ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том X 1974 Вып. 3

УДК 621.391.1:519

ЗАКОНЫ СОХРАНЕНИЯ (НЕВОЗРАСТАНИЯ) ИНФОРМАЦИИ И ВОПРОСЫ ОБОСНОВАНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Л. А. Левин

В статье приводится новый вариант определения алгоритмического количества информации, введенного А. Н. Колмогоровым. Доказывается невозрастание этой величины в случайных и некоторых других процессах. Установленные свойства применяются для изучения вопросов обоснования теории вероятностей, связанных с подходом работ [1, 2].

§ 1. Введение

В работе [1] А. Н. Колмогоров предложил новый подход к определению понятия количества информации. Этот подход был основан на рассмотрении новой характеристики конечных объектов — их алгоритмической сложности. Там же А. Н. Колмогоров выдвинул программу построения на тех же идеях оснований теории вероятностей. Эта программа вызвала появление большого числа работ, развивающих и уточняющих ее идеи (см. в частности, [1–7]). При этом рассматривались только вычислимые законы теории вероятностей. Для прикладных целей это ограничение не существенно, но оно может причинить неудобство при рассмотрении некоторых теоретических вопро́сов. Кроме того, оно может вызвать неудовлетворение у некоторых «классически воспитанных» вероятностников, которым отделение вычислимых законов от законов, задаваемых иными привычными математическими конструкциями (например, предельным переходом), покажется искусственным с физической точки зрения.

В настоящей работе, в частности, показывается, что произвольный теоретико-вероятностный закон вытекает из универсального вычислимого закона Мартин-Лёфа и некоторого закона сохранения информации. Эти законы сохранения информации формулируются и изучаются в § 4 данной статьи. Они имеют очень общий характер и, по-видимому, соблюдаются не только в случайных, но и во всех вообще физических процессах (эта гипотеза обсуждается в § 4 и формулируется в виде специального тезиса). В частности, эти законы не зависят от распределения вероятностей. Поэтому их естественно вынести за скобки теории вероятностей, и тогда внутри этой теории все законы сведутся к вычислимым. В этом смысле подход работ [1, 2] покрывает классическую теорию вероятностей в полном объеме.

Содержание §§ 3, 4 — определение $I(\alpha : \beta)$, ее свойства, устанавливаемые теоремами 1–3, и тезис 1 — предствляет и самостоятельный интерес для теории информации и других математических областей. (В частности, в теории алгоритмов, по-видимому, интересно рассматривать понятие алгоритмической независимости последовательностей ($I(\alpha : \beta) < \infty$). Мы также собираемся в дальнейшем опубликовать результаты применения этого круга понятий к изучению некоторых вопросов интуиционистской логики.)

