40 YEARS OF THE ORIGIN OF KOLMOGOROV RANDOMNESS THEORY

A.L. Semenov, An.A. Muchnik [1]

Institute of New Technologies

Введение

Разработка Колмогоровым того, что ныне называется „колмогоровской сложностью“, началась с его статьи [1]. Мы развиваем этот подход; в частности, мы решаем проблему нахождения точных оценок сложности частотных тестов случайности.

Как писал Колмогоров ([1]), „теоретико-множественная аксиоматика теории вероятностей ... разрешила большинство формальных трудностей ... настолько успешно, что проблема поиска обоснования практических применений ... стала казаться второстепенной многим исследователям.“

Мы рассматриваем простейшую ситуацию последовательности испытаний с исходами $\{0,1\}$ и хотим проверить, что цифры появляются независимо с вероятностями $1/2$. В качестве критерия этого фон Мизес ([5]) предложил близость частоты нулей к $1/2$ в самой последовательности и во всех её подпоследовательностях, выбранных допустимыми правилами. Понятие допустимого правила было уточнено Чёрчем ([6]). И фон Мизес, и Чёрч говорили о бесконечных последовательностях. Колмогоров в [1] предложил своё понятие допустимого правила для конечных последовательностей[2] and сформулировал такое (частотное) определение случайности конечной последовательности: частоты нулей нулей и единиц во всех достаточно длинных подпоследовательностях, выбранных достаточно простыми правилами близки к $1/2$. Конечно, мы должны заменить слова „достаточно длинные“, „достаточно простые“, „близки“ на

[1]Institute of New Technologies, 10 Nizhnyaya Radishevskaya, Moscow, 109004, Russia.
Phone: +7(095)9156296.
Fax: +7(095)9156963.
E-mail: alsemenov@mtu-net.ru, muchnik@lpcs.math.msu.ru

[2]Естественный аналог этого понятия для бесконечных последовательностей инвариантен относительно вычислимых перестановок цифр последовательности.

конкретные оценки (которые могут зависеть от длины исходной последовательности). Проблема из [1], которая решена авторами (Теорема 4), заключалась в нахождении точных оценок для основных параметров, при которых существуют случайные последовательности.

Обсудим язык, на котором мы будем ставить и решать интересующие нас задачи. Построение теории информации было начато Шенноном на основе теории вероятностей. Однако существуют разные ситуации, когда разумно поставить вопрос о количестве информации в конкретном объекте, но не видно, как эту информацию можно связать с каким-нибудь распределением вероятностей. Другой (комбинаторный) подход состоит в рассмотрении какого-то свойства объекта. Пусть, например, нас интересует свойство объекта $\mathbf{t}$ „быть двоичной последовательностью длины L, в которой доля нулей заключена между числами p и q“. Это свойство имеет смысл независимо от каких бы то ни было вероятностных предположений. Если L, p, q известны заранее, то неформально количеством информации в $\mathbf{t}$ можно назвать

$$\left\lceil \log_2 \left(C_L^{\lceil pL \rceil} + C_L^{\lceil pL \rceil + 1} + \ldots + C_L^{\lfloor qL \rfloor} \right) \right\rceil$$

(столько битов требуется для двоичной записи номера элемента во множестве всех объектов, удовлетворяющих указанному свойству). Заметим, что один объект обладает разными свойствами. К примеру, $\mathbf{t}$ имеет свойство „быть равным $\mathbf{t}$“. Поэтому фактически при комбинаторном подходе количество информации приписывается не объектам, а свойствам.

