О роли закона больших чисел в теории случайности*

Алексей Семенов† Андрей Мучник‡

Институт новых технологий,
109004, Москва, ул. Нижняя Радищевская, 10.

Аннотация

В первой части статьи решена проблема уточнения условий существования датчика случайных чисел, поставленная А. Н. Колмогоровым в 1963 году в [1].

Колмогоровская теория сложности впервые позволила строго определить понятие случайности индивидуальной последовательности нулей и единиц. При этом для бесконечных последовательностей речь идёт о двузначном свойстве: последовательность случайна или последовательность неслучайна, в то время как для конечных последовательностей можно говорить только о непрерывном свойстве — мере их случайности.

Можно ли мерить случайность последовательности $\mathbf{t}$ по тому, насколько выполнен закон больших чисел во всех подпоследовательностях, полученных из $\mathbf{t}$ „допустимым способом“? Ситуация для бесконечных последовательностей была изучена в [7].

В качестве меры случайности (а точнее, неслучайности) конечной последовательности мы рассматриваем удельный дефект случайности δ (определение 5). Во второй части настоящей работы показано, что функция $\delta/\ln(1/\delta)$ характеризует связь между случайностью конечной последовательности и выполнением закона больших чисел.

*Работа выполнена при поддержке Российского фонда фундаментальных исследований, гранты N 01-01-00505, N 02-01-10904, N 02-01-22001.

†e-mail: alsemenov@mtu-net.ru, fax: +(095)9156963.

‡e-mail: muchnik@lpcs.math.msu.ru, fax: +(095)9156963.

Ключевые слова: сложность описания, колмогоровская энтропия, датчик случайных чисел, допустимое правило выбора, устойчивость частот, дефект случайности.

Введение

В 1930-е годы Андрей Николаевич Колмогоров обосновал теорию вероятностей с помощью теории меры. В [1] он пишет об этом: „теоретико-множественное обоснование теории вероятностей ... оказалось настолько эффективным с математической точки зрения и с точки зрения практических приложений, что многие исследователи стали относиться к философскому обоснованию связи между математическими результатами теории вероятностей и их практическими применениями как к чему-то второстепенному“.

Однако сам Колмогоров считал вопрос об этом обосновании принципиальным. В 1962 году, во время пребывания в Индии, он начал разрабатывать новый подход к нему[1] — теорию сложности описания. За прошедшее время начатые Колмогоровым исследования выросли в очень богатую теорию, имеющую важные связи не только с теорией вероятностей, но и с теорией алгоритмов, теорией кодирования, теорией матроидов и другими областями математики.

Что касается практических применений, то здесь основные результаты, вероятно, ещё впереди. Для их получения требуется серьёзно учитывать не только сложность описания программы, но и объём используемых ею ресурсов. Этот учёт, во многих случаях, связан с нерешёнными проблемами теории сложности вычислений.

Рассмотрим последовательность независимых испытаний с двумя равновероятными исходами 0 и 1. Простейший и в то же время важнейший признак случайности для последовательности исходов — приблизительное равенство количества нулей и количества единиц. Очевидно, одного этого требования недостаточно (например, последовательность 0101010101 ... по нашим представлениям неслучайна). Однако если *указанный признак выполняется для всех последовательностей, получаемых из исходной с помощью „допустимых“ правил выбора*, то такую последовательность уже можно рассматривать как датчик случайных

[1] Первая публикация [1] появилась в 1963 году в индийском журнале Sankhyã: The Indian Journal of Statistics.

чисел. Разумеется, понятие допустимого правила должно быть математически уточнено[2].

Является ли сформулированный частотный критерий случайности универсальным? Ряд известных фактов о бесконечных последовательностях скорее свидетельствовал об обратном. К примеру (см. [5]),

(i) существует такое множество S бесконечных двоичных последовательностей меры 0, что для любого счётного семейства R допустимых правил выбора найдётся $s \in S$, для которого закон больших чисел выполняется во всех бесконечных подпоследовательностях, выбранных из s с помощью правил, принадлежащих R.

При этом построение множества S и его свойство иметь меру нуль достаточно конструктивны, и следовательно, все его элементы интуитивно неслучайны.

Колмогоров придавал особое значение анализу конечных последовательностей, а не только предельных закономерностей. Все результаты нашей статьи лежат в конечной области. Аналог (i) для конечных последовательностей мог бы звучать так:

(ii) Пусть фиксировано натуральное L. Существует такое множество S двоичных последовательностей длины L, что мощность S достаточно мала по сравнению с 2^L и для любого не слишком большого семейства R допустимых правил выбора найдётся $s \in S$, для которого закон больших чисел выполняется с достаточной точностью во всех не слишком коротких подпоследовательностях, выбранных из s с помощью правил, принадлежащих R.

Разумеется, последняя формулировка нуждается в уточнении. Что значит „достаточно мала“, „не слишком большого“, „с достаточной точностью“, „не слишком коротких“? Замечательно, что существует естественное уточнение этих параметров, при котором истинно отрицание (ii). Тем самым удаётся установить положительную связь между *частотной* и *универсальной* случайностью (теорема 4).

[2]Пример допустимого правила — выбрать цифры, стоящие на чётных местах. Пример недопустимого правила — выбрать цифры, стоящие на тех местах, где в исходной последовательности нули.

1. Анализ статьи Колмогорова 1963 года „О таблицах случайных чисел“

1.1. Философская мотивировка

По словам Колмогорова из [1], в течение долгого времени он полагал, что:

1) предельные теоремы, описывающие бесконечные последовательности испытаний, не имеют отношения к практике;

2) связь с практикой вероятностных теорем о конечных последовательностях испытаний нельзя формализовать чисто математически.

Мнение Колмогорова в отношении первого тезиса не изменилось[3]. По поводу второго тезиса Колмогоров пришёл к выводу, что, используя частотную концепцию случайности и уточнение понятия сложности программы, можно чисто математически сформулировать критерии применимости теории вероятностей к практике.

В работе [1] сообщается, что строгое определение сложности программы будет дано в другой статье. При этом в [1] используется только тот факт, что количество простых объектов не может быть слишком велико. Такой подход позже Колмогоров назвал комбинаторным (см. [2]). Определения, соответствующие алгоритмическому подходу, были даны Колмогоровым в 1965 году (см. [2]). Результатам из [1] можно сопоставить естественные алгоритмические аналоги, и мы их приводим. Параллель между комбинаторными и алгоритмическими формулировками распространяется очень широко. Поэтому в нашей статье почти все теоремы сформулированы в двух вариантах — комбинаторном и алгоритмическом — и имеют параллельную нумерацию (алгоритмические варианты имеют номер со штрихом).

Конечная двоичная последовательность рассматривается Колмогоровым как датчик (таблица) случайных чисел, если во всех достаточно длинных подпоследовательностях, полученных с помощью не слишком сложных правил выбора, частота нулей и единиц мало отличается от $\frac{1}{2}$.

[3]Хотя в других работах Колмогоров подчёркивал, что исследование бесконечных последовательностей имеет большое эвристическое значение.

Математическая цель статьи [1] — оценить, насколько большую сложность можно допустить для правил выбора, чтобы заведомо существовал датчик случайных чисел.

1.2. Определения

В работе [1] Колмогоров определил понятие (немонотонного) правила выбора из конечной двоичной последовательности $\mathbf{t}$.

