0 &\longleftrightarrow 5\\
II &\longleftrightarrow 6\\
000 &\longleftrightarrow 7\\
00I &\longleftrightarrow 8\\
&\cdots\cdots\cdots
\end{aligned}$$

(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 $x = x_1 x_2 \ldots x_n$ (x_i - 0 или I), то

$$\bar{x} = x_1 x_1 x_2 x_2 \ldots x_n x_n 01 . \qquad (0.I)$$

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 $\pi_1(z) = \Lambda$, $\pi_2(z) = \Lambda$ I).

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

$$l(xy) = l(x) + l(y), \qquad (0.2)$$

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

I) Можно было бы устроить более стандартную нумерацию пар (x, y), однако для нас важно, чтобы выполнялось свойство (0.II) (см. ниже)

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

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

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

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 $\ell(\omega) \leqslant n$, 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.

$$\Gamma_x = \{\omega: (\omega)_{\ell(x)} = x\}. \tag{0.6}$$

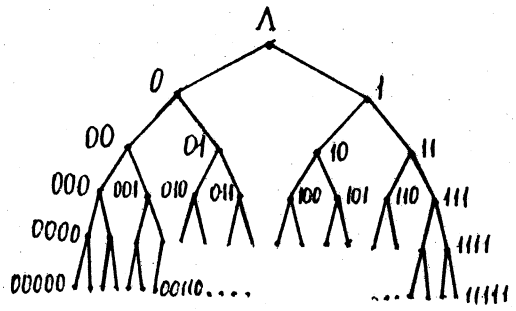

Рис. I

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. I).

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 чиоло 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 $x_1, \ldots, x_n$

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

на более короткую фразу (использующую новое обозначение):

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

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

Аналогично определяется отношение $\succeq$; $F \asymp G$ тогда и только тогда, когда $F \preceq G$ и $G \preceq F$. Очевидно, отно-

шения $\prec$, $\succ$ и $\asymp$ транзитивны. Очевидно также,

$$\ell(x) \asymp \log_2 x \quad \text{для } x > 0, \tag{0.9}$$

$$\ell(\bar{x}) \asymp 2\,\ell(x), \tag{0.10}$$

$$\ell(\bar{x}y) \asymp \ell(y) \quad (x \text{ входит как параметр}) \tag{0.11}$$

и т.д.

ГЛАВА I

ВВЕДЕНИЕ *)

§ I. Общая конструкция сложности

Круг вопросов, которым посвящена настоящая диссертация, возник в 1964 году с определения А.Н.Колмогоровым понятия сложности конструктивного объекта . (Близкие понятия независимо были рассмотрены А.А.Марковым и Р.Дж.Соломоновым). Сложностью слова x по алгоритму A А.Н.Колмогоров называл минимальную длину двоичного слова p , кодирующего x (т.е. такого, что $A(p) = x$). Определенная таким образом величина сильно зависит от вида алгоритма A , и центральным результатом, вызвавшим к жизни все дальнейшие исследования, явилась теорема, установленная независимо А.Н.Колмогоровым и (в несколько иных терминах) Р.Дж.Соломоновым. Она утверждает существование оптимального алгоритма A , дающего наименьшее по сравнению с любым другим алгоритмом B значение сложности с точностью до аддитивной константы C_B (не зависящей от x). Сложность слова x по произвольному оптимальному алгоритму является уже достаточно инвариантной величиной и фундаментальной характеристикой рассматриваемых объектов. Эта величина нашла много применений, породила круг вопросов, который быстро превратился в довольно развитую теорию

*) В диссертации используется терминология и обозначения статьи [6]

(см. например, обзорную работу [6]).

В процессе развития этой теории оказалось полезным ввести некоторые другие величины, аналогичные сложности (хотя и не совпадающие с ней). Так, А.А.Марков и Д.Ловеланд рассматривали сложность разрешения конечных двоичных последовательностей, П.Мартин-Леф построил величину, характеризующую "случайность" последовательностей, автор вводил "априорную вероятность" и т.п. К настоящему времени известно около десятка таких функций.

Возникает необходимость навести порядок в этом многообразии и взглянуть на все указанные величины с общей точки зрения. Это до некоторой степени делается следующими построениями.

ОПРЕДЕЛЕНИЕ. Финитной функцией мы будем называть таблицу, задающую функцию на некотором конечном множестве $A \subset S$ со значениями в S (будем считать, что на $A \setminus S$ эта функция принимает значение ∞).

ОПРЕДЕЛЕНИЕ. Объемным ограничением мы будем называть всякое перечислимое семейство V финитных функций, такое, что

1) Если $f \geqslant g$ и $g \in V$, то $f \in V$;

2) $\exists C \; \forall f, g \in V \quad (C + min\{f, g\}) \in V$.

Для простоты впредь будем считать $C = 1$.

ОПРЕДЕЛЕНИЕ. Пусть V - объемное ограничение. V-мажорантой будем называть всякую функцию $F(x)$, такую что

1) множество точек, лежащих над ее графиком, перечис-

лимо и

2) для всякой финитной функции g, если $g \geq F$, то $g \in V$.

ТЕОРЕМА I. Для любого объемного ограничения V существует минимальная с точностью до аддитивной константы V-мажоранта $K_V(x)$, т.е. такая, что для всякой V-мажоранты $L(x)$ $K_V(x) \preccurlyeq L(x)$.

ДОКАЗАТЕЛЬСТВО. Если есть конечное множество пар чисел $\mathcal{M}$, то взяв на каждой вертикали самую нижнюю его точку (если такая есть), мы получим график финитной функции. Будем называть ее нижней границей множества .

Пусть ч.р. функция $U(i,t)$ при каждом i перечисляет i-ое перечислимое множество пар (x,a). Зададим функцию $U'(i,t)$ таким алгоритмом; при каждом i она перечисляет i-е перечислимое множество, но медленнее, чем U и, возможно, не до конца. А именно каждый следующий элемент этого перечислимого множества она выдает только убедившись, что нижняя граница множества перечисленных элементов будет принадлежать семейству V.

Очевидно, что при каждом i, множество, перечисляемое функцией U', определяет некоторую V-мажорируемую функцию. Причем ни одна из них не окажется "забытой". Пусть теперь $\mathcal{M}$-множество пар, лежащих над парами вида $(x, a+Ci)$, где C - константа из пункта 2 определения объемного ограничения, а пара (x,a) принадлежит i-му множеству, перечисляемому функцией U'.