Принципиальным достоинством алгоритмического подхода, предложенного А. Н. Колмогоровым, является определение для каждых отдельно взятых конструктивных объектов $\mathbf{t}$, $\mathbf{s}$ количества информации $K(\mathbf{t}|\mathbf{s})$, необходимой для нахождения $\mathbf{t}$ при известном $\mathbf{s}$. Если $\mathbf{s}$ — пустое слово, мы получаем количество информации в $\mathbf{t}$ (обозначаемое $K(\mathbf{t})$). Основное соотношение для энтропии Шеннона (применяемой к распределениям вероятностей)

$$H(\langle x, y \rangle) = H(x) + H(y|x)$$

выполнено по теореме Колмогорова–Левина с очень большой точностью и для функции K (названной Колмогоровым энтропией конструктивных объектов):

$$K(\langle \mathbf{t}, \mathbf{s} \rangle) = K(\mathbf{t}) + K(\mathbf{s}|\mathbf{t}) + O(\log_2(K(\langle \mathbf{t}, \mathbf{s} \rangle))) \,. \tag{1}$$

Колмогоров доказал невычислимость функции K. Изучались и некоторые способы определения вычислимой энтропии, но для них не удалось доказать основное соотношение (1).

Как важное свойство энтропии Колмогоров отметил следующее. Пусть $\mathbf{t}$ — двоичная последовательность длины L, тогда

$$\begin{array}{c} K(\mathbf{t}|L) \text{ не намного меньше } L \\ \Downarrow \\ \text{часто́ты нулей и единиц во всех достаточно длинных подпоследовательностях, выбранных из } \mathbf{t} \text{ достаточно простыми правилами, близки к } 1/2. \end{array} \tag{2}$$

(Конечно, следует заменить выражения “не намного меньше”, “достаточно длинные”, “достаточно простые”, “близки” на конкретные оценки.) Замечательно, что импликацию (2) можно обратить (см. теорема $1'$). Таким образом, устойчивость частот при переходе к подпоследовательностям является не только необходимым, но и достаточным признаком случайности. Подчеркнём, что для получения этого результата нужно рассматривать подпоследовательности, цифры которых выбирались не обязательно в том порядке, в котором они стояли в исходной последовательности (см. теорема $3'$). Использованное нами определение немонотонных правил выбора было введено Колмогоровым в [1].

Колмогоров писал в [3]: „Теория информации должна предшествовать теории вероятностей, а не опираться на неё. Основы теории информации имеют по самому существу этой дисциплины финитный комбинаторный характер. ... Естественно, что такой подход к делу не мешает тому, чтобы теория вероятностей как часть математики развивалась как специализация общей теории меры.“ Формулировки всех теорем нашей работы строго финитны, но их доказательства (которые опубликованы в [4]), будучи столь же финитными по сути, используют для упрощения изложения вероятностные рассуждения (так же, как интегралы и другие понятия бесконечной математики).

Наша статья разделена на две части, соответствующие комбинаторному и алгоритмическому подходам к теории информации (в смысле [2]). Для каждой теоремы комбинаторной части в алгоритмической части есть теорема (имеющая тот же номер со штрихом), в которой изучается близкий вопрос, сформулированный на языке другого подхода.

Комбинаторный подход

Мы рассматриваем проблему определения меры случайности конечной двоичной последовательности относительно бернуллиевского распределения. Иногда некоторое интуитивное понятие может быть формализовано разными способами. (Например, есть различные неэквивалентные

определения размерности компакта.) Интуитивно, последовательность $\mathbf{t}$ длины L можно назвать неслучайной, если кто-нибудь, кому известно L, но неизвестно $\mathbf{t}$, укажет множество последовательностей длины L, содержащее $\mathbf{t}$ и имеющее достаточно малую меру. Такие множества называются тестами. При $\epsilon > 0$ закону больших чисел соответствует тест, состоящий из последовательностей, в которых доля нулей отстоит от $1/2$ более чем на ϵ. С точки зрения таких тестов интуитивно неслучайная последовательность $01010101\ldots$ оказывается максимально случайной. Чтобы уйти от этой трудности, Колмогоров определил понятие правила выбора подпоследовательности и каждой паре $\langle$*правило,* $\epsilon\rangle$ сопоставил тест, соответствующий закону больших чисел.