§ 2. Некоторые термины и обозначения

Статья будет понятна читателю, знакомому с основными понятиями теории алгоритмов, теории информации, теории вероятностей и, в частности, с алгоритмическим подходом к понятиям информации и случайности (см. [1–3]). Мы будем рассматривать множество S натуральных чисел и пространство Ω конечных и бесконечных последовательностей натуральных чисел. Понятно, что любые конечные объекты допускают естественную нумерацию и могут рассматриваться вместо натуральных чисел. По аналогичным соображениям вместо Ω могут рассматриваться и другие топологические пространства со счетной базой. Через Γ_x обозначим множество последовательностей, начинающихся с конечной последовательности x. Распределение вероятностей $\mathcal{P}$ на Ω называется вычислимым, если перечислимо множество пар (x, r), где x — конечная последовательность, а r — рациональное число, меньшее $\mathcal{P}(\Gamma_x)$, и множество $\{(x, r) : r > \mathcal{P}(\Gamma_x)\}$. Распределение $\mathcal{P}$ называется полувычислимым, если перечислимо только первое из этих множеств. Вычислимость относительно последовательности означает вычислимость с помощью алгоритмического оператора. Напомним также определение теста случайности по Мартин-Лёфу (см. [2]). Тестом называется множество d пар (x, m), где x — конечная последовательность, m — натуральное число. Дефектом случайности последовательности ω по тесту d называется величина $d(\omega)$, равная $\max\{m : \exists x (x, m) \in d;\ \omega \in \Gamma_x\}$. Там, где это не внесет путаницы, мы не будем различать тест d и функцию $d(\omega)$. Тест корректен относительно распределения вероятностей $\mathcal{P}$, если при всяком m имеем $\mathcal{P}\{\omega : d(\omega) \geqslant m\} \leqslant 2^{-m}$. Тест называется вычислимым, если его множество пар (x, m) перечислимо. Очевидно, что для всякого $\mathcal{P}$-корректного теста выдерживающие его последовательности ω (т. е. такие, что $d(\omega) < \infty$) образуют множество полной меры $\mathcal{P}$. Верно и обратное: для всякого множества полной меры $\mathcal{P}$ существует $\mathcal{P}$-корректный тест, который выдерживается только последовательностями из этого множества (возможно, не всеми). Таким образом, понятие $\mathcal{P}$-корректного теста соответствует понятию закона теории вероятностей (и последнее выражение мы в дальнейшем будем употреблять именно в этом смысле) для распределения $\mathcal{P}$. Мы будем писать $f \lesssim g$, если для функций f и g существует константа c, такая, что $f \leqslant g + c$. Аналогично будем понимать $f \gtrsim g$ и $f \asymp g$. Мартин-Лёф доказал, что среди вычислимых тестов, корректных относительно вычислимого распределения $\mathcal{P}$, существует универсальный $d_0(\omega)$, такой, что для любого другого d $d_0(\omega) \gtrsim d(\omega)$. Мы будем считать, что расходящиеся ряды положительных чисел принимают значение ∞. Неравенства с ∞ будем понимать естественным образом. Под парой натуральных чисел мы будем подразумевать ее номер в какой-нибудь естественной нумерации. Парой последовательностей (α, β) мы будем называть последовательность пар $\gamma_n = (\alpha_n, \beta_n)$. Изложение ведется с использованием теоретико-множественной терминологии, но может быть легко конструктивизировано.

§ 3. Определение информации

В [1] А. Н. Колмогоров ввел понятие алгоритмического количества информации. Нам придется несколько видоизменить его определение по ряду причин. Во-первых, оно не удовлетворяет с нужной нам точностью некоторым важным соотношениям (в частности, невозрастанию в простейших случайных процессах). Нарушения этих соотношений малы по сравнению с $K(x, y)$, но могут быть очень велики по сравнению с самим количеством информации в x об y. Во-вторых, это определение плохо обобщается на слу-

чай бесконечных последовательностей (что, правда, менее существенно, хотя неудобно). Поэтому мы предложим некоторое видоизменение колмогоровского определения.

В [6, 7] подробно рассматривалась величина $KP(x)$. Она отличается от колмогоровской сложности $K(x)$ тем, что декодирующий алгоритм A обладает следующим свойством «префиксности»: если $A(p_1)$ и $A(p_2)$ определены и различны, то p_1 не может быть начальным фрагментом p_2. По каждому такому префиксному алгоритму A рассматривается функция $K_A(x)$ (так же как в [1]) и среди них выбирается оптимальная с точностью до аддитивной константы. Она обозначается $KP(x)$. Через $P(x)$ мы обозначим величину $2^{-KP(x)}$ — «априорную вероятность числа x». Аналогично определяется $KP(x|\alpha)$ и $P(x|\alpha)$, где x — число, а α — последовательность натуральных чисел. Для этого вместо обычных алгоритмов рассматриваются алгоритмы, использующие в своей работе последовательность α.