Докажем, что множество $\mathcal{M}$ задает V -мажоранту (аддитивная оптимальность ее очевидна из построения). То есть покажем, что любая финитная функция f , график которой лежит в $\mathcal{M}$ принадлежит семейству V . Из определения $\mathcal{M}$ ясно, что можно выбрать конечное семейство финитных функций $g_i \in V$ $i \leq n$ такое, что $f \geq \min\limits_{i \leq n}(g_i + c\,i)$. Отсюда следует, что $f \in V$. Действительно, пусть

$$h_k = \min_{i>k} (g_i + c\,(i-k))$$

Тогда $f \geq h_0$, $h_{k-1} = c + min(h_k, g_k)$ и индукцией по k от n до I получаем требуемое.

Теорема доказана.

Если объемное ограничение V разрешимо, то существует алгоритм, вычисляющий по x $m_V(x) = \min\limits_{f \in V} f(x)$. Очевидно, что любая V -мажоранта будет больше $m_V(x)$. Для простоты можно изучать не сами мажоранты $K_V(x)$, а разности $K_V(x) - m_V(x)$. Они будут V' -мажорантами, где $f \in V' \Leftrightarrow (f + m_V) \in V$. Очевидно, $m_{V'}(x) = 0$. Такие объемные ограничения V' мы будем называть "приведенными". К ним очевидно сводятся все разрешимые объемные ограничения.

ТЕОРЕМА 2. Среди приведенных объемных ограничений существует самое "узкое". Соответствующая ему универсальная мажоранта $p(x)$ будет самой большой *). Это ограничение за-

*) Эта мажоранта будет логарифмом наибольшего (с точностью до константы) полувычислимого распределения вероятностей на множестве натуральных чисел

дается условием $f \in V \Leftrightarrow \sum_x 2^{-f(x)} \leqslant 1$.

ДОКАЗАТЕЛЬСТВО. Очевидно, что V является приведенным объемным ограничением. Если V' - произвольное объемное ограничение и $f \in V$, то использовав конечное число раз равенство 2 из определения объемного ограничения убеждаемся, что $f \in V'$. Оказывается, что эта "предельная" мажоранта $p(x)$ не очень далеко ушла от обычной сложности $K(x)$ (введенной в работе [9]) (а значит $K(x)$ близка к пределу).

ТЕОРЕМА 3.

$$K(x) \leqslant p(x) \leqslant K(x) + 2 \log_2 K(x)$$

ДОКАЗАТЕЛЬСТВО. В одну сторону очевидно, что $K(x) \leqslant p(x)$ по теореме 2 (см. также теорему 4а). Для доказательства неравенства

$$p(x) \leqslant K(x) + 2 \log_2 K(x)$$

достаточно показать, что любая финитная функция $f(x) \geqslant K(x) + 2 \log_2 K(x)$ принадлежит объемному ограничению V (из теоремы 2). Действительно, $\sum_x 2^{-f(x)} = \sum_a \sum_{x\ K(x)=a} 2^{-f(x)} \leqslant$

$\leqslant \sum_a \sum_{x\ K(x)=a} 2^{-K(x)-2\log_2 K(x)} = \sum_a d\{x: K(x)=a\} \cdot \frac{1}{2^a a^2}$. Поскольку $d\{x: K(x)=a\} \leqslant 2^a$, то рассматриваемое выражение не превышает $\sum_a \frac{2^a}{2^a \cdot a^2} \leqslant 1$, что и дает требуемое.

§ 2. Примеры мажорант

Определение (А.Н.Колмогоров). Пусть F^1 - произвольная ч.р. функция. Тогда сложность x по F' есть

$$K_{F'}(x) \overset{def}{=} \begin{cases} \min\limits_{F(p)=x} l(p) \\ \infty \text{ , если такого } p \text{ не существует} \end{cases}$$

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

ОПРЕДЕЛЕНИЕ (А.Н.Колмогоров). Пусть F^2 - ч.р. функция. Тогда условная сложность слова x при известном y по F^2 есть

$$K_{F^2}(x/y) \overset{def}{=} \begin{cases} \min l(p) : F^2(p, y) = x \\ \infty \text{ , если } \forall p \; F^2(p, y) \neq x \end{cases}$$

ОПРЕДЕЛЕНИЕ (Д.Ловеланд, А.А.Марков). Сложность разрешения слова x по ч.р. функции F^2 есть

$$K_{F^2}(x) \overset{def}{=} \begin{cases} \min l(p) : \forall i < l(x) \; F_2(p, i) = x_i \\ \infty \text{ , если такого } p \text{ не существует} \end{cases}$$

здесь x_i — i -й знак слова x .

Мы привели три величины, играющие важную роль в теории сложностей. Покажем, что они вписываются в общее понятие V - мажорант. В частности, тогда из теоремы I получатся знаменитые теоремы оптимальности, открытие которых А.Н.Колмогоровым

и Р.Дж.Соломоновым заложило основы многих исследований по сложностям.

Пусть V_1 - множество финитных функций, таких что прообраз любой точки a имеет мощность, не больше 2^a. Пусть V_2 - множество финитных функций $f(x,y)$, определенных на парах чисел (x,y) (правильнее было бы сказать "на номерах пар"), таких что при каждом a, y_0 количество x, для которых $f(x,y_0) \leq a$ не превосходит 2^a.

Пусть V_3 - множество финитных функций, таких что количество ветвей в дереве слов, для которых $f(x) \leq a$ не превосходит 2^a.

Легко проверить, что V_1, V_2, V_3 являются объемными ограничениями. Классы A и B будем называть эквивалентными, если для любой функции f из одного из них, найдется $g \preceq f$ из другого.

ТЕОРЕМА 4. а) Класс V_1 -мажорант эквивалентен классу сложностей по различным алгоритмам.

б) Класс V_2 -мажорант эквивалентен классу условных сложностей по различным функциям.

в) Класс V_3 -мажорант эквивалентен классу сложностей разрешения.

ДОКАЗАТЕЛЬСТВО. Докажем теорему 4. а. Теоремы 4.б. и 4.в. доказываются аналогично.

Легко видеть, что сложность по любому алгоритму является V_1 -мажорантой функции. Обратно. Пусть F V_1 -мажоранта. Тогда всем точкам (x,n) над ее графиком можно

ставить в соответствие различные пары (x, p) , где $l(p) = n+1$, из графика некоторого алгоритма A . Из ограничения V_1 вытекает, что различных пар (x, p) хватит. Сложность по построенному алгоритму будет на I больше $F(x)$, что и требуется доказать.

§ 3. Инвариантные функции и сложность

Сложность обладает важным свойством инвариантности, а именно имеет место

ЗАМЕЧАНИЕ: При любом ч.р. изоморфизме между двумя перечислимыми множествами сложность их элементов изменяется не более чем на константу.