Правилом выбора r на последовательностях длины L называется пара функций $\langle r^1, r^2\rangle$, определённых на двоичных последовательностях длины от 0 до $L-1$. Значения r^1 принадлежат $\{1,\ldots,L\}$, причём на последовательности и на её собственном продолжении значения всегда различны. Значения r^2 принадлежат {“S”,“N”}. Пусть фиксирована последовательность $\mathbf{t} = t_1 \ldots t_L$. Для каждого i от 0 до L построим индукцией по i последовательности $\mathbf{u}_i$ и $\mathbf{s}_i$. Положим $\mathbf{u}_0 = \mathbf{s}_0 = \Lambda$, $\mathbf{u}_{i+1} = \mathbf{u}_i t_{r^1(\mathbf{u}_i)}$, и если $r^2(\mathbf{u}_i) =$ “N”, то $\mathbf{s}_{i+1} = \mathbf{s}_i$, а если $r^2(\mathbf{u}_i) =$ “S”, то $\mathbf{s}_{i+1} = \mathbf{s}_i t_{r^1(\mathbf{u}_i)}$.

Говорим, что правило r выбрало из $\mathbf{t}$ подпоследовательность $\mathbf{s}_L$, обозначаемую $r[\mathbf{t}]$.

Иногда полезно рассматривать более узкие классы правил. В *монотонных правилах* $r^1(\mathbf{u}_i)$ всегда равно $i+1$. *Неадаптивные правила* — такие монотонные правила, в которых $r^2(\mathbf{u}_i)$ зависит только от i.

Каждой длине L, правилу r и числу $\epsilon > 0$ сопоставляется *частотный* тест, состоящий из последовательностей $\mathbf{t}$ длины L, у которых доля нулей в $r[\mathbf{t}]$ отстоит от $1/2$ не менее чем на ϵ. Частотный тест, порождённый монотонным (неадаптивным) правилом, сам будет называться монотонным (неадаптивным).

Сводимость всякого закона теории вероятностей к закону больших чисел в самой сильной форме означала бы, что произвольный тест можно вложить в частотный тест не слишком бóльшего размера. Рассмотрим простой пример. Каждую двоичную последовательность чётной длины L будем также представлять как последовательность пар двоичных цифр длины $L/2$. Пусть $\epsilon < 1/4$; интересующий нас тест V состоит из тех последовательностей, для которых в соответствующей последовательности пар доля вхождений 00 отстоит от $1/4$ не менее чем на ϵ. То есть мы имеем дело с законом больших чисел, но для другого распределения вероятностей (бернуллиевского с четырьмя равновероятными исходами). Как доказал Вилль, этот тест не вкладывается ни в какой монотонный тест, порождённый правилом, у которого длины выборок всегда больше

$1/\epsilon$. Естественность требования, чтобы выборки были не слишком короткими, видна из самого названия закона больших чисел (например, в выборке длины единица отклонение доли нулей от $1/2$ всегда равно $1/2$).

Докажем сделанное про тест V утверждение. Пусть дано монотонное правило r. Построим по очереди цифры последовательности $\mathbf{t} = t_1 \dots t_L$. Мы используем обозначения $\mathbf{u}_i$ и $\mathbf{s}_i$ из определения $r[\mathbf{t}]$. Предположим, что $\mathbf{u}_i$ уже построено. Если $r^2(\mathbf{u}_i) =$ "N", то $t_{i+1} = 1$. Если $r^2(\mathbf{u}_i) =$ "S", то t_{i+1} отличается от последней цифры $\mathbf{s}_i$ и $t_{i+1} = 1$, когда $\mathbf{s}_i = \Lambda$.

С одной стороны, в последовательности пар, соответствующей построенному $\mathbf{t}$, нет ни одного вхождения 00 (то есть с точки зрения закона больших чисел для четырёх исходов $\mathbf{t}$ очень неслучайна). С другой стороны, выборка $r[\mathbf{t}]$ имеет вид $1010101\dots$ (то есть случайна с точки зрения закона больших чисел для двух исходов).