Определение 1*. Количеством информации в последовательности α о последовательности β назовем величину

$$I(\alpha:\beta)=\log_2 \sum_{x,y}\left(P(x,y)\frac{P(x|\alpha)}{P(x)}\frac{P(y|\beta)}{P(y)}\right).$$

Можно доказать, что $I(\alpha:\beta)$ — невычислимая функция, даже для конечных α и β (как, впрочем, и все другие аналогичные функции). Однако, если нам про некоторые числа x и y из каких-нибудь соображений станет известна оценка снизу величины $KP(x)$ и $KP(y)$ (или сверху $P(x)$ и $P(y)$), то мы можем дать оценку $I(\alpha:\beta)$ снизу (в принципе сколь угодно хорошую). На практике оценку $KP(x)$ для некоторых x (которые предположительно могут внести большой вклад в $I(\alpha:\beta)$) можно извлекать из различных физических соображений и гипотез. Например, если x получено в случайном процессе с просто задаваемым распределением вероятностей $\mathcal{P}(x)$, то $KP(x)$ с почти единичной вероятностью близка к $|\log_2 \mathcal{P}(x)|$ (см. [5]).

Теорема 1. *Пусть α и β — конечные объекты (натуральные числа). Тогда* $I(\alpha:\beta)\asymp KP(\alpha)+KP(\beta)-KP(\alpha,\beta)$.

Отсюда следует, что если $K(\alpha,\beta)$ сравнимо с количеством информации в α о β, то $I(\alpha:\beta)$ почти не отличается от информации, определенной в [1]. Эта теорема не является тривиальной, и в доказательстве используется техника работы [7].

---

* Если вместо β рассматривать конечный объект (число), то можно выбрать определение информации более похожее на колмогоровское, а именно $[KP(x)-KP(x|\alpha)]$. Эта величина удовлетворяет законам сохранения (теоремы 2, 3 настоящей статьи).

Однако, как показал П. Гач [7], она не обладает свойством коммутативности, с которым связано свойство монотонности $(I(\alpha:(x,y))\geqslant I(\alpha:x))$, нужное для обобщения на случай последовательности β. Автору неизвестно, в какой мере окончательным является определение 1, но, по-видимому, любое хорошее выражение представимо в виде

$$\log_2 \sum_{x,y} g(x,y)\frac{R(x|\alpha)R(y|\beta)}{R(x)R(y)},$$

где x, y — двоичные слова, R — априорная вероятность слова в смысле [3], $g(x,y)$ — какое-нибудь семейство коэффициентов со сходящимся рядом $\sum_{x,y} g(x,y)$. Любая величина такого вида удовлетворяет теоремам 2 и 3. Интересен случай $g(x,y)=P(x,y)$. Существует и более простой вариант определения информации, для которого справедливы все теоремы настоящей статьи, кроме, возможно, теоремы 1. Это — $\log_2 \sum_x P(x|\alpha)P(x|\beta)/P(x)$. Не исключено, что это выражение эквивалентно тому, которое приведено в основном тексте статьи.

§ 4. Сохранение информации

Теперь мы можем сформулировать обещанные теоремы о невозрастании количества информации.

Теорема 2. *Пусть $\mathcal{P}$ — произвольное распределение вероятностей на множестве последовательностей, ρ — последовательность, относительно которой $\mathcal{P}$ вычислимо (такая последовательность всегда существует). Тогда для каждой последовательности α существует $\mathcal{P}$-корректный тест d_α, такой что при всех ω*

$$I((\rho, \omega) : \alpha) \leqslant I(\rho : \alpha) + d_\alpha(\omega).$$

Это и есть первая теорема о сохранении информации. Если на некоторой ω информация об α возрастает (по сравнению с $I(\rho : \alpha)$), то для этой ω велик дефект случайности $d_\alpha(\omega)$, и появление таких ω в случайном процессе почти невероятно. Эта теорема распространяется также на случай, когда $\mathcal{P}$ полувычислима относительно ρ.

Теорема 3. *Пусть A — произвольный алгоритмический оператор. Тогда*

$$I(A(\omega) : \alpha) \leqslant I(\omega : \alpha) + KP(A).$$

Эта теорема является следствием определения $I(\omega : \alpha)$ и показывает, что информация не возрастает и в алгоритмических процессах.

Для математиков, принимающих тезис Чёрча, теорема 3 свидетельствует о том, что информация не может возрастать ни в каких физически реализуемых детерминированных процессах. Принимая во внимание также теорему 2, можно заключить, что она не возрастает также и в любых комбинациях случайных и детерминированных физических процессов. Этот факт вместе с рядом менее четких соображений побуждает нас высказать естественно-научную гипотезу о том, что информация о заранее выбранной α не может появляться вообще ни в каком физическом процессе. Постараемся уточнить формулировку этого тезиса.