Кроме того, сложность обладает "информационной корректностью", а именно имеет место

ЗАМЕЧАНИЕ: Существует такая вычислимая нумерация пар

$$f(x, y) = i, \quad \pi_1(i) = x, \quad \pi_2(i) = y,$$

что сложность пары не меньше сложности ее элементов с точностью до аддитивной константы (это верно даже для любой вычислимой нумерации).

Сложность принимает значение порядка логарифма значений аргумента, т.е. она дает не очень много информации о словах, но оказывается, что более "богатой" инвариантной функции не существует даже среди функций любой природы.

ТЕОРЕМА 5. Любая инвариантная, информационно корректная (в данном выше смысле) функция $F(x)$ будет не больше функции $K(x)$ с точностью до мультипликативной константы. Сделать точность аддитивной нельзя, т.к. уже переход к словам в другом алфавите меняет сложность мультипликативно.

ДОКАЗАТЕЛЬСТВО. Алгоритм, по которому измеряется сложность, можно представить в виде суперпозиции функций $\mathcal{A}_1(x)$ и обратимой функции. Тогда из условия теоремы вытекает, что $F(x) \leq F(p)$, если $A(p) = x$.

Осталось показать, что $F(p) \leq C \cdot \ell(p)$. Это получается путем построения четырех изоморфизмов натурального ряда, комбинируя которые можно за n шагов получить из 0 любое слово длины не большей n. Теорема доказана.

§ 4. Вычислимые мажоранты сложности

Очевидно, если мы знаем само слово x и его сложность, то можно эффективно (хотя бы перебором) найти одну из программ наименьшей длины, кодирующих слово x. Более того, если мы знаем слово x и какое-нибудь число $S > K(x)$, 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 сложность по оптимальному алгоритму - невычислимая функция, то на практике приходится довольствоваться вычислимыми ее мажорантами, которые дают длину, хотя и не самого короткого кода, но зато эффективно вычислимого. В исследованиях Барздиня, Петри, Кановича

было показано, что в некоторых случаях эти мажоранты лишь очень грубо оценивают сложность, однако имеет место

ТЕОРЕМА 6. Любая "информационно корректная" (в смысле § 3) функция меньшая (с точностью до аддитивной константы) всякой вычислимой мажоранты сложности будет меньше (с той же точностью) и самой сложности.

ДОКАЗАТЕЛЬСТВО. Легко показать, что любой алгоритм можно представить в виде суперпозиции обратимого алгоритма, принимающего все значения и функции $\mathcal{T}_1(x)$. Сложность по обратимой функции является просто логарифмом ее обращения и, значит, вычислима. Из этого замечания следует утверждение теоремы.

§ 5. Сложность разрешения

При изучении сложности последовательностей (а не законченных слов) по многим причинам не совсем естественно пользоваться величиной $K(x)$ или $K(x/\ell(x))$. Поэтому А.А.Марковым и Д.Ловеландом была введена величина $KR(x)$, которая оказалась очень плодотворной. Например, очевидно

ЗАМЕЧАНИЕ. Последовательность ω тогда и только тогда является вычислимой, когда $KR((\omega)_n)$ ограничена.

Для $K(x)$ это неверно, а для $K(x/\ell(x))$ это не очевидно, и некоторое время назад было проблемой. Однако она была положительно решена автором независимо от Колодия, Ловеланда (США) и Мишина. Этот факт следует из теоремы, которая

устанавливает некоторую связь между $KR(x)$ и $K(x/\ell(x))$

ТЕОРЕМА 7. Для любой $\omega \in \Omega$ величина $KR((\omega)_n)$ ограничена тогда и только тогда, когда ограничена $K(\omega_n/n)$ *)

ДОКАЗАТЕЛЬСТВО. В одну сторону утверждение очевидно: если последовательность вычислима, то существует общерекурсивная функция $F'(n) = (\omega)_n$. Положим $F^2(p,n) = F'(n)$, тогда

$$K_{F_2}(\omega_n/n) = \ell(\Lambda) = 0$$

так как $F^2(\Lambda, n) = (\omega)_n$, следовательно,

$$K(\omega_n/n) \leq 0$$

то есть

$$K(\omega_n/n) \leq C.$$

Докажем обратное утверждение. Пусть $K((\omega)_n/n) \leq C$. 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 таблицу, как показано на рисунке 4 из [6]: в n-м столбце против слова p стоит $F_0^2(p,n)$ (см.

---

*) Однако, как показал Петри, не существует эффективного способа оценки $KR(\omega_n)$ по константе, ограничивающей $K(\omega_n/n)$ т.е. первая может быть очень велика.

(I.6) из [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

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

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 $K(\omega_n/n) \leq C$ равно K. Обозначим через m_2 минимальное число, такое, что все m_2 - фрагменты этих 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 $K \leq l$). Пусть $m = max(m_1, m_2)$ *)

---

*) Приводимое построение алгоритма использует числа l, K, m. Это построение не эффективно, так как не приводится эффективной процедуры построения чисел l, K и m. Мы лишь доказываем, что искомый алгоритм существует (интуционист выразился бы: "Не может не существовать"), поэтому для нас достаточно лишь самого факта существования чисел l, K, m.

Выберем из U бесконечное разрешимое подмножество U'. Пусть

$$V = U' \cap \{n : n > m\}.$$

Алгоритм, разрешающий i-ю (в лексикографическом порядке) из наших последовательностей, действует так: пусть мы хотим определить j-й знак i-й последовательности. Выберем минимальное $n_z \in V$ такое, что $n_z > j$, и начнем заполнять n_z-й столбец (то есть строить слова $F_z(p, n_z)$, $l(p) \leqslant c$). Как только окажется, что уже построено l слов, мы останавливаемся: мы получили все слова из A_{n_z}. Следующий шаг: выберем из A_{n_z} слова длины n_z, множество этих слов обозначим R_{n_z}. Дальше, аналогично построим множество $B_{n_{z+1}}$ и выберем из $B_{n_{z+1}}$ слова, являющиеся продолжениями слов из R_{n_z}; множество этих слов обозначим $C_{n_{z+1}}$. Далее, из $R_{n_{z+2}}$ выберем слова, являющиеся продолжением слов из $C_{n_{z+1}}$ – они образуют множество $C_{n_{z+2}}$; $C_{n_{z+3}}$ – множество слов из $B_{n_{z+3}}$, являющихся продолжением слов из $C_{n_{z+2}}$, 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 $K(\omega_n/n) \leqslant c$. Выберем из слов в C_{n_s} i-е по величине слово, и найдем его j-й знак; он будет искомым.