Однако уже двух монотонных тестов (не слишком больших по размеру) достаточно, чтобы вложить V в их объединение. В дальнейшем построении обозначим эти тесты F_1 и F_2. Тест F_1 является даже неадаптивным. Порождающее его правило r_1 включает в выборку все цифры, стоящие на нечётных местах (то есть $r_1^2(\mathbf{u}_i) =$ "S" $\Leftrightarrow$ i чётно). Монотонное правило r_2 включает в выборку цифры, стоящие на чётных местах, у которых предшествующая цифра равна нулю (то есть $r_2^2(\mathbf{u}_i) =$ "S" $\Leftrightarrow$ [i нечётно и $t_i = 0$]). Для того, чтобы обеспечить достаточно большую длину всех выборок, рассмотрим следующее монотонное правило $\bar{r}_2$. Пусть $\nu(\mathbf{u})$ — количество нулей, стоящих на нечётных местах в слове $\mathbf{u}$. Тогда $\bar{r}_2^2(\mathbf{u}_i) =$ "S" $\Leftrightarrow$ [$r_2^2(\mathbf{u}_i) =$ "S" или [i нечётно и $\nu(\mathbf{u}_i)+(L-i)/2 \le L/5$]].

Как показывает простая проверка, если паре $\langle r_1, \epsilon(1-\epsilon)\rangle$ сопоставить тест F_1, а паре $\langle \bar{r}_2, \epsilon(1-\epsilon)\rangle$ сопоставить тест F_2, то $V \subseteq F_1 \cup F_2$. При этом для $L > 34/\epsilon^3$

$$|F_1 \cup F_2| < 2^L \cdot \mathrm{e}^{-L\epsilon^2/5}\,.$$

Отметим, что тест, соответствующий закону больших чисел для последовательности длины L, тоже имеет размер порядка $2^{(1-\delta)L}$.

Теперь построим один немонотонный тест G, небольшой по размеру и содержащий V. Неформально говоря, правило r, порождающее тест G, состоит в последовательном применении правил r_1 и r_2. Первая компонента правила r зависит только от длины аргумента. А именно,

$$r^1(\mathbf{u}) = \begin{cases} 2\ell(\mathbf{u}) + 1 & \text{при } \ell(\mathbf{u}) < L/2\,, \\ 2\ell(\mathbf{u}) - L + 2 & \text{при } \ell(\mathbf{u}) \ge L/2\,. \end{cases}$$

При $\ell(\mathbf{u}) < L/2$ всегда $r^2(\mathbf{u}) =$ "S". Пусть $\ell(\mathbf{u}) \ge L/2$. Обозначим через $\mathbf{w}$ начало длины $L/2$ слова $\mathbf{u}$. Если доля нулей в $\mathbf{w}$ отклоняется от $1/2$

более чем на $\epsilon/8$, то $r^2(\mathbf{u}) =$ "N". Иначе полагаем $r^2(\mathbf{u}) =$ "S" $\Leftrightarrow$ цифра $\mathbf{u}$ с номером $(\ell(\mathbf{u}) - L/2 + 1)$ равна нулю.

Ясно, что длина $r[\mathbf{t}]$ всегда не меньше $L/2$. Просто проверить, что если тест G сопоставить паре $\langle r, \epsilon/8\rangle$, то $V \subseteq G$. При этом для $L > 256/\epsilon^3$

$$|G| < 2^L \cdot \mathrm{e}^{-L\epsilon^2/128}\,.$$

Прежде чем от рассмотренного примера перейти к произвольным тестам мы должны выбрать удобную шкалу для измерения "малости" теста. (Выбор подходящей шкалы может делать математические утверждения более наглядными, а иногда и более информативными. Например, в центральной предельной теореме естественной единицей измерения для отклонения частоты от вероятности является [количество испытаний]$^{-1/2}$.)

Удельным дефектом теста U, состоящего из последовательностей длины L, называется величина $1 - \frac{\log_2 |U|}{L}$.

Заметим, что эта же единица измерения используется во многих результатах теории кодирования.