Неформальное описание такое. Представим, что имеется набор карт в количестве, равном длине $\mathbf{t}$. Цифры последовательности $\mathbf{t}$ нарисованы на лицевой стороне карт, оборотная сторона у всех карт одинакова. Вначале карты лежат лицевой стороной вниз в том порядке, в котором идут цифры в $\mathbf{t}$. Правило решает, какую карту перевернуть и (ещё до того, как эта карта перевёрнута) включить ли цифру, нарисованную на этой карте, в подпоследовательность. При принятии очередного решения правило может учитывать цифры, нарисованные на ранее перевёрнутых картах. Цифры, выбранные правилом, располагаются в подпоследовательности в том порядке, в котором они выбирались, а не в порядке следования в исходной последовательности.

Определение 1. *Правилом на последовательностях длины* L называется функция r, определённая на двоичных последовательностях длины от 0 до $L-1$, значениями которой являются пары из множества $\{1,\ldots,L\} \times \{\text{„включить“}, \text{„не включить“}\}$, причём первые компоненты значений функции на последовательности и на её собственном продолжении всегда различны. Пусть фиксирована последовательность $\mathbf{t}$ длины L. Для каждого i от 0 до L построим индукцией по i последовательность $\mathbf{s}_i$ двоичных цифр длины i. Положим $s_0 = \Lambda$, а s_{i+1} получается из s_i присоединением в конце цифры последовательности $\mathbf{t}$ с номером $\pi_1(r(\mathbf{s}_i))$. Подпоследовательность, выбранная правилом r из $\mathbf{t}$, получается, если из $\mathbf{s}_L$ удалить цифры с номерами i, для которых $\pi_2(r(\mathbf{s}_{i-1})) = \text{„не включить“}$. (Здесь π_1, π_2 обозначают первый и второй члены пары.)

Подпоследовательность, выбранная правилом r из последовательности $\mathbf{t}$, будет обозначаться $r[\mathbf{t}]$.

Иногда полезно рассматривать более узкие классы правил. *Монотонные правила* всегда переворачивают карты в их исходном порядке[4].

[4]Монотонные правила впервые рассмотрел Чёрч в 1940 году.

Неадаптивные правила сразу указывают некоторое множество карт, и подпоследовательность состоит из цифр, нарисованных на этих картах и расположенных в исходном порядке.

Определение 2. 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 $\frac{1}{2}$ менее чем на ε.

Абсолютная величина разности между $\frac{1}{2}$ и долей нулей в последовательности называется *отклонением* (доли нулей).

Заметим, что для каждого $p \in [0, 1]$ можно рассматривать (n, ε, p)-датчики, для которых доля нулей в подпоследовательностях должна быть близка не к $\frac{1}{2}$, а к p.

Как перевести на строгий язык слова о не слишком сложных правилах выбора?

При комбинаторном подходе *сложностью* конечного множества называется двоичный логарифм его мощности.

В алгоритмическом подходе мы будем использовать *простую энтропию* конструктивного объекта, которая была определена Колмогоровым в 1965 году[5]. Фактически был определён класс функций энтропии, каждые две из которых различаются не более чем на константу. При этом Колмогоров надеялся, что удастся предложить какой-то „естественный“ язык программирования, для которого функция энтропии не более чем на 100 превышала бы функцию энтропии любого другого „естественного“ языка программирования. Так как наши алгоритмические результаты в равной степени применимы к любой функции энтропии, в их формулировках может присутствовать константа C, зависящая только от выбора языка программирования. Отметим, что будет использоваться и понятие *условной* энтропии, также введённое Колмогоровым.

Аналогия между комбинаторным и алгоритмическим подходами основана на следующих двух утверждениях, доказанных Колмогоровым.

[5]Сравнение простой энтропии с префиксной энтропией и другими вариантами определения можно найти в [6].

Утверждение 1. *Количество объектов, условная энтропия которых относительно фиксированного объекта меньше m, не превышает $2^m - 1$.*

Утверждение 2. *Если множество перечисляется программой энтропии меньше α и имеет мощность меньше 2^m, то энтропия его элементов меньше $m + \alpha + 2\operatorname{lb}\alpha + C$. (Здесь и в дальнейшем выражение* lb *будет обозначать двоичный логарифм.)*

Закон больших чисел влечёт, что множество последовательностей, из которых фиксированное правило выбирает длинную подпоследовательность с большим отклонением, имеет малую мощность. Однако если мы удалим требование большого отклонения, мощность всё равно может остаться малой. В качестве примера можно рассмотреть монотонное правило, которое включает цифры в подпоследовательность, если уже известные цифры образуют начало последовательности $0101010101\ldots$; при длине выборки равной длине последовательности и без ограничений на отклонение, соответствующее множество содержит всего два элемента.

Для преодоления этой трудности введём понятие нормального правила и будем рассматривать только множества, порождённые такими правилами.

Определение 3. Правило r, действующее на последовательностях длины L, назовём *нормальным*, если из всех последовательностей r выбирает подпоследовательности одной и той же положительной длины.

Для нормального правила r, выбирающего подпоследовательности длины n, и числа $\varepsilon \in [0, \frac{1}{2}]$, определим множество $A_{r,\varepsilon}$ последовательностей, в которых r выбирает подпоследовательности с отклонением не менее ε. Такие множества назовём *правильными*.

Введём обозначение, которое неоднократно понадобится нам:

$$d(n, \varepsilon) \rightleftharpoons 2n\varepsilon^2 \operatorname{lb} \mathrm{e}\,.$$

Следующие факты (подробно обсуждаемые в дальнейшем) мотивируют введённое обозначение. Если нормальное правило r выбирает подпоследовательности длины n, то мощность правильного множества $A_{r,\varepsilon}$ зависит только от L, n и ε. Когда ε достаточно мало, а n достаточно велико по сравнению с $1/\varepsilon$, то $|A_{r,\varepsilon}| \approx 2^{L-d(n,\varepsilon)}$.

Существует эффективная операция (назовём её *n-нормализацией*), превращающая каждое правило в нормальное с длиной выборки n, причём так, что выборки, уже имевшие длину ровно n, остаются неизменными.

Опишем процедуру n-нормализации. Новое правило действует как и старое, но с двумя исключениями. Если старое правило уже выбрало n цифр в подпоследовательность, то в новом правиле на этом выбор прекращается. Если старое правило уже выбрало $k < n$ цифр и остались неизвестными $n - k$ цифр, то новое правило выбирает их все.

1.3. Достаточные условия и необходимые условия существования датчика случайных чисел

Теорема 1 (Колмогоров, 1963). *Рассмотрим произвольные числа L (длина последовательности), $\varepsilon > 0$ (отклонение частоты от вероятности) и $n \geqslant \varepsilon^{-4}$ (длина выборки). Для любого множества правил $\mathcal{R}_L$, сложность которого меньше*

$$d(n, \varepsilon)(1 - \varepsilon)\,,$$

существует (n, ε)-датчик случайных чисел.

Доказательство. В статье [1] Колмогоров только наметил доказательство этой теоремы, и для полноты изложения мы его приведём полностью.

Очевидно, можно считать, что $\varepsilon \leqslant \frac{1}{2}$. На последовательностях длины L рассмотрим равномерную бернуллиевскую меру (все цифры независимы и с вероятностью $\frac{1}{2}$ равны 0 или 1). Вероятность каждой последовательности равна 2^{-L}.