Г Л А В А II

МЕРЫ И ПРОЦЕССЫ

В этой главе рассматриваются детерминированные и недетерминированные процессы, производящие последовательности. Центральным результатом является построение универсальной полувычислимой меры и выяснение ее связи со сложностью, в конце главы результаты ее прилагаются к изучению возможностей вероятностных машин.

§ I. Определения. Эквивалентность мер

ОПРЕДЕЛЕНИЕ I. 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 до тех пор, пока это возможно (то есть пока F определена). В результате мы будем получать фрагменты некоторой новой последовательности (возможно конечной, или даже пустой *) - результата применения процесса F к ω (т.е. процесс F отоб-

*) Если для некоторого n $F((\omega)_n)$ определена и все $F((\omega)_m)$, $m > n$ совпадут 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.

ражает Ω в Ω^*). В этом случае будет использоваться также обозначение $\rho = F(\omega)$.

ЗАМЕЧАНИЕ. Существует универсальный процесс, то есть частично рекурсивная функция $H(i, x)$ такая, что для любого i $H(i, x)$ есть процесс и для любого процесса F существует i такое, что $H(i, x) \equiv F(x)$. Функцию $H(i,x)$ легко построить из универсальной частично рекурсивной функции. Без ограничения общности можно считать, что $H(\Lambda,\Lambda)=\Lambda$ (это понадобится в дальнейшем).

Процессы F и G будем называть эквивалентными, если для любой $\omega \in \Omega$ $F(\omega) = G(\omega)$.

ЗАМЕЧАНИЕ. Для любого процесса существует эквивалентный ему примитивно рекурсивный процесс.

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

ЗАМЕЧАНИЕ. Любой процесс на множестве тех последовательностей, к которым он применим, является непрерывной функцией (относительно естественной топологии пространства бесконечных двоичных последовательностей *).

ОПРЕДЕЛЕНИЕ. Будем называть процесс быстрорастущим (быстроприменимым к последовательности ω), если существует монотонная неограниченная общерекурсивная функция $\Phi(n)$, такая, что для любых x (для любых x, являющихся фрагмента-

---

*) В этой топологии Ω гомеоморфно канторовскому совершенному множеству.

ми ω) и n таких, что $\ell(x) = n$ 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. Легко показать, что процесс, применимый ко всем ω, является общерекурсивным и быстрорастущим. Очевидно, верно и обратное.

ОПРЕДЕЛЕНИЕ. Пусть P - вероятностная мера на Ω. Будем говорить, что процесс P -регулярен, если множество последовательностей, к которым он применим, имеет P -меру I.

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

ОПРЕДЕЛЕНИЕ. Назовем меру P на Ω вычислимой, если существуют общерекурсивные функции $F(x,n)$ и $G(x,n)$ такие, что рациональное число $\alpha_P(x,n) = \frac{F(x,n)}{G(x,n)}$ приближает число $P(\Gamma_x)$ с точностью до 2^{-n}.

ЗАМЕЧАНИЕ. 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

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

Эта мера соответствует испытаниям Бернулли с вероятностью $p = \frac{1}{2}$; она же - мера Лебега на отрезке $0,1$. Мера L , очевидно, вычислима.

ТЕОРЕМА 8 ж). а) Для любой вычислимой меры P и любого P - регулярного процесса F мера

$$Q\{\Gamma_y\} = P\{U\Gamma_x : (F(x) \supset y)\}$$

(то есть 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 FG).

ДОКАЗАТЕЛЬСТВО. а) Нам нужно уметь вычислять $Q\{\Gamma_y\}$ с точностью до 2^{-n} , то есть находить $\alpha_Q(y,n)$ жж). Выберем m такое, чтобы

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

ж) Несколько более слабый факт независимо доказал Манн (США)

жж) Мы не будем строить приближение с избытком, а построим произвольное приближение; сделать из него приближение с избытком легко

(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 $2^{-(m+n+1)}$, то есть положим

$$\alpha_Q(y,n) = \sum_{\substack{x:\, l(x)=m \\ F(x) \supset y}} \alpha_P(x, m+n+1)$$

Тогда наша ошибка (то есть $\alpha_Q(y,n) - Q\{\Gamma_y\}$) 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 двичным 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. 5 [6] 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.

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} (то есть $\alpha_Q(y, 2n)$). Выделим те слова z длины n, для которых

$$\sum_{y \geq z} \left(\alpha_Q(y, 2n) - 2^{-2n}\right) \geq 1 - \alpha''_n \qquad (1)$$

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

$$\sum_{y \geq z} \alpha_Q(y, 2n) \geq \alpha'_n \qquad (2)$$

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

II. Множества $\bigcup \Gamma_z$ согласно (1) и (2), являются отрезками, содержащими (при каждом n) g - прообраз точки α, поэтому, если процесс F применим к α, то его резуль-

татом будет $g^{-1}(\alpha)$ (прообразом точек $\alpha \in [\sigma', \sigma'']$ будем считать точку γ - см. рис. 5 из [6]). Для того чтобы доказать, что процесс F искомый, достаточно тем самым доказать его L -регулярность.

I) 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 (I) и (2) следует, что $Q\{U\Gamma_z\} \to 0$ при $n \to \infty$, откуда, если α не является точкой типа ρ, соответствующей отрезку меры 0, то и сами отрезки $U\Gamma_z$, 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

3) если $\beta = \frac{m}{2^k}$, то отрезки $U\Gamma_z$ могут всегда содержать как последовательности, лежащие слева от $\frac{m}{2^k}$ и, следовательно, начинающиеся на слово $m-1$, так и последователь-

ности, лежащие справа от $\frac{m}{2^k}$ и, следовательно, начинающиеся на слово m. В этом случае наиболее длинный общий фрагмент всех выделенных слов z будет иметь длину меньше k.

Итого, процесс F может быть неприменим только к последовательностям типа ρ (см. рис. 5 из [6]), типа σ' и σ'' (см. рис. 5), а также к последовательностям, имеющим двоично-рациональные прообразы. Очевидно, что множество таких последовательностей не более чем счетно, следовательно, процесс $F-L$ -регулярен.

Ш. 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 вследствие теоремы II будет доказан более общий результат) и, во-вторых (может быть), к последовательностям β, на которых функция g принимает двоично-рациональные значения (аналогично П(3)). Если процесс F применим к этим двоично-рациональным значениям α, то наши последовательности β вычислимы (как F - образы двоично-рациональных); если же процесс F не применим 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. Полувычислимой мерой *) называется мера, по которой распределены результаты применения произвольного (не обязательно регулярного) процесса к последовательностям, распределенным по некоторой вычислимой мере.