В новых терминах можно сказать, что для больших L тест V из примера Вилля имеет удельный дефект приблизительно $\frac{4}{3\ln 2}\epsilon^2$, а покрывающий его частотный тест G имеет удельный дефект больше $\frac{1}{128\ln 2}\epsilon^2$.

Постановка разбираемых далее задач естественно вписывается в контекст комбинаторной математики, где исследуются такие вопросы, как "*Чему равен минимальный диаметр шара, которым можно покрыть произвольное множество диаметра* 1*?*", "*Каково минимальное количество шаров диаметра* 1*, которыми можно покрыть произвольное множество диаметра* 1*?*" и т. д.

Авторами доказан следующий результат (см. [4, теорема 4]).

Теорема 1. *Для каждого достаточно малого положительного δ можно указать натуральное число $R(\delta)$, для которого выполнено следующее. Для любого теста U из последовательностей длины $L > (1/\delta)^5$ если удельный дефект U больше δ, то существует $R(\delta)$ монотонных частотных тестов, объединение которых покрывает U и имеет удельный дефект больше*

$$\frac{\delta}{\ln(1/\delta)}(1 - \beta(\delta))\,,$$

где $\beta(\delta) \to 0$ при $\delta \to 0$.

Как видно из формулировки, количество тестов в покрытии не зависит от L. Конструкция Вилля из [7] (упрощённый вариант которой был разобран нами выше) показывает, что теорему 1 нельзя усилить, заменив $R(\delta)$ на константу.

Открытая проблема: сохранится ли утверждение теоремы 1, если заменить $R(\delta)$ на константу и убрать требование монотонности частотных тестов?

Авторы доказали, что оценка на удельный дефект покрытия в теореме 1 близка к точной (см. [4, теорема 5]).

Теорема 2. *Для каждого достаточно малого положительного δ можно указать натуральное число $R(\delta)$, для которого выполнено следующее. Для любого $L > (1/\delta)^5$ существует такой тест U из последовательностей длины L, что удельный дефект U больше δ и всякое покрытие U частотными тестами с удельным дефектом больше*

$$\frac{2\delta}{\ln(1/\delta)}$$

состоит более чем из $\mathrm{e}^{L/R(\delta)}$ тестов.

Отметим, что в теореме 2 даже нет требования монотонности частотных тестов.

Авторами доказано, что теорему 1 нельзя усилить, заменив требование монотонности частотных тестов на требование неадаптивности (см. [4, теорема 6]).

Теорема 3. *Для любого $\sigma > 0$ и для любого $L > 12 + 6\log_2(1/\sigma)$ существует такой тест U из последовательностей длины L, что удельный дефект U не меньше $1/3$ и всякое покрытие U неадаптивными частотными тестами с удельным дефектом не меньше σ состоит не менее чем из $2^{\sigma L/2}$ тестов.*

Обратимся теперь к вопросу о датчиках (таблицах) случайных чисел, который был поставлен в статье Колмогорова [1].

Пусть $\mathcal{R}_L$ — некоторое множество правил на последовательностях длины L. Последовательность $\mathbf{t}$ длины L называется *(n,ϵ)-датчиком случайных чисел для $\mathcal{R}_L$*, если каждое правило из $\mathcal{R}_L$ выбирает в $\mathbf{t}$ подпоследовательность $r[\mathbf{t}]$ с таким свойством:

- при условии, что длина подпоследовательности не меньше n, доля нулей в ней отличается от $1/2$ менее чем на ϵ.

Каждой длине L и правилу r сопоставляется *(n,ϵ)-тест*, состоящий из последовательностей $\mathbf{t}$ длины L, у которых $\ell(r[\mathbf{t}]) \geq n$ и доля нулей в $r[\mathbf{t}]$ отстоит от $1/2$ не менее чем на ϵ.