Зафиксируем некоторое правило выбора r. Обозначим k-нормализацию правила r через r_k. Заметим, что для r_k вероятность выбрать данную подпоследовательность длины k, равна 2^{-k}, поскольку независимость цифр сохраняется при изменении порядка их расстановки.

Оценим вероятность того, что последовательность $\mathbf{t}$ не является (n, ε)-датчиком для правила r (то есть длина выборки больше или равна n, а отклонение больше или равно ε). Если мы рассмотрим наименьшее $k \geqslant n$, для которого отклонение в начале длины k выборки $r[\mathbf{t}]$ больше или равно ε, то станет понятным, что

$$\Pr\{\mathbf{t} \text{ не } (n,\varepsilon)\text{-датчик для } r\} \leqslant \Pr\{\text{отклонение в } r_n[\mathbf{t}] \ \geqslant \varepsilon\} + \\ 2 \cdot \sum_{k=n}^{L} \Pr\left\{\text{количество нулей в } r_k[\mathbf{t}] \ = \left\lceil k\left(\frac{1}{2}+\varepsilon\right)\right\rceil\right\} = \\ 2 \cdot \sum_{j=\lceil n(\frac{1}{2}+\varepsilon)\rceil}^{n} \binom{n}{j} \cdot 2^{-n} + 2 \cdot \sum_{k=n}^{L} \binom{k}{\lceil k(\frac{1}{2}+\varepsilon)\rceil} \cdot 2^{-k}.$$

Для $j \neq k$ используем следующую оценку, вытекающую из формулы Стирлинга:

$$\binom{k}{j} \leqslant \frac{\mathrm{e}^{k \cdot h(j/k)}}{\sqrt{2\pi j(k-j)/k}},$$

где $h(x) = -x \ln x - (1-x)\ln(1-x)$ — шенноновская функция энтропии. Знаменатель этой оценки не может быть меньше $\sqrt{2\pi \cdot (1-1/k)} \geqslant \sqrt{2\pi \cdot 15/16}$, поскольку $k \geqslant n \geqslant \varepsilon^{-4} \geqslant 16$. Знак производной $h(x)$ показывает, что при $x \geqslant \frac{1}{2}$ функция h убывает; поэтому $h(j/k) \leqslant h(\frac{1}{2}+\varepsilon)$. Дважды дифференцируя, можно доказать, что $h(\frac{1}{2}+\varepsilon) \leqslant \ln 2 - 2\varepsilon^2$.

Таким образом, верхняя оценка вероятности для правила r не превышает

$$2^{1-n} + \sqrt{\frac{8}{15\pi}}\mathrm{e}^{-2n\varepsilon^2} n + \sqrt{\frac{8}{15\pi}}\frac{2\mathrm{e}^{-2n\varepsilon^2}}{1-\mathrm{e}^{-2\varepsilon^2}}.$$

Простые вычисления показывают, что последняя оценка строго меньше $\mathrm{e}^{-2n\varepsilon^2(1-\varepsilon)}$ для $\varepsilon \leqslant \frac{1}{2}$ и $n \geqslant \varepsilon^{-4}$ (при этом используется неравенство $1 - \mathrm{e}^{-x} \geqslant x/\sqrt{\mathrm{e}}$ для $0 \leqslant x \leqslant \frac{1}{2}$).

Чтобы оценить вероятность того, что последовательность не является (n,ε)-датчиком хотя бы для одного правила из $\mathcal{R}_L$, умножим на количество правил. Получим, что вероятность того, что последовательность не является (n,ε)-датчиком для $\mathcal{R}_L$, строго меньше единицы, следовательно, существует последовательность, являющаяся (n,ε)-датчиком. $\square$

Замечание. Теорема Колмогорова, как и следующие результаты, обобщается на случай, когда 0 и 1 появляются с частотами, близкими к p и $1-p$ во всех длинных подпоследовательностях, выбранных простыми правилами.

При алгоритмическом подходе получается следующая

Теорема 1$'$. *Рассмотрим произвольные числа L (длина последовательности), $\varepsilon > 0$ (отклонение частоты от вероятности) и $n \geqslant \varepsilon^{-4}$ (длина выборки). Тогда для множества $\mathcal{R}_L$, состоящего из всех правил, условная энтропия которых при известном L меньше*

$$d(n,\varepsilon)(1-\varepsilon)\,,$$

существует (n,ε)-датчик случайных чисел.

Доказательство. Из утверждения 1 следует, что множество правил энтропии меньше $d(n,\varepsilon)(1-\varepsilon)$ имеет сложность также меньше $d(n,\varepsilon)(1-\varepsilon)$, и алгоритмический аналог тривиально вытекает из комбинаторного результата. $\square$

Теорема 2 (Колмогоров, 1963). *Рассмотрим произвольные числа L (длина последовательности), $\varepsilon \in (0, 1/20)$ (отклонение частоты от вероятности) и $n \in [\varepsilon^{-3}, L/2]$ (длина выборки). Существует множество неадаптивных правил $\mathcal{R}_L$, сложность которого меньше*

$$4n\varepsilon(1+5\varepsilon)\,,$$

для которого не существует (n,ε)-датчика случайных чисел.

Доказательство. Построим такое множество неадаптивных правил выбора, что для любой последовательности $\mathbf{t}$ в нём найдётся правило r, которое выбирает длинную подпоследовательность $r[\mathbf{t}]$ с большим отклонением.

Положим

$$m = \left\lfloor \frac{1}{4\varepsilon} + \frac{1}{2} \right\rfloor\,, \quad L' = 2m \cdot \left\lceil \frac{n}{2m-1} \right\rceil\,.$$

Просто проверить, что при заданных в теореме условиях на ε и n будет выполнено $L' < L$.

Наши правила будут выбирать подпоследовательности только из первых L' цифр последовательности. А именно, начало длины L' разобьём на m отрезков равной длины L'/m. Правило помещает в подпоследовательность ровно половину цифр одного из этих отрезков и все цифры остальных. Таким образом, правило задаётся номером отрезка и $\frac{L'}{2m}$-элементным подмножеством множества $\{1, \ldots, L'/m\}$.

Каждое правило выбирает подпоследовательность длины

$$n' = (2m-1) \cdot \left\lceil \frac{n}{2m-1} \right\rceil \geqslant n\,.$$

Зафиксируем последовательность $\mathbf{t}$ и докажем, что найдётся правило r, для которого в подпоследовательности $r[\mathbf{t}]$ отклонение больше ε. Обозначим начало $\mathbf{t}$ длины L' через $\mathbf{t}'$. Рассмотрим три возможных случая.

1) Среди отрезков $\mathbf{t}'$ есть два таких, что в первом не меньше половины цифр составляют нули, а во втором — единицы.

   Рассмотрим два правила, r_1 и r_2. Для их задания достаточно указать, какие цифры они *не* выбирают.

   - Для r_1 все невыбранные цифры лежат в первом отрезке и являются нулями.
   - Для r_2 все невыбранные цифры лежат во втором отрезке и являются единицами.