ЗАМЕЧАНИЕ. 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. 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 8б).

ТЕОРЕМА 9. Мера P является полувычислимой тогда и только тогда, когда существуют общерекурсивные функции $F(x,t)$ и $G(x,t)$ такие, что функция $\beta_P(x,t) = \dfrac{F(x,t)}{G(x,t)}$ монотонно не убывает по t, и

$$\lim_{t\to\infty} \beta_P(x,t) = P\{\Gamma_x\} \qquad (3)$$

Из этой теоремы вытекает, что класс полувычислимых мер (точ-

*) Название "полувычислимая" оправдывается теоремой 9

нее их логарифмов) эквивалентен классу V -мажорируемых функций, где V - множество финитных функций, для которых $\sum_{x\in M} 2^{-f(x)} \leqslant 1$ при всяком M , в котором никакие два слова не являются началом друг друга.

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 , и, обозначив результат за $F_t(y)$ (если он еще не получился, то $F_t(y)=\Lambda$, положим

$$B_P(x,t)=L\{\cup\Gamma_y : x\subset F_t(y)\}$$

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 при всех t

$$\beta_P(x,t) \geq \beta_P(x0,t) + \beta_P(x1,t)$$

(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) 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 $x < y$ 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 $l(x) \leq l(z)$ и $t \leq l(z)$, 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 для $z' \subset z''$ выполняется $x' \subset x''$.

§ 3. Универсальная полувычислимая мера

ТЕОРЕМА 10. Существует универсальная полувычислимая мера R, то есть полувычислимая мера, удовлетворяющая следу-

ющему условию: для любой полувычислимой меры Q найдется константа C такая, что

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

для любого x ж).

ДОКАЗАТЕЛЬСТВО. Согласно замечанию на стр. 21 существует универсальный процесс $H(i,x)$. Положим

$$F(z) = H(\pi_1(z), \pi_2(z))$$

Легко показать, что $F(z)$ - процесс. 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 $F(x)$ переводит в Γ_x те же последовательности с приписанным к ним слева словам $\bar{i}$, где i - номер процесса G (то есть $H(i,x) = G(x)$ для всех x) и, может быть, некоторые другие последовательности. Поэтому мера Γ_x не может уменьшиться более чем в C раз, где в качестве C можно взять $2^{l(\bar{i})} = i^2$.

ЗАМЕЧАНИЕ. 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 пре-

---

ж) Иными словами, Q -абсолютно непрерывна относительно R, причем производная Радона-Никодима ограничена константой C.

небречь 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".

Представляет интерес следующий факт:

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

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

Таким образом, эта вероятность имеет порядок примерно $\frac{1}{n}$ ж).

ж) 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 1/3 650 000. Это было бы верно, если бы наша информация о Солнце исчерпывалась указанным фактом.

Доказательство этого утверждения легко вытекает из теоремы II, если учесть, что сложность разрешения слова, состоящего из n нулей и единицы, не превосходит $\log_2 n + c$, 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 II.

$$\left| KR(x) - (-\log_2 R\{\Gamma_x\}) \right| \lesssim 2 \log_2 KR(x)$$

ДОКАЗАТЕЛЬСТВО. Пусть $KR(x) = i$, то есть существует слово p, $\ell(p) = i$ такое, что для всякого $n \leq \ell(x)$

$$G_0^2(p, n) = x_n$$

(здесь G_0^2 из теоремы 8 [ ]). Тогда легко построить процесс, который любую последовательность, начинающуюся на слово $\overline{\ell(p)}\, p$, переводит в последовательность, начинающуюся на слово x: он должен сначала выделить $\overline{\ell(p)}$; по нему восстановить $\ell(p)$; зная $\ell(p)$, "прочитать" само слово p и 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 меньше, чем $2^{-\ell(\overline{\ell(p)}\, p)}$. Поэтому согласно теореме 10

$$R\{\Gamma_x\} \geq \ell \cdot 2^{-\ell(\overline{\ell(p)}\, p)}$$

откуда

$$-\log_2 R\{\Gamma_x\} \leq l(\overline{l(p)}\ p) = l(\overline{l(p)}) + l(p) \leq$$

$$\leq l(p) + 2\,l(l(p)) = i + 2\,l(i) = KR(x) + 2\,l(KR(x))$$

Пусть теперь $R\{\Gamma_x\} = q$. Обозначим

$$l(q) = [-\log_2 q] .$$

Оценим сложность разрешения слова x ; для этого мы покажем, что любой знак слова x можно восстановить по информации, задаваемой тройкой слов $l(q)$, k, i (или, что то же самое, одним словом $\overline{l(q)}, \bar{k}\, i$), где $K = 0$ или I, а $i \leq 2^{l(q)+1}$. Наш алгоритм будет действовать так: по слову $l(q)$ он начнет выстраивать дерево (см. рис. 2 из [6]) слов y таких, что

$$R\{\Gamma_y\} > 2^{-l(q)-1}$$

(Для этого нужно вычислять $\beta_R(y,t)$ для все больших значений t и y и подстраивать слово y к дереву, как только для какого-то t станет

$$\beta_K(y,t) > 2^{-l(q)-1}$$

Слово x принадлежит этой совокупности. На каждом шагу ал-

горитма мы будем выделять в уже выстроенной части дерева совокупность "максимальных" слов, то есть 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. 6 из [6]) 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 пристроили второй в порядке получения росток, количество максимальных слов увеличится на 1 и станет равным i).

При этом,

$i \leq 2^{l(q)+1}$, то есть $l(i) \leq l(q)+1$

в итоге $KR(x) \leq l(\overline{l(q)}\,\bar{k}\,i) \leq 2\,l(l(q)) + l(i) \leq$

$2\,l(l(q)) + l(q) \leq -\log_2 R\{\Gamma_x\} + 2\log_2(-\log_2 R\{\Gamma_x\})$.

Но, по доказанному ранее,

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

$\leq 2\log_2 KR(x)$,

откуда $KR(x) \leq -\log_2 R\{\Gamma_x\} + 2\log_2 KR(x)$, что и завершает доказательство.

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

$$P(\omega_n) \geq c \cdot R(\omega_n)$$

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

В качестве следствия получаем известную теорему де-Леу-Мура-Шапиро-Шениона о вероятностных машинах:

СЛЕДСТВИЕ. Последовательность имеет положительную вероятность по некоторой (а, следовательно, и по универсальной) полувычислимой мере тогда и только тогда, когда она вычислима.