Колмогорова интересовало, для какого количества правил заведомо существует (n,ϵ)-датчик. Переведём это на язык покрытий: каким количеством (n,ϵ)-тестов можно покрыть множество всех последовательностей длины L? Колмогоров доказал в [1], что при естественных ограничениях на n, ϵ это количество меньше $\mathrm{e}^{2n\epsilon^2(1-\epsilon)}$ и больше $2^{4n\epsilon(1+5\epsilon)}$ (точные

формулировки и полные доказательства см. [4, теоремы 1,2]). Поставленная Колмогоровым проблема состояла в устранении разрыва между степенями ϵ в приведённых оценках. Авторы доказали, что верхняя оценка Колмогорова близка к точной (см. [4, теорема 3]).

Теорема 4. *Пусть n — целое положительное число и $\epsilon \in (0, 1/3)$. Если $2n \le L \le 2^{n\epsilon^3/2}$, то множество всех последовательностей длины L можно покрыть (n, ϵ)-тестами в количестве*

$$\mathrm{e}^{2n\epsilon^2(1+\epsilon)/(1-n/(L-1))}. \tag{3}$$

Наличие в последней теореме условия $L \ge 2n$ и члена $(1 - n/(L-1))$ в формуле (3) неудивительно, так как всякая последовательность с одинаковым количеством нулей и единиц не принадлежит ни одному (n, ϵ)-тесту при $L < n(1 + 2\epsilon)$.

Следует заметить, что Колмогоров для своей нижней оценки явно построил покрывающее семейство тестов. Наше же доказательство вероятностное. Ситуация, когда вероятностное доказательство даёт лучшую оценку, чем все известные явные построения, далеко не в первый раз встречается в теории информации (например, теорема Шеннона о помехоустойчивом кодировании).

Сейчас мы опишем в самом общем виде полезный в теории информации приём, применённый в теореме 4. Рассмотрим конечное множество A и семейство B подмножеств A (обозначим $a = |A|$, $b = |B|$). Пусть каждый элемент из A принадлежит по крайней мере $c > 0$ множествам из B. Тогда в B существует подсемейство мощности $\lceil (b/c) \ln a \rceil$, покрывающее всё A. Выберем случайно (относительно равномерного на B распределения) элемент из B, и повторим эту операцию независимо $\lceil (b/c) \ln a \rceil$ раз (в выборке могут оказаться повторения). Фиксируем $x \in A$. Вероятность того, что один случайный элемент из B покроет x, не меньше c/b. Вероятность противоположного события не больше $1 - c/b$. Вероятность того, что ни один из элементов выборки не покроет x, не больше $(1 - c/b)^{\lceil (b/c) \ln a \rceil}$. Вероятность того, что хотя бы один элемент из A не покрыт множествами из выборки не больше $a \cdot (1 - c/b)^{(b/c) \ln a} < 1$, поскольку $(1 - c/b)^{b/c} < \mathrm{e}^{-1}$. Тем самым вероятность того, что выборка покроет всё A, положительна, и следовательно, искомое подсемейство в B существует.

Алгоритмический подход

В алгоритмическом подходе[3] каждой паре L, δ сопоставляется ровно один (в отличие от комбинаторного подхода) тест $U_{L,\delta}$ из последовательностей длины L, имеющий удельный дефект больше δ. Он состоит из всех последовательностей длины L, у которых при известном L условная энтропия меньше $L(1-\delta)$. Неформально говоря, $U_{L,\delta}$ содержит последовательности, которые могут быть сжаты без потери информации в $(1-\delta)$ раз.

Мы должны ввести понятия *алгоритмического правила выбора* и *алгоритмического частотного теста*, которые являются алгоритмическими аналогами для понятий правила выбора и частотного теста. Необходимость этих новых понятий будет мотивирована ниже. Алгоритмическим правилом выбора r на последовательностях длины L называется программа, вычисляющая пару частичных функций $\langle r^1, r^2\rangle$, определённых на двоичных последовательностях длины от 0 до $L-1$. Значения r^1 принадлежат $\{1,\ldots,L\}$, причём на последовательности и на её собственном продолжении значения всегда различны. Значения r^2 принадлежат {“S”,“N”}. Пусть фиксирована последовательность $\mathbf{t} = t_1 \ldots t_L$. Последовательности $\mathbf{u}_i$ и $\mathbf{s}_i$ строятся так же, как и в определении (не алгоритмического) правила выбора. Если для какого-нибудь i значение $r^1(\mathbf{u}_i)$ или $r^2(\mathbf{u}_i)$ неопределено, то подпоследовательность $r[\mathbf{t}]$ тоже неопределена, иначе $r[\mathbf{t}] = \mathbf{s}_L$.