Грубо этот тезис может быть сформулирован так. С точностью до вероятности порядка 2^{-m} никакой физический процесс не может породить последовательность, содержащую более m *бит* информации об α, каким бы высокоорганизованным он ни был и как бы долго ни продолжался. Однако в этой формулировке неясно, как понимать вероятность для физического процесса произвольной природы и, кроме того, не исключена ситуация, когда α выбирается в зависимости от результатов процесса ω. Последнего легко избежать, ограничившись случаем, когда α индивидуально определена каким-нибудь конечно формулируемым математическим свойством (не обязательно допускающим алгоритмическое вычисление α).

Вероятность же можно физически интерпретировать в обычном частотном смысле, рассматривая длинную серию физических процессов. Тогда тезис будет звучать более точно:

Тезис 1. *Пусть каким-нибудь математическим свойством индивидуально задана последовательность α и пусть указана растущая серия физических процессов, каждый из которых порождает последовательность ω_i. Тогда при любом m доля тех процессов i, для которых $I(\omega_i : \alpha) \geqslant m$, по порядку не превзойдет 2^{-m}.*

Во избежание недоразумения заметим, что если α вычислима с помощью некоторого алгоритма A, то, конечно, есть физический процесс, порождающий $\omega = \alpha$. Но в этом случае $I(\omega : \alpha)$ все равно мала (ограничена сложностью A), так как для вычислимых α имеет место неравенство $I(\alpha : \alpha) \leqslant KP(A)$, следовательно, в этом случае тезис выполняется тривиальным образом. Нетривиален этот тезис, когда α — индивидуально задаваемая невычислимая последовательность — например, последователь-

ность всех истинных утверждений теории чисел в порядке возрастания длины формулировки. Тогда наш тезис противоречит убеждению некоторых математиков, что истинность любого справедливого утверждения может быть установлена в процессе развития науки с помощью неформальных методов (формальными методами этого нельзя сделать в силу теоремы Гёделя) *.

§ 5. Применение к обоснованию теории вероятностей

Рассмотрим вначале для простоты случай вычислимого распределения вероятностей $\mathcal{P}$. Для него, как показано в [2], существует универсальный вычислимый тест (см. также по этому поводу [5]). Этот тест задает множество последовательностей $\mathcal{P}$-меры 0, которое мы обозначим через $U_{\mathcal{P}}$. Кроме того, с каждой последовательностью α по теореме 2 связан тест d_α. Он также задает множество $\mathcal{P}$-меры 0, которые мы обозначим через D_α. Тогда справедлива

Теорема 4. *Пусть A — произвольное множество последовательностей $\mathcal{P}$-меры 0. Тогда существует α такая, что $A \subset (D_\alpha \cup U_{\mathcal{P}})$.*

Таким образом, оказывается, что произвольный закон теории вероятностей для вычислимого распределения покрывается универсальным вычислимым законом и законом сохранения информации.

Покажем, как этот результат обобщается на случай произвольного распределения вероятностей $\mathcal{P}$ на множестве последовательностей натуральных чисел **. Пусть ρ — последовательность, относительно которой $\mathcal{P}$ вычислима (такая ρ всегда существует). Если $I(\rho : \alpha)$ конечна, то осмысленно выражение $I((\omega, \rho) : \alpha) - I(\rho : \alpha)$. Это выражение по смыслу означает «количество информации в ω об α при известном ρ» и в силу теоремы 2 мажорируется некоторым корректным относительно $\mathcal{P}$ тестом $d(\omega)$. Если же обе величины, входящие в разность, бесконечны, то разность теряет смысл. Мы приведем технически более совершенное определение, имеющее смысл всегда. При конечном ρ оно будет равно $I((\rho, \omega) : \alpha) - I(\rho : \alpha)$ с точностью до логарфима сложности ρ.