ДОКАЗАТЕЛЬСТВО. Из теоремы II следует, что мера всех фрагментов 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, недоступных 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 Π, разрешима с помощью вероятностных машин, если универсальная мера R таких последовательностей больше нуля. Следующая теорема показывает, что такие задачи действительно можно решать со сколь угодно большой надежностью, на машинах, использующих датчик случайных чисел, причем весьма экономно (употребляя мало знаков датчика).

Мы будем называть функции $f(n)$ и $g(n)$ асимптотически равными (и писать $f(n) \backsim g(n)$, если $\frac{f(n)}{\log_2 f(n)} \asymp \frac{g(n)}{\log_2 g(n)}$ аналогично будут пониматься неравенства).

* и понятие массовой проблемы в работе Ю.Т. Медведева в ДАН том 104, №4

ТЕОРЕМА I2. Пусть $A \subseteq \Omega$ $R\{A\} > 0$. Тогда для любого $\varepsilon > 0$ существует быстрорастущий (т.е. слаботабличный) со скоростью роста $l(F(x)) \gtrsim l(x)$ процесс, который, будучи применен к последовательностям, распределенным по мере L выдает последовательности из A с вероятностью не меньше $1-\varepsilon$ *).

Очевидно, что, например, задачу получения какой-нибудь максимально сложной последовательности решить процессом, растущим быстрее нельзя, т.к. при применении процесса сложность слов не может возрастать. Если же сложность последовательностей из A мала, то должны существовать короткие программы, дающие фрагменты этих последовательностей. Но мыслима ситуация, когда эти программы очень уникальны и специфичны (из-за специфики A) и использовать вместо них столь же короткие случайные числа нельзя и значит процесс, решающий задачу A , будет расти медленно. Однако можно доказать, что в этом случае существует и "быстрый" процесс.

---

*) Заметим, что, во-первых, построение этого процесса по ε не всегда эффективно, во-вторых, как показал Н.Петри, этот процесс не всегда может быть заменен табличным (т.е. быстрорастущим и общерекурсивным).

ТЕОРЕМА 13. Пусть g - общерекурсивная монотонная функция. Задача получения последовательности из множества A тогда и только тогда разрешима с помощью процесса, растущего со скоростью асимптотически $\geq g(n)$, когда существует множество *) $B \subseteq A$, $R\{B\} > 0$ такое, что

$$KR(x) \lesssim n \quad , \text{ где } x = (\omega)_{g(n)}, \omega \in B$$

ДОКАЗАТЕЛЬСТВО. В одну сторону она, очевидно, вытекает из замечаний предыдущего абзаца. Докажем ее в другую сторону. Пусть $B \subseteq A$, $R(B) > 0$ $\forall \omega \in B$, сложность ее $g(n)$ -фрагментов $\geq n + c \log n$.

Сначала построим полувычислимую меру $P(x)$ такую, что

$$P(B) \geq 0 \text{ и } \forall x\ l(x) = g(n) \quad P(x) = \frac{a}{2 \cdot 2^n \cdot n^{c+u}}$$

где a - целое число. Для этого сначала округлим R в точках значений $g(n)$ и "срежем" по неравенству $P(x) \geq P(x0) + P(x1)$. Легко показать, что полученная мера на последовательностях из B уменьшилась не более чем на $\frac{1}{2 \cdot 2^n \cdot n^{c+2}}$. В то же время $KR(x) \geq n + c \log n$, значит по теореме II $R(x) \geq \frac{1}{2^n \cdot n^{c+2}}$, и значит на

*) Это множество B всегда может быть выбрано замкнутым.

последовательностях из B $P(x) \geqslant \frac{R(x)}{2}$. Отсюда $P(B) > 0$. Далее для P строится процесс по доказательству теоремы 9, причем множества, соответствующие паре (x,t) (где $l(x) = g(n)$) выбираются состоящими из интервалов длины кратной $\frac{1}{2 \cdot 2^n \cdot n^{c+n}}$. Очевидно, что этот процесс требуемый.

Приведем один результат, касающийся возможностей решения на вероятностных машинах стандартных алгоритмических задач. Первый интересный результат такого характера был получен Барздинем.

Будем называть бесконечное множество натуральных чисел иммунным, если оно не содержит никакого бесконечного перечислимого подмножества.

ПРЕДЛОЖЕНИЕ (Барздинь). Существует иммунное множество такое, что задача получения последовательности, характеристической для некоторого его бесконечного подмножества, разрешима с помощью вероятностной машины.

Интересной разновидностью иммунных множеств являются гипериммунные множества. Однако для них имеет место

ТЕОРЕМА I4.*) Каково бы ни было гипериммунное множество M , задача получения последовательности, характеристической для какого-нибудь его бесконечного подмножества, не-

---

*) Независимо от автора диссертации доказано также В.Н.Агафоновым

разрешима с помощью вероятностной машины.

ДОКАЗАТЕЛЬСТВО. Пусть есть машина, решающая с положительной вероятностью эту задачу. Тогда по теореме I2 существует машина, решающая ее с вероятностью $p > \frac{2}{3}$. Построим функцию $f(i)$, вычисляемую таким алгоритмом: он применяет машину на дереве последовательностей, пока на мере $\geqslant \frac{2}{3}$ она не получит последовательностей, имеющих по крайней мере i единиц и тогда берет максимум мест этих единиц. Очевидно, что она будет мажорировать прямой пересчет множества M, что и требовалось доказать.

В дальнейшем Петри показал, что если множество M не фиксировать, то задача получения последовательности, характеристической для гипериммунного множества разрешима на вероятностных машинах. Однако имеет место

ТЕОРЕМА I5. Назовем сильно-гипериммунным множество, прямой подсчет которого мажорирует, начиная с некоторого места любую вычислимую функцию. Задача получения последовательности такой, что множество, для которого она характеристическая, сильно гипериммунно, не разрешима на вероятностных машинах.

Доказательство аналогично предыдущей теореме.

Г Л А В А Ш

ТЕОРИЯ ИНФОРМАЦИИ

§ I. Определение и простейшие свойства

Сложность $K(x)$ интуитивно означает количество информации, необходимое для восстановления текста x. Условная сложность $K(x/y)$ интуитивно означает 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 (А.Н.Колмогоров)количества информации в y об x:

$$I(y:x) \overset{def}{=} K(x) - K(x/y)$$

ЗАМЕЧАНИЕ. а) $I(x:y) \gtrsim 0$

б) $I(x:x) \asymp K(x)$.