Каждой длине L, алгоритмическому правилу r и $\epsilon > 0$ сопоставляется алгоритмический частотный тест, состоящий из последовательностей $\mathbf{t}$ длины L, для которых $r[\mathbf{t}]$ определено и доля нулей в $r[\mathbf{t}]$ отстоит от $1/2$ не менее чем на ϵ.

В алгоритмическом подходе для каждой пары R, σ исследуется вопрос о том, покрывается ли $U_{L,\delta}$ фиксированным (в отличие от комбинаторного подхода) семейством $\mathcal{F}_{L,R,\sigma}$. Это семейство состоит из всех алгоритмических частотных тестов с удельным дефектом больше σ, порождённых алгоритмическими правилами, у которых при известном L условная энтропия меньше $\log_2 R$ (количество таких правил меньше R).

Если бы тот же вопрос о покрытии исследовался для (не алгоритмических) частотных тестов, то во всех интересных случаях ответ оказался бы отрицательным. Действительно, частотные тесты — это конструктивные объекты. Рассмотрим частотные тесты G, которые состоят из последовательностей длины L, имеют удельный дефект больше σ и $K(G|L) < \log_2 R$. Пусть даны числа L, R, σ и известно количество соот-

[3]Переменные, которые в комбинаторном подходе принимали вещественные значения, в алгоритмическом подходе принимают только рациональные значения.

ветствующих частотных тестов G. Если существует последовательность $\mathbf{t}$ длины L, не принадлежащая объединению этих частотных тестов, то её можно найти перебором. Условная энтропия $\mathbf{t}$ при известном L не зависит от L, следовательно, при больших L тест $U_{L,\delta}$ содержит $\mathbf{t}$.

Доказательства следующих двух наших результатов см. [4, теоремы $4', 5'$].

Теорема 1′. *Для каждого достаточно малого $\delta > 0$ существует такое L_0, что для любого натурального $L > L_0$, $R = (1/\delta)^{4\ln 2/\delta}$ и $\sigma = \frac{\delta}{\ln(1/\delta)}(1-\beta(\delta))$, где $\beta(\delta)$ — фиксированная функция, стремящаяся к 0 при $\delta \to 0$, выполнено*

$$U_{L,\delta} \subseteq \bigcup \mathcal{F}_{L,R,\sigma}\,.$$

Теорема 2′. *Для каждого достаточно малого $\delta > 0$, натурального $L > (1/\delta)^5$, $R = 2^{L\delta^4/70}$ и $\sigma = \frac{2\delta}{\ln(1/\delta)}(1+3\delta)$ выполнено*

$$U_{L,\delta} \not\subset \bigcup \mathcal{F}_{L,R,\sigma}\,.$$

Обратим внимание, что в теореме $1'$, в отличие от теоремы 1, частотные тесты из покрывающего семейства порождаются не только монотонными правилами. Это обстоятельство неустранимо, как показывает следующий доказанный авторами результат (см. [4, теорема $6'$]).

Обозначим через $\mathcal{M}_{L,R,\sigma}$ семейство всех алгоритмических частотных тестов с удельным дефектом больше σ, порождённых алгоритмическими монотонными правилами, у которых при известном L условная энтропия меньше $\log_2 R$.

Теорема 3′. *Для каждого достаточно большого натурального L, числа $\delta = 1/\sqrt{L}$ и $R = 2^{\sqrt{L}/4}$ выполнено*

$$U_{L,1/2} \not\subset \bigcup \mathcal{M}_{L,R,\delta}\,.$$

Проблема Колмогорова из [1] при алгоритмическом подходе решается так (см. [4, теорема $3'$]).