   Количество нулей в $r_1[\mathbf{t}]$ равно количеству нулей в $\mathbf{t}'$ минус $L'/2m$, а количество нулей в $r_2[\mathbf{t}]$ равно количеству нулей в $\mathbf{t}'$. Таким образом, количества нулей в этих выборках различаются на $L'/2m$, следовательно, одно из них отстоит от $n'/2$ не менее чем на $L'/4m$. Либо в $r_1[\mathbf{t}]$, либо в $r_2[\mathbf{t}]$ отклонение не меньше

$$\frac{L'/4m}{n'} = \frac{1}{2(2m-1)} \geqslant \varepsilon\,.$$

2) Во всех отрезках $\mathbf{t}'$ не меньше половины цифр составляют нули.

   Тогда рассмотрим правило, которое из первого отрезка выбирает половину цифр, причём только нули. Количество нулей в этой выборке не меньше

$$\frac{L'}{2m} + (m-1)\frac{L'/m}{2} = \frac{L'}{2}\,.$$

   Отклонение не меньше

$$\frac{L'/2 - n'/2}{n'} = \frac{2m-(2m-1)}{2(2m-1)} \geqslant \varepsilon\,.$$

3) Во всех отрезках $\mathbf{t}'$ не меньше половины цифр составляют единицы.

   Этот случай полностью аналогичен предыдущему (с заменой нулей на единицы).

Итак, мы доказали, что для построенного множества правил не существует (n, ε)-датчика. Теперь оценим количество правил. Оно равно

$$m \cdot \binom{L'/m}{L'/2m} \leqslant m \cdot \sqrt{\frac{2}{\pi L'/m}} \cdot 2^{L'/m}\,.$$

Осталось заметить, что $m \cdot \sqrt{2m/\pi L'} < 1/4$ при $n \geqslant \varepsilon^{-3}$ и $\varepsilon < 1/20$, а $\frac{L'}{m} < 2 + 4n\varepsilon(1 + 5\varepsilon)$ при $\varepsilon < 1/20$. □

Теорема 1 даёт нижнюю оценку, а теорема 2 даёт верхнюю оценку на максимальное число τ, для которого всякому L и всякому множеству правил, имеющему сложность меньше τ, соответствует хотя бы один (n, ε)-датчик случайных чисел длины L. Поскольку $d(n, \varepsilon) = 2n\varepsilon^2 \operatorname{lb} \mathrm{e}$ гораздо меньше $4n\varepsilon$ при малых ε, Колмогоров стремился устранить разрыв между степенями ε в нижней и верхней оценках. Как отмечено в [1], ему это не удалось. В предисловии к переводу [1] на русский язык, опубликованному в [3], Колмогоров напоминает, что указанная проблема ждёт своего решения.

Оказалось, что нижняя оценка, полученная Колмогоровым, практически точна (даже для неадаптивных правил).

Теорема 3. *Рассмотрим произвольные числа $L \geqslant 2$ (длина последовательности), $\varepsilon \in (0, 1/3)$ (отклонение частоты от вероятности) и $n \in [2\varepsilon^{-3} \operatorname{lb} L, L/2]$ (длина выборки). Существует множество неадаптивных правил $\mathcal{R}_L$, сложность которого меньше*

$$d(n, \varepsilon)\frac{1 + \varepsilon}{1 - n/(L - 1)}\,,$$

для которого не существует (n, ε)-датчика случайных чисел.

Доказательство. Существование такого множества правил докажем, как и в теореме 1, вероятностно. Однако теперь мы рассмотрим распределение вероятностей на правилах и покажем, что событие „для множества правил $\mathcal{R}_L$ существует (n, ε)-датчик“ имеет вероятность меньше единицы.[6]

Нужное множество правил мы будем искать среди неадаптивных правил, которые выбирают подпоследовательности длины ровно n, то есть, правило задаётся n-элементным подмножеством множества $1, \ldots, L$. Всего таких правил $\binom{L}{n}$; на них мы введём равномерное распределение вероятностей.

Зафиксируем некоторую последовательность $\mathbf{t}$ длины L и оценим снизу вероятность того, что она не является (n, ε)-датчиком для случайно взятого правила r, то есть, отклонение в $r[\mathbf{t}]$ не меньше ε. Предположим, что нулей в $\mathbf{t}$ не меньше, чем единиц (противоположный случай рассматривается симметрично).

Сначала рассмотрим ситуацию, когда числа $L/2$ и $n(\frac{1}{2} + \varepsilon)$ являются целыми, а в конце рассуждения объясним, что делать в остальных случаях. Поскольку нам нужна оценка снизу, достаточно оценить вероятность такого отклонения, когда доля нулей в выборке по крайней мере на ε больше половины; очевидно, что эта вероятность минимальна, если нулей и единиц в $\mathbf{t}$ поровну. Достаточно оценить вероятность отклонения, точно равного ε. Итак, оцениваем снизу вероятность выбрать из последовательности, содержащей $L/2$ нулей и $L/2$ единиц, подпоследовательность длины n, в которой ровно $(\frac{1}{2} + \varepsilon)n$ нулей; очевидно, эта вероятность равна

$$\frac{\binom{L/2}{(\frac{1}{2} - \varepsilon)n}\binom{L/2}{(\frac{1}{2} + \varepsilon)n}}{\binom{L}{n}},$$

(считаем, что $\varepsilon < \frac{1}{2}$ и $n \leqslant L/2$).

Как и в теореме 1, используем следующую оценку на биномиальные коэффициенты, вытекающую из формулы Стирлинга:

$$\frac{\mathrm{e}^{k \cdot h(j/k)}}{\sqrt{8j(k-j)/k}} \leqslant \binom{k}{j} \leqslant \frac{\mathrm{e}^{k \cdot h(j/k)}}{\sqrt{2\pi j(k-j)/k}}\,.$$

[6] В математике можно указать не один пример, когда неявный способ построения даёт лучшую оценку, чем все известные явные способы. В книге [4] на страницах 257–261 и 273–280 очень подробно разбирается предложенное Колмогоровым доказательство частного случая теоремы Шеннона о помехоустойчивом кодировании. Там особенно подчёркивается роль „неэффективности“ в рассуждении.

Получим, что искомая вероятность не меньше

$$\mathrm{e}^{L\left(\frac{1}{2}h((1-2\varepsilon)\gamma)+\frac{1}{2}h((1+2\varepsilon)\gamma)-h(\gamma)\right)} \times \frac{\sqrt{2\pi}/4}{\sqrt{(1-4\varepsilon^2)(1-4\varepsilon^2\frac{\gamma^2}{(1-\gamma)^2})(1-\gamma)n}}\,,$$

где $\gamma = n/L$.

Дважды дифференцируя, убеждаемся, что при $\varepsilon \leqslant \frac{1}{\sqrt{8}}$ и $\gamma \leqslant \frac{1}{2}$ выполнено

$$\frac{1}{2}h((1-2\varepsilon)\gamma) + \frac{1}{2}h((1+2\varepsilon)\gamma) - h(\gamma) \geqslant -\frac{2\gamma\varepsilon^2}{1-\gamma}(1+4\varepsilon^2/3)\,. \qquad (1)$$

Простая проверка показывает, что второй сомножитель в оценке вероятности больше $1/\sqrt{\mathrm{e}n}$.