Определение. Количеством информации в ω об α при известном ρ назовем по определению величину

$$I_\rho(\omega : \alpha) = \log_2 \sum_{x,y} \left(P(x, y|\rho) \frac{P(x|\omega, \rho)}{P(x|\rho)} \frac{P(y|\alpha, \rho)}{P(y|\rho)} \right).$$

Это выражение отличается от $I(\omega : \alpha)$ тем, что все априорные вероятности взяты при условии ρ.

По аналогии с теоремой 1 для конечных ω и α справедливо соотношение $I_\rho(\omega : \alpha) \prec KP(\omega|\rho) + KP(\alpha|\rho) - KP(\omega, \alpha|\rho)$. Аналог теоремы 3 также очевиден. Для нас будет важен аналог теоремы 2.

---

* Легко построить серию из n утверждений, для которой $KP(i_1, \ldots, i_n) \approx n$, где i_k — истинность k-го утверждения. Тогда в силу тезиса 1, какими бы методами исследования мы ни пользовались и сколько бы времени на них ни потратили, мы можем получить n правильных ответов лишь случайно и с такой вероятностью, как если бы мы решали эти вопросы с помощью бросания монеты. Вот как можно построить эти утверждения. Пусть $P(x, t_1, \ldots, t_k)$ — диофантово уравнение четвертой степени, разрешимость которого при каждом x эквивалентна тому, что x принадлежит универсальному перечислимому множеству. Пусть d — количество чисел $x < 2^n$, для которых уравнение разрешимо. Тогда k-е утверждение нашей серии будет состоять в том, что $[d/2^k]$ — четное число.

** Конечно, вместо множества последовательностей можно брать и другие топологические пространства со счетной базой (см., например, [3]).

Теорема 2′. *При любом $\mathscr{P}$, вычислимом относительно* ρ, *существует константа* c *такая, что* $(I_\rho(\omega : \alpha) - c)$ — *корректный относительно* $\mathscr{P}$ *тест, при всех* α.

С этим тестом, как и для случая вычислимой меры, связано множество последовательностей, имеющее $\mathscr{P}$-меру 0. Обозначим его $D_{\rho,\alpha}$. Так же как в случае вычислимой меры, строится универсальный, вычислимый относительно ρ корректный тест. Связанное с ним множество меры 0 обозначим $U_{\mathscr{P}}$. Тогда справедлива

Теорема 4′. *Пусть* A — *произвольное множество* $\mathscr{P}$-*меры* 0. *Тогда существует* α, *такое, что* $A \subset (D_{\rho,\alpha} \cup U_{\mathscr{P}})$.

Таким образом, программа, обещанная во введении, выполнена. Установлены законы сохранения информации и доказано, что любой другой вероятностный закон покрывается одним из этих законов и универсальным вычислимым законом. Вынося утверждения о сохранении информации за рамки теории вероятностей или признавая, что все последовательности, встречающиеся в природе, им удовлетворяют, мы получаем, что для таких последовательностей все законы теории вероятностей сводятся к универсальному вычислимому закону. Этот закон хорошо изучен (см. [2, 4, 5]).

В заключение автор выражает благодарность П. Гачу и А. К. Звонкину за обсуждение изложенных в статье вопросов.

ЛИТЕРАТУРА

1. *Колмогоров А. Н.* Три подхода к определению понятия «количество информации». Проблемы передачи информации, 1965, 1, *1*, 3–7.
2. *Martin-Löf P.* Definition of Random Sequences. Inform. and Control, 1966, 9, *6*, 602–619.
3. *Звонкин А. К., Левин Л. А.* Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов. Успехи матем. наук, 1970, 25, *6*, 85–128.
4. *Schnorr P. K.* Zufälligkeit und Wahrscheinlichkeit. Berlin, Springer Verlag, 1970.
5. *Левин Л. А.* О понятии случайной последовательности. Докл. АН СССР, 1973, 12, *3*, 548–550.
6. *Левин Л. А.* О различных видах алгоритмической сложности конечных объектов (в печати).
7. *Гач П.* О симметрии алгоритмической информации (в печати).

Поступила в редакцию
9 января 1974 г.