Теорема 4′. *Пусть n — целое положительное число и рациональное $\epsilon \in (0, 1/3)$. Если $2n \le L \le 2^{n\epsilon^3/2}$, то множество всех последовательностей длины L можно покрыть (n,ϵ)-тестами, порождёнными неадаптивными правилами, у которых при известных L, n, ϵ условная энтропия меньше*

$$(2/\ln 2)\, n\epsilon^2 \frac{1+\epsilon}{1-n/L} + C\,,$$

где C — константа, зависящая только от выбора оптимального языка программирования.

Заключение

Как известно, Андрей Николаевич Колмогоров в России и Рэй Соломонофф в США независимо пришли к основным понятиям и фактам, касающимся сложности конечных объектов, в середине 60-тых годов. В 60-е и 70-е годы исследования в этой области велись в основном учениками Колмогорова и членами московского математического сообщества, обосновавшегося в Московском государственном университете (им. Ломоносова): Л. Левиным, В. Вьюгиным, А. Звонкиным, Н. Петри, П. Гачем. Г. Чэйтин внёс существенный вклад в популяризацию этой теории на западе; повлияла на исследования в России и рукопись Соловэя.

С конца 70-х годов до своей смерти в 1987 году Колмогоров был заведующим кафедрой математической логики. Сейчас её возглавляет его бывший студент Владимир Андреевич Успенский, который также работает в области колмогоровской сложности. В самом начале своей работы в качестве заведующего нашей кафедрой профессор Колмогоров предложил первому из авторов настоящей статьи начать семинар по сложности. (Сейчас он называется колмогоровским семинаром по сложности определений и вычислений.) С первых шагов работы семинара сложилась традиция заслушивать на нём иссследовательские сообщения и обзоры ещё и по практическому программированию и сложности алгоритмов. В своих первых докладах на семинаре Колмогоров представил программу исследований по частотному подходу к случайности и другим темам. Эта программа реализовывалась в работах В. Вовка, А. Шеня, Н. Верещагина, Ан. Мучника и других. В последующие годы к руководству семинаром присоединились А. Шень и Н. Верещагин, ставшие его основной движущей силой после смерти Колмогорова. Большой вклад в развитие семинара внёс профессор Успенский. В пленарном докладе Колмогорова и Успенского на бернуллиевском конгрессе были подведены итоги работы московской группы за всё то время, что её возглавлял Колмогоров.

Основное содержание настоящей статьи было доложено на колмогоровском семинаре весной 2002 года.

Acknowledgments

Авторам очень важно общение с участниками Колмогоровского семинара. Особую признательность мы выражаем Николаю Константиновичу Верещагину, интересные обсуждения с которым вопросов колмогоровской теории послужили для нас одним из стимулов начать настоящую работу.

Список литературы

[1] A. N. Kolmogorov. On tables of random numbers. *Sankhyã: The Indian Journal of Statistics*, Series A, 1963, v. 25, part 4. (Reprinted in *Theoretical Computer Science*, 1998, vol. 207, pp. 387–395).

[2] А. Н. Колмогоров. Три подхода к определению понятия "количество информации". *Проблемы передачи информации*, 1965, т. 1, N 1, с. 3–11.

[3] А. Н. Колмогоров. Комбинаторные основания теории информации и исчисления вероятностей. *Успехи математических наук*, 1983, т. 38, N 4, с. 27–36.

[4] Ан. А. Мучник, А. Л. Семенов. О роли закона больших чисел в теории случайности. *Проблемы передачи информации*, 2003, т. 39, N 1, сс. 134–165.

[5] R. von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 1919, v. 5, pp. 52–99.

[6] A. Church. On the concept of a random sequence. *Bulletin of American Mathematical Society*, 1940, v. 46, pp. 130–135.

[7] J. Ville. *Étude Critique de la Notion de Collectif*. Gauthier-Villars, 1939.