Окончательно получим, что искомая вероятность больше e^{-K}, где

$$K = \frac{2n\varepsilon^2}{1-\gamma}(1+4\varepsilon^2/3) + \frac{1}{2}(1+\ln n)\,.$$

Таким образом, вероятность того, что из фиксированной последовательности одно правило не выберет подпоследовательность с отклонением не менее ε, оказывается меньше $1-\mathrm{e}^{-K}$. Теперь укажем независимо N случайных правил[7]. Вероятность того, что фиксированная последовательность $\mathbf{t}$ является (n,ε)-датчиком для множества указанных правил, будет меньше

$$(1-\mathrm{e}^{-K})^N < \mathrm{e}^{-N\mathrm{e}^{-K}}$$

(здесь использовано неравенство $(1-\frac{1}{x})^x < \mathrm{e}^{-1}$, верное при всех $x > 1$). Умножив на количество последовательностей длины L, получим строгую верхнюю оценку вероятности того, что для множества указанных правил есть хотя бы один (n,ε)-датчик, а именно

$$2^L\mathrm{e}^{-N\mathrm{e}^{-K}} = \mathrm{e}^{L\ln 2 - N\mathrm{e}^{-K}}\,,$$

что не превышает единицы для $N = \lceil \mathrm{e}^K L\ln 2\rceil < \mathrm{e}^K L$. При этом сложность множества указанных правил меньше

$$\frac{2n\varepsilon^2}{1-n/L}(1+4\varepsilon^2/3)\,\mathrm{lb}\,\mathrm{e} + 2\,\mathrm{lb}\,L\,.$$

[7]Некоторые из этих правил могут совпасть друг с другом.

Легко убедиться, что

$$d(n,\varepsilon)\frac{1+\varepsilon}{1-n/L} > \frac{2n\varepsilon^2}{1-n/L}(1+4\varepsilon^2/3)\,\mathrm{lb}\,\mathrm{e} + 2\,\mathrm{lb}\,L \qquad (2)$$

при $\varepsilon < 1/3$ и $n \geqslant 2\varepsilon^{-3}\,\mathrm{lb}\,L$.

Теперь вернёмся к нашему предположению о том, что числа $L/2$ и $n(\frac{1}{2}+\varepsilon)$ — целые. Если число L нечётное, то можно провести всё рассуждение для начал исходных последовательностей длины $L-1$. При этом в окончательной формуле n/L заменится на $n/(L-1)$. Число $n(\frac{1}{2}+\varepsilon)$ всегда можно сделать целым, заменив ε на $\varepsilon' \geqslant \varepsilon$ так, что $\varepsilon' - \varepsilon < \frac{1}{n}$. Из условий теоремы просто следует, что если $\varepsilon < \frac{1}{3}$, то $\varepsilon' < \frac{1}{\sqrt{8}}$. Поэтому истинность неравенства (1) сохраняется при замене ε на ε', истинность же неравенства (2) сохраняется, если его левую часть оставить прежней, а в правой заменить ε на ε'. $\square$

Замечание. В доказательстве теоремы 2 множество $\mathcal{R}_L$ полиномиально вычислимо, то есть построен алгоритм, который по номеру правила из $\mathcal{R}_L$ и его аргументу за полиномиальное время вычисляет результат применения этого правила к аргументу. Для множества $\mathcal{R}_L$ из теоремы 3 это может оказаться не так, поскольку доказательство вероятностное.

Открытая проблема: можно ли усовершенствовать доказательство теоремы 3 так, чтобы множество $\mathcal{R}_L$ было полиномиально вычислимо (неформально говоря, чтобы $\mathcal{R}_L$ стало явно заданным)?

Теорема 3$'$. *Рассмотрим произвольные числа $L \geqslant 2$ (длина последовательности), рациональное $\varepsilon \in (0, 1/3)$ (отклонение частоты от вероятности) и $n \in [2\varepsilon^{-3}\,\mathrm{lb}\,L, L/2]$ (длина выборки). Для множества $\mathcal{R}_L(n,\varepsilon)$ всех неадаптивных правил, у которых условная энтропия при известных L, n, ε меньше*

$$d(n,\varepsilon)\frac{1+\varepsilon}{1-n/L} + C\,,$$

не существует (n,ε)-датчика случайных чисел.

Доказательство. По предыдущей теореме для некоторого множества неадаптивных правил, имеющего сложность меньше $d(n,\varepsilon)\frac{1+\varepsilon}{1-n/(L-1)}$, не существует (n,ε)-датчика длины L. Покажем, что, зная L, n, ε, множество правил с таким свойством можно построить алгоритмически.

Действительно, по каждому множеству неадаптивных правил $\mathcal{R}_L$ и последовательности $\mathbf{t}$ длины L можно эффективно узнать, является ли $\mathbf{t}$ (n,ε)-датчиком для $\mathcal{R}_L$ (надо применить к $\mathbf{t}$ каждое правило из $\mathcal{R}_L$ и вычислить отклонение). Перебрав все последовательности длины L, можно проверить, существуют ли (n,ε)-датчики для $\mathcal{R}_L$. Перебирая все множества неадаптивных правил данного размера, мы найдём нужное (если множеств с нужным свойством несколько, мы возьмём первое в нашем перечислении).

По L, n, ε найденное множество правил можно перечислить программой, энтропия которой зависит только от языка программирования. Из утверждения 2 (релятивизованного относительно L, n, ε) следует, что условная энтропия каждого из этих правил не превосходит сложности множества плюс некоторая константа. При добавлении правил (всех остальных с энтропией меньше $d(n,\varepsilon)\frac{1+\varepsilon}{1-n/(L-1)}+C$) требуемое свойство (отсутствие (n,ε)-датчика) не может нарушиться. Остаётся отметить, что разность $\left(d(n,\varepsilon)\frac{1+\varepsilon}{1-n/(L-1)} - d(n,\varepsilon)\frac{1+\varepsilon}{1-n/L}\right)$ меньше 1. $\square$

2. Сравнение универсальной и частотной случайности

2.1. Философская мотивировка

Для уточнения представлений о случайности в математической статистике традиционно используются тесты. Пусть перед нами конечная последовательность карт (см. абзац перед определением 1), повёрнутых лицевой стороной вниз. На лицевой стороне каждой карты нарисована двоичная цифра. Последовательности карт в этой формулировке соответствует (неизвестная нам) последовательность нулей и единиц. Неслучайность (относительно равномерного распределения) этой последовательности неформально означает возможность сделать нетривиальное предсказание о её поведении. Если кто-нибудь назовёт последовательность, и после переворачивания карт нарисованная на них последовательность совпадёт с названной, то такую последовательность можно считать максимально неслучайной. Если кто-нибудь назовёт множество, содержащее сравнительно мало последовательностей, и после переворачивания карт нарисованная на них последовательность попадёт в названное множество, то и такую последовательность можно считать неслучайной (причём мера её неслучайности будет тем меньше, чем больше будет мощность названного множества). Такие множества принято называть тестами.

На практике бывает удобно рассматривать тесты какого-то специального вида. Особенно важными являются частотные тесты. Каждому правилу r, выбирающему из последовательностей длины L подпоследовательности длины n, и отклонению ε сопоставляется множество[8] (частотный тест), содержащее те последовательности $\mathbf{t}$, для которых $\ell(\mathbf{t}) = L$, $\ell(r[\mathbf{t}]) = n$ и доля нулей в $r[\mathbf{t}]$ отстоит от $\frac{1}{2}$ по крайней мере на ε. Можно было бы сказать, что общее представление о случайности сводится к частотному, если бы любой тест U можно было покрыть небольшим набором частотных тестов $F_1, \ldots, F_m$. При этом важно, чтобы m было очень мало по сравнению с мощностью U и чтобы мощность каждого F_i не была гораздо больше мощности U. В теореме 4 даются конкретные оценки такого типа. В теореме 5 доказывается, что оценки из теоремы 4 нельзя существенно улучшить. Из теоремы 6 следует, что частотные тесты F_i, вообще говоря, невозможно задать с помощью правил выбора более узкого (чем в теореме 4) класса.