ДОКАЗАТЕЛЬСТВО. а) Пусть

$$F^2(p, x) = F_0^1(p)$$

Тогда, если $F_0^1(p_0) = y$ и $K(y) = l(p_0)$, то, так как $F^2(p_0, x) = y$, имеем:

$$K(y/x) \lesssim K_{F_2}(y/x) = K(y)$$

б) Пусть $F^2(p, x) = x$. Тогда и $F^2(\Lambda, x) = x$, откуда

$$K(x/x) \leq K_{F^2}(x/x) = l(\Lambda) = 0.$$

Замечая, что $I(x:x) = K(x) - K(x/x)$, получаем требуемое.

§ 2. Коммутативность информации

Классическое Шенноновское количество информации в одной случайной величине о другой удовлетворяет условию коммутативности, то есть

$$\mathcal{J}(\xi:\eta) = \mathcal{J}(\eta:\xi)$$

Для колмогоровского количества информации в одном тексте о другом точного равенства, вообще говоря, не будет.

ПРИМЕР. Очевидно, что для любого l_0 существует слово x длины l_0 такое, что

$$K(x/l(x)) \geq l(x) - 1.$$

Согласно теореме 4 б существуют сколь угодно большие l_0 такие, что

$$K(l_0) \geq l(l_0) - 1.$$

Для так выбираемых пар слов x и l_0 ($l(x) = l_0$) имеем

$$I(x:\ell_0) = K(\ell_0) - K(\ell_0/x) \geq \ell(\ell_0)$$

$$I(\ell_0 : x) = K(x) - K(x/\ell_0) \leq \ell_0 - \ell_0 = 0.$$

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 в 1967 году А.Н.Колмогоров и Л.Левин указанный порядок является для нее предельным и, следовательно, если пренебречь величинами бесконечно малыми по сравнению с информацией, содержащейся в обоих словах, величина $I(x:y)$ будет все же коммутативной. А именно, А.Н.Колмогоров и Л.Левин доказали следующую теорему

ТЕОРЕМА I6 ж). а) $\left|I(x:y) - I(y:x)\right| \leq 12\,\ell(K(\bar{x}y))$

б) $\left|I(x:y) - [K(x) + K(y) - K(\bar{x}y)]\right| \leq 12\,\ell(K(\bar{x}y))$

ДОКАЗАТЕЛЬСТВО. Будем доказывать неравенство только в одну сторону:

$$I(x:y) \geq I(y:x) - 12\,\ell(K(xy)) \qquad (I)$$

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

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 $K(y) \leq b$, $K(x/y) \leq c$. Существование такой функции следует из теоремы 0.4 [ ] (с учетом замечания). Обозначим через j количество таких y (j невычислимо зависит от x, b, c). Функция F^4 определена для всех $n \leq j$ и только для них. Следовательно, предикат $\Pi(b, c, d, x)$, утверждающий, что вышеопределенное 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

$$K(x) \leq a, \quad \Pi(b, c, d, x)$$

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 $m \leq i$ 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 $K(y') \leq b$ и $K(x/y') \leq c$ (j зави-

сит от x, b, c), а i есть количество слов x' таких, что $K(x') \leq a$ и соответствующее число $j' \geq 2^{l(i)}$. Легко видеть, что $i \cdot 2^{l(j)}$ не превосходит количества пар (x', y') таких, что $K(y') \leq b$, $K(x'/y') \leq c$, которых в свою очередь не больше чем 2^{b+c+2}, откуда

$$l(i) + l(j) \leq b + c \qquad (2)$$

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

$$K(y/x) \leq l(\bar{b}\bar{c}n) \leq 2l(b) + 2l(c) + l(j) \qquad (3)$$

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

$$a = K(x) = l(\bar{a}\bar{b}\bar{c}\bar{d}m) \leq 2l(a) + 2l(b) + 2l(c) + 2l(d) + l(i) \qquad (4)$$

Из неравенства (2); (3), (4), а также из того, что каждая из величин $l(a)$, $l(b)$, $l(c)$, $l(d) = l(l(j))$ не превосходит $l(K(\bar{x}y))$, легко получаем

$$K(y/x) \leq b + c - a + 12\, l(K(\bar{x}y))$$

откуда и следует (a)

б) Очевидно, что

$$K(\bar{x}\,\bar{x}\,y) \lesssim K(\bar{x}\,y)$$

отсюда, согласно пункту а) настоящей теоремы

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

то есть

$$K(\bar{x}y) - K(\bar{x}y/x) - K(x) + K(x/\bar{x}y) \lesssim 12\,\ell(K(\bar{x}y))$$

или

$$\left|K(\bar{x}y) - K(x) - K(y)\right| + K(y) - K(\bar{x}y/x) - K(x/\bar{x}y) \lesssim 12\,\ell(K(\bar{x}y))$$

откуда, замечая что

$$K(x/\bar{x}y) \gtrsim 0$$

$$K(\bar{x}y/x) \gtrsim K(y/x)$$

получаем искомое.

§ 3. Энтропия произвольных динамических систем (стационарных случайных процессов) и алгоритмическое количество информации

А.Н.Колмогоров показал, что для процессов независимых испытаний алгоритмическое количество информации асимптотически совпадает с классическим (вероятностным) (см. [ ],

т. 25). Принимая во внимание теорему I6 б, легко понять, что для этого достаточно установить связь между алгоритмической сложностью и вероятностной энтропией.

Дж.Т.Шварц поставил вопрос, имеет ли место аналогичный факт в случае произвольного эргодического стационарного процесса (т.е. такого, для которого определена энтропия). Автор решил этот вопрос положительно в следующей теореме

ТЕОРЕМА I7. Пусть $\{\xi_i\}$, $i = 1,2,\ldots$ произвольный эргодический стационарный случайный процесс со значениями $\xi_i \in \Omega$, P - мера на его траекториях $u \in \Omega^N$, задающая этот процесс, а H - его энтропия. Обозначим через $\alpha_n^i(\omega)$ слово $(\bar{\xi}_1)_n (\bar{\xi}_2)\ldots(\bar{\xi}_i)_n$. Тогда для P - почти всех u

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

Очевидно, что требование эргодичности здесь не существенно. Разница лишь в том, что в случае не эргодического процесса нужно брать не его среднюю энтропию H, а "энтропию в точке", которая является функцией, измеримой относительно σ - алгебры инвариантных множеств и интеграл ее по любому такому множеству равен его средней энтропии.

Это легко следует из того, что произвольный стационарный случайный процесс можно "разложить" на эргодические. Вернемся к эргодическому случаю. Сразу заметим, что утверждение теоремы достаточно доказать для процессов с дискретными зна-

чениями $(\xi)_n$. Общий случай следует из этого, если перейти к пределу по n.