Стоит отметить, что комбинаторные результаты первой части статьи тоже можно интерпретировать как исследование вопроса о возможности покрытия множества U частотными тестами $F_1, \ldots, F_m$. Только там в качестве U берётся множество всех последовательностей.

При комбинаторном подходе, как и в классической теории вероятностей, нельзя уточнить использованное нами выражение „кто-нибудь, ещё не зная последовательности, назовёт множество малой мощности, которому эта последовательность принадлежит“. При алгоритмическом подходе в качестве этого множества изучается множество последовательностей малой (относительно длины) энтропии. Другими словами, изучается один („универсальный“) тест. При этом плюсом является то, что мы можем говорить о мере случайности индивидуальной последовательности. Минусом является то, что универсальный тест не задан явно (он перечислим, но не разрешим).

Как и в первой части статьи, теоремы алгоритмического подхода имеют номера параллельных теорем комбинаторного подхода с добавленным штрихом.

[8]В определении 3 такие множества названы правильными.

2.2. Комбинаторный подход

Определение 4. *Дефектом* $\Delta(S)$ множества S двоичных последовательностей некоторой длины L называется разность между L и сложностью множества S, т. е., $\Delta(S) = L - \mathrm{lb}\,|S|$. *Удельным дефектом* $\delta(S)$ называется величина $\delta(S) = \Delta(S)/L$.

Удельный дефект характеризует меру нетривиальности множества, если его использовать как тест.

Теорема 4. *Пусть $\delta \in (0, \mathrm{e}^{-\mathrm{e}^8})$ и фиксировано натуральное $L \geqslant (1/\delta)^5$. Будем рассматривать множества двоичных последовательностей длины L. Для произвольного множества S с удельным дефектом $\geqslant \delta$ существует семейство правильных множеств с удельным дефектом больше*

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

которое покрывает S и содержит не более $L\mathrm{e}^{1/\delta'}$ множеств (а в качестве β можно взять $\frac{2\ln\ln(1/\delta)}{\ln(1/\delta)}$).

Доказательство. Для произвольного множества S последовательностей длины L рассмотрим следующую игру. Математик и Природа по очереди делают L ходов: на i-ом ходу Математик указывает ставку $x_i \in [0,1]$ на цифру 0 или 1, а Природа указывает элемент $t_i \in \{0,1\}$, причём так, чтобы после L-ого хода построенная последовательность $\mathbf{t}$ принадлежала S. В начале капитал Математика равен нулю, затем на каждом ходу он возрастает на величину ставки, если Математик угадал следующую цифру t_i, и уменьшается на ту же величину в противоположном случае. Капитал может быть отрицательным, поэтому такая игра называется игрой „в кредит“. Покажем, что для любого множества S существует стратегия Математика, позволяющая выиграть не менее $\Delta(S)\ln 2$.

Введём сначала несколько обозначений. Если $\mathbf{s}$ — продолжение последовательности $\mathbf{t}$, будем писать $\mathbf{s} \sqsupset \mathbf{t}$ (при этом возможно $\mathbf{s} = \mathbf{t}$). $S_{\mathbf{t}} = \{\mathbf{s} \mid \mathbf{s} \sqsupset \mathbf{t}, \mathbf{s} \in S\}$ — множество продолжений $\mathbf{t}$, принадлежащих множеству S. Пусть $\mathbf{t}_{1:i}$ — часть последовательности $\mathbf{t}$, полученная перед $(i+1)$-м ходом.

Стратегия Математика заключается в следующем. Если $|S_{\mathbf{t}_{1:i}0}| \geqslant |S_{\mathbf{t}_{1:i}1}|$ (то есть, если большинство лежащих в S продолжений $\mathbf{t}_{1:i}$ начинается на 0), то Математик ставит на 0, а в противном случае —

на 1; величина ставки x определяется соотношением $\frac{1+x}{2} = \frac{|S_{\mathbf{t}_{1:i}0}|}{|S_{\mathbf{t}_{1:i}}|}$ или, соответственно, $\frac{1+x}{2} = \frac{|S_{\mathbf{t}_{1:i}1}|}{|S_{\mathbf{t}_{1:i}}|}$.

Покажем, что для этой стратегии не убывает величина $K_i + \Delta_i \ln 2$, где K_i — капитал Математика после i-го хода, а $\Delta_i = \Delta(S_{\mathbf{t}_{1:i}}) - i$ является дефектом множества возможных концовок игры. Поскольку в начальный момент $K_0 = 0$, $\Delta_0 = \Delta(S)$, а после последнего хода $\Delta(S_{\mathbf{t}_{1:L}}) - L = L - \operatorname{lb} 1 - L = 0$, отсюда будет следовать, что выигрыш Математика K_L не меньше $\Delta(S) \ln 2$. Мы хотим доказать, что сумма $(K_{i+1} - K_i) + (\Delta_{i+1} - \Delta_i) \ln 2$ всегда неотрицательна. Второе слагаемое равно $-\ln 2 - \ln \frac{|S_{\mathbf{t}_{1:(i+1)}}|}{|S_{\mathbf{t}_{1:i}}|}$. Первое слагаемое равно ставке x или $(-x)$. В первом случае получается $x - \ln 2 - \ln \frac{1+x}{2} = x - \ln(1+x) \geqslant 0$ (в силу известного неравенства для логарифма). Во втором случае получается $-x - \ln 2 - \ln \frac{1-x}{2} = -x - \ln(1-x) \geqslant 0$. Таким образом, мы доказали, что выигрыш составит не меньше $\delta L \ln 2$, где $\delta = \delta(S)$.

Описанную стратегию мы используем для построения множества правил, которые дадут покрытие S правильными множествами. В общих чертах конструкция выглядит так: по S мы строим правильные множества большого дефекта, покрывающие „почти всё“ S, а для оставшейся части мы повторяем ту же конструкцию; оценка количества покрытых последовательностей проводится с помощью нашей выигрышной стратегии (которая будет слегка модифицирована, чтобы уменьшить количество возможных размеров ставок).

Прежде всего модифицируем стратегию. Для этого возьмём ближайшее сверху к $\ln(1/\delta)$ такое число B, что $M = \frac{B}{\delta \ln 2}$ — натуральное число, и округлим с уменьшением все ставки до ближайшего числа вида k/M (где $k \in \mathbb{N}$). Заметим, что уменьшение выигрыша на каждом ходу меньше $\frac{1}{M}$, и стратегия теперь выигрывает больше $\delta L \ln 2(1 - \frac{1}{B})$. При этом ненулевые ставки могут принимать только M различных значений.

Двоичной цифре (0 или 1) и множеству допустимых размеров ставок — подмножеству $\{1/M, 2/M, \ldots, 1\}$ — сопоставляем следующее правило. Оно читает последовательность по порядку и применяет к прочитанному отрезку выигрышную стратегию; очередная цифра выбирается в подпоследовательность, если стратегия предписывает поставить ставку допустимого размера на заданную цифру. Общее количество правил выбора такого вида не превышает $2 \cdot 2^M$. К каждому правилу применим n-нормализацию для всех n, после этого количество нормальных правил не превышает $2L \cdot 2^{B/(\delta \ln 2)} < 2L \cdot 2^{1+\mathrm{lb}(1/\delta)/\delta} = 4L(1/\delta)^{1/\delta}$.

Для почти каждой последовательности из S найдём нормальное правило r описанного вида и число ε, такие что соответствующее правильное множество $A_{r,\varepsilon}$ содержит эту последовательность и имеет достаточно большой дефект. Из $\varepsilon_1 > \varepsilon_2$ следует $A_{r,\varepsilon_1} \subseteq A_{r,\varepsilon_2}$, поэтому для каждого r можно брать наименьшее ε, при котором дефект $A_{r,\varepsilon}$ достаточно велик. Бо́льшая часть S окажется покрытой не более чем $4L(1/\delta)^{1/\delta}$ множествами. А для малого исключительного множества $\widetilde{S} \subset S$ повторим всю конструкцию; рекурсивный подсчёт покажет, что суммарное количество множеств в покрытии будет невелико.

Зафиксируем последовательность $\mathbf{t}_0 \in S$. Наша стратегия выигрывает на ней более $\delta L \ln 2(1 - \frac{1}{B})$, ставя как на нуль, так и на единицу. При ставках только на 0 или только на 1 выигрыш хотя бы в одном случае составит более $\frac{1}{2}\delta L \ln 2(1-\frac{1}{B})$. Выберем соответствующую цифру (будем считать, что это 0) и далее будем рассматривать ставки только на 0 (а ставки на 1 игнорируем).

Для произвольного начала $\mathbf{t}$ последовательности из S обозначим через $d_i(\mathbf{t})$ разность количества выигрышей и проигрышей на ставках размера i/M (ставки только на 0), а через $n_i(\mathbf{t})$ — общее количество таких ставок (в ситуации, когда Природа ходила в соответствии с $\mathbf{t}$).

Выигрыш на $\mathbf{t}_0$ равен

$$\sum_{i=1}^{M} \frac{i}{M} d_i(\mathbf{t}_0)\,.$$

Исключим те ставки, для которых $n_i(\mathbf{t}_0) \leqslant \frac{L}{M^2}$. Выигрыш на одном ходу не превышает размера ставки, то есть 1, поэтому общий выигрыш уменьшится самое большее на $1 \cdot \frac{L}{M^2} \cdot M = L/M = \delta L \ln 2/B$. Таким образом,

$$\sum_{i\,:\, n_i(\mathbf{t}_0) > \frac{L}{M^2}} \frac{i}{M} d_i(\mathbf{t}_0) \quad > \quad \frac{1}{2}\delta L \ln 2 \left(1 - \frac{3}{B}\right)\,. \tag{3}$$

Для каждого $\mathbf{t}_1 \in S$ рассмотрим множество

$$S(\mathbf{t}_1, i) = \left\{ \mathbf{t} \in S \,\middle|\, n_i(\mathbf{t}) = n_i(\mathbf{t}_1),\, d_i(\mathbf{t}) \leqslant \frac{i}{M} n_i(\mathbf{t}_1) \left(1 - \frac{1}{B}\right) \right\}.$$

Предложение 4.1. *Если $n_i(\mathbf{t}_1) > L/M^2$, то $|S(\mathbf{t}_1, i)| < |S| \mathrm{e}^{-L/2B^2M^4}$.*

Докажем это предложение позже, а сейчас продолжим доказательство теоремы.

Положим

$$\widetilde{S} \quad = \bigcup_{\substack{\mathbf{t}_1 \in S \\ i\colon n_i(\mathbf{t}_1) > L/M^2}} S(\mathbf{t}_1, i)\,.$$

Заметим, что множество $S(\mathbf{t}_1, i)$ полностью определяется двумя параметрами — i и $n_i(\mathbf{t}_1)$, которые могут принимать соответственно M и L различных значений. Поэтому

$$|\widetilde{S}| < LM|S|\mathrm{e}^{-L/2B^2M^4}\,.$$

Рассмотрим два возможных случая:

1) $\mathbf{t}_0 \in \widetilde{S}$;

2) $\mathbf{t}_0 \notin \widetilde{S}$.

В первом случае $\mathbf{t}_0$ попало в исключительное множество, которое мы не пытаемся покрыть правильными множествами ранее описанного типа. Вместо этого мы повторяем всю описываемую конструкцию для множества $\widetilde{S}$ с одним и тем же δ (дефект $\widetilde{S}$ больше дефекта S). В результате получаем цепочку уменьшающихся множеств $S^{(0)} = S$, $S^{(1)} = \widetilde{S^{(0)}}$, $S^{(2)} = \widetilde{S^{(1)}}, \ldots$. Эта цепочка оборвётся, поскольку мощность множества — целое число — на каждом шаге уменьшается более чем в $\mathrm{e}^{L/2B^2M^4}/LM$ раз. Длина цепочки не превышает

$$1 + \frac{\ln|S|}{\ln(\mathrm{e}^{L/2B^2M^4}/LM)} < 1 + \frac{L\ln 2}{L/2B^2M^4 - \ln L - \ln M}\,.$$

Учитывая границы, наложенные в условии теоремы на L и δ, длину цепочки можно оценить сверху как $\frac{25\ln^6(1/\delta)}{\delta^4}$. Так как количество правил для одного члена цепочки не больше $4L(1/\delta)^{1/\delta}$, то всего их получается меньше $L\mathrm{e}^{1/\delta'}$.

Разберём теперь второй случай. Построим для $\mathbf{t}_0$ покрывающее правильное множество, то есть, укажем правило (совокупность размеров ставок) и отклонение.

Для краткости обозначим $n_i = n_i(\mathbf{t}_0)$, $d_i = d_i(\mathbf{t}_0)$. Из $\mathbf{t}_0 \notin \widetilde{S}$, в частности, следует, что при $n_i > L/M^2$ имеет место $\mathbf{t}_0 \notin S(\mathbf{t}_0, i)$, то есть, $d_i > \frac{i}{M} n_i (1 - \frac{1}{B})$. Отсюда и из (3)

$$\sum_{i\,:\, n_i > \frac{L}{M^2}} \frac{d_i^2}{n_i} > \sum_{i\,:\, n_i > \frac{L}{M^2}} d_i \frac{i}{M} \left(1 - \frac{1}{B}\right) > \frac{1}{2} \delta L \ln 2 \left(1 - \frac{4}{B}\right) .$$

Рассмотрим правильное множество, которое задаётся совокупностью ставок $J \subseteq \{i \mid n_i > \frac{L}{M^2}\}$, длиной выбираемой подпоследовательности $n = \sum_{i\in J} n_i$ и отклонением $\varepsilon = \frac{1}{2n} \sum_{i \in J} d_i$. Это множество содержит $\mathbf{t}_0$ (напомним, что d_i — разность количества выигрышей и проигрышей). По оценке Чернова удельный дефект такого множества не меньше $2n\varepsilon^2 \operatorname{lb} \mathrm{e}/L$. Теорема будет доказана, если мы покажем, что для некоторого J эта величина больше $\frac{\delta}{\ln(1/\delta)}(1-\beta)$.

Предположим противное, то есть, для каждого $J \subseteq \{i \mid n_i > \frac{L}{M^2}\}$

$$\frac{\left(\sum_{i\in J} d_i\right)^2}{\left(\sum_{i \in J} n_i\right) \cdot 2L \ln 2} \leqslant \frac{\delta}{\ln(1/\delta)}(1-\beta) .$$

Перепишем это в виде

$$\left(\sum_{i\in J} \frac{d_i}{n_i} n_i\right)^2 \leqslant Z \cdot \left(\sum_{i \in J} n_i\right) ,$$

где $Z = \frac{\delta}{\ln(1/\delta)}(1-\beta) \cdot 2L \ln 2$.

Предложение 4.2. *Если* $\sum_{i\in J} \frac{d_i}{n_i} n_i \leqslant \sqrt{Z \sum_{i \in J} n_i}$ *для всех* J, *то*

$$\sum \frac{d_i^2}{n_i} \leqslant Z + \frac{Z}{4}(\ln \sum n_i - \ln Z) ,$$

где суммирование идёт по всем i *таким, что* $n_i > \frac{L}{M^2}$.

Доказательство этого предложения помещено чуть ниже.

Вспоминая полученную ранее оценку на $\sum d_i^2/n_i$, имеем

$$\frac{1}{2} \delta L \ln 2 \left(1 - \frac{4}{B}\right) < Z + \frac{Z}{4} (\ln L - \ln Z)$$

(здесь учтено $\sum n_i \leqslant L$). После простых преобразований получим

$$\left(1-\frac{4}{\ln(1/\delta)}\right) < (1-\beta)\left(1+\frac{\ln\ln(1/\delta)+4-\ln\ln 4-\ln(1-\beta)}{\ln(1/\delta)}\right),$$

что неверно при $\beta = \frac{2\ln\ln(1/\delta)}{\ln(1/\delta)}$ и $\delta < \mathrm{e}^{-\mathrm{e}^8}$. □

Осталось доказать два использованных предложения.

Доказательство предложения 4.2. Мы хотим из оценки сумм $\sum \frac{d_i}{n_i} n_i$ получить оценку суммы $\sum \frac{d_i^2}{n_i} = \sum \left(\frac{d_i}{n_i}\right)^2 n_i$. Заменим эти суммы интегралами от кусочно-постоянных функций. Поскольку упорядочение индексов i несущественно, будем считать, что $i = 1, \ldots, I$, а $\frac{d_i}{n_i}$ монотонно невозрастают. На полуинтервале $[0, \sum_{i=1}^{I} n_i)$ рассмотрим невозрастающую функцию f, определённую равенством:

$$f(x) = \frac{d_k}{n_k}, \text{ если } N_{k-1} \leqslant x < N_k,$$

где $N_k = \sum_{i=1}^{k} n_i$. Очевидно, что рассматриваемые суммы по i от 1 до k равны соответственно интегралам f и f^2 от 0 до N_k. По условию, для любого k

$$\int_0^{N_k} f(x)\,dx \leqslant \sqrt{Z \cdot N_k}.$$

Пользуясь выпуклостью корня, продолжим это неравенство на произвольные значения верхнего предела интегрирования. Действительно, для $N_k \leqslant u \leqslant N_{k+1}$

$$\int_0^{u} f(x)\,dx = \int_0^{N_k} f(x)\,dx + \frac{d_{k+1}}{n_{k+1}}(u - N_k).$$

Из неравенства при $u = N_{k+1}$ получаем для всех u

$$\frac{d_{k+1}}{n_{k+1}}(u - N_k) \leqslant \left(\sqrt{ZN_{k+1}} - \int_0^{N_k} f(x)\,dx\right)\frac{u - N_k}{N_{k+1} - N_k}.$$

Отсюда

$$\int_0^{u} f(x)\,dx \leqslant \sqrt{ZN_{k+1}}\frac{u - N_k}{N_{k+1} - N_k} + \sqrt{ZN_k}\left(1 - \frac{u - N_k}{N_{k+1} - N_k}\right) \leqslant \sqrt{Zu}.$$

Мы хотим сделать так, чтобы в правой части неравенства был интеграл некоторой ограниченной невозрастающей функции g.

Заметим, что $f(x) \leqslant 1$, так как $d_k \leqslant n_k$. Положим

$$g(x) = \begin{cases} 1, & \text{если } x < Z, \\ \frac{\sqrt{Z}}{2\sqrt{x}}, & \text{если } x \geqslant Z. \end{cases}$$

Отметим, что если $u \geqslant Z$, то $\int_0^u g(x)\,dx = \sqrt{Zu}$. Тогда

$$\int_0^u (g(x) - f(x))\,dx \geqslant 0$$

при всех $u \in [0, N_I]$. Выведем отсюда, что $\int_0^u (g^2(x) - f^2(x))\,dx \geqslant 0$.

Вторая теорема о среднем для интегралов утверждает, что для любой интегрируемой функции $\psi(x)$ и любой неотрицательной невозрастающей функции $\varphi(x)$ найдётся точка $\xi \in [a, b]$ такая, что

$$\int_a^b \varphi(x)\psi(x)\,dx = \varphi(a) \int_a^\xi \psi(x)\,dx\,.$$

Для $\psi(x) = g(x) - f(x)$ и $\varphi(x) = g(x) + f(x)$ условия теоремы выполнены, откуда

$$\int_0^u (g^2(x) - f^2(x))\,dx = (g(0) + f(0)) \int_0^\xi (g(x) - f(x))\,dx \geqslant 0\,.$$

Проверим, что $Z < N_I$. В самом деле, $\sum n_i = \sum \frac{n_i^2}{n_i} \geqslant \sum \frac{d_i^2}{n_i} > \frac{1}{2}\delta L \ln 2(1 - \frac{4}{B}) > \frac{\delta}{\ln(1/\delta)}(1 - \beta) \cdot 2L \ln 2 = Z$ (использовано, что δ мало).

Таким образом,

$$\sum_i \frac{d_i^2}{n_i} = \int_0^{N_I} f^2(x)\,dx \leqslant \int_0^{N_I} g^2(x)\,dx = Z + \frac{Z}{4}(\ln N_I - \ln Z)\,.$$

$\square$

Доказательство предложения 4.1. Рассмотрим множество $S(\mathbf{t}_1, i)$. Заметим, что условие

$$d_i(\mathbf{t}) \leqslant \frac{i}{M} n_i(\mathbf{t}_1) \left(1 - \frac{1}{B}\right)$$

можно переписать в виде

$$\left(\frac{1}{2}+\frac{i}{2M}\right)n_i - \frac{in_i}{2BM} \geqslant \frac{n_i+d_i}{2}\,.$$

Чтобы оценить размер $S(\mathbf{t}_1, i)$, введём две вспомогательные вероятностные меры, $\Pr_S$ и $\Pr$. Мера $\Pr_S$ равномерно распределена на S, то есть $\Pr_S(X) = |X\cap S|/|S|$. В качестве меры $\Pr$ возьмём бернуллиевскую меру с вероятностью нуля $\frac{1}{2}+\frac{i}{2M}$.

Пусть $X_{\nu,\sigma}$ — множество последовательностей длины ν, в которых не более σ нулей. Пусть