Рассмотрим множество 2^n-ичных последовательностей u, представляющих из себя реализации нашего случайного процесса. На нем определено преобразование T, осуществляющее сдвиг времени на единицу и T - инвариантная эргодическая мера, задающая наш процесс. За единицу времени может получиться $2^{n\cdot k}$ различных последовательностей длины к X_i^k Очевидно, что для любого ε, найдется K такое, что

$$-\sum_{i<2^{n\cdot k}} P(X_i^k)\log_2 P(X_i^k) \leq k\cdot(H+\varepsilon)$$

Поскольку T^k, как и T, сохраняет меру P, то по центральной эргодической теореме (Ц.Э.Т.) для P - почти любой последовательности существует при любом e предел частоты таких m, что последовательность $T^{mk+e}(\omega)$ начинается на X_i^k.

Выберем любую такую ω и обозначим эти пределы для нее через $P_{i,e}$. Из Ц.Э.Т. для T и эргодичности последнего вытекает, что почти всегда $\sum_{e\leq k}\frac{P_{ie}}{k} = P(X_i^k)$. Таким образом, у нас есть k распределений вероятностей на конечном множестве X_i^k и одно среднее между ними с энтропией $\leq k(H+\varepsilon)$.

Используя выпуклость энтропии, можно заключить, что по

крайней мере у одного из распределений - слагаемых энтропия меньше $K(H+\varepsilon)$. Таким образом, для нашей u существует l такое, что энтропия частот P_{il}, с которыми число m удовлетворяет условию "$T^{mk+l}(\omega)$ начинается на X_i^k" не превышает $K(H+\varepsilon)$. Отсюда, по теореме А.Н.Колмогорова (см. теорему 5.I из [6]) вытекает, что "удельная сложность" не превышает $H+\varepsilon$. Поскольку это доказательство проведено для любого ε и почти любого u, то можно заключить, что "удельная сложность" почти всех u не превышает H, что дает "половину" нашей теоремы.

Теперь докажем, что удельная сложность не может и быть меньше H.

Для этого нам придется привлечь некоторые результаты главы П. Рассмотрим набор X_i^τ значений какой-нибудь реализации ω процесса за первые k моментов времени и сравним четыре величины: энтропию H, логарифм вероятности этого набора, деленный на τ $\dfrac{\log P(X_i^\tau)}{\tau}$, логарифм априорной вероятности его (см. опр. R), тоже деленный на τ $\dfrac{\log R(X_i^\tau)}{\tau}$ и удельную сложность $\dfrac{K(x_i^\tau)}{\tau}$.

Покажем, что пределы при $\tau \to \infty$ этих величин совпадают. Для первых двух величин это вытекает из теоремы Шеннона-Макмиллана-Бреймана, для последних двух из теоремы II настоящей диссертации, а для двух средних из замечания на стр. Теорема доказана.

Л И Т Е Р А Т У Р А

I. В.Н.Агафонов, Об алгоритмах, частоте и случайности, Кандидатская диссертация, Новосибирск, I970

2. Я.М.Барздинь, Сложность и частотное решение некоторых алгоритмически неразрешимых массовых проблем, препринт, I970

3. Я.М.Барздинь, Сложность программ, распознающих принадлежность натуральных чисел, не превышающих n , рекурсивно перечислимому множеству, ДАН I82 (I968), I249-I252

4. Я.М.Барздинь, О частотном решении алгоритмически неразрешимых массовых проблем, ДАН I9I (I970), 967-970

5. Я.М.Барздинь, О вычислимости на вероятностных машинах, ДАН I89 (I969), 699-702

6. А.К.Звонкин и Л.А.Левин, Сложность конечных объектов и ~~объектов и~~ обоснование понятий информации и случайности с помощью теории алгоритмов, УМН, I970, вып. 6, стр.85

7. М.И.Канович, О сложности перечисления и разрешения предикатов, ДАН I90 (I970), 23-26

8. М.И.Канович, Н.В.Петри, Некоторые теоремы о сложности нормальных алгорифмов и вычислений, ДАН I84 (I969), I275-I276

9. А.Н.Колмогоров, Три подхода к определению понятия "количество информации" ,Проблемы передачи информации I:I (I965) 3-7

I0. А.Н.Колмогоров, К логическим основам теории информации и теории вероятностей, Проблемы передачи информации 5:3 (I969), 3-7

II. М.И.Канович, О сложности разрешения алгоритмов, ДАН I86, (I969), I008-I009

I2. А.Н.Колмогоров, Несколько теорема об алгоритмической энтропии и алгоритмическом количестве информации, УМН, 23:2 (I968), 20I

I3. К. де Леу, Э.Ф.Мур, К.Шеннон, Н.Шапиро, Вычислимость на вероятностных машинах, Автоматы (сб. переводов), М., ИЛ, I956

I4. Г.Б.Маранджан, О некоторых свойствах асимптотически оптимальных рекурсивных функций, Изв. Арм. АН ССР 4:I (I969), 3-22

I5. А.А.Марков, О нормальных алгорифмах, связанных с вычислением булевских функций и предикатов, Изв. АН, сер. матем. 3I (I967), I6I-208

I6. А.А.Марков, О нормальных алгорифмах, вычисляющих булевы функции, ДАН I57 (I964), 262-264

I7. П.Мартин-Леф, О колебании сложности бесконечных двоичных последовательностей, препринт, I970

I8. П.Мартин-Леф, О понятии случайной последовательности, теория вероятн. и ее примен. II (I966), I98-200

I9. Н.В.Петри, Сложность алгорифмов и время их работы, ДАН I86 (I969), 30-3I

20. Н.В.Петри, Об алгорифмах, связанных с предикатами и булевыми функциями, ДАН I85 (I969), 37-39

21. Б.А.Трахтенброт, Сложность алгоритмов и вычислений, Новосибирск, 1967

22. С.В.Яблонский, Об алгоритмических трудностях синтеза минимальных схем, Проблемы кибернетики, 2, 1959, 75-121

23. 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)

24. A.Kolmogoroff, Logical basis for information theory and probability theory, IEEE Trans., IT-14 (1968), 662-664

25. D.W.Loveland, A variant of the Kolmogorov notion of complexity, препринт, 1970

26. Mann I. Probabilistic recursive functions, J. Symbolic Logik 31 (1966), No. 4, 698

27. P.Martin-Löf, The definition of random sequences, Information and Control 9 (1966), 602-619

28. P.Martin-Löf, Algorithms and random sequences, University of Erlangen, Germany, 1966

29. P.K.Schnorr, Eine neue Charakterisierung der Zufälligkeit von Folgen, препринт, 1970

30. R.J.Solomonoff, A formal theory of inductive iference, Information and Control 7:1 (1964), 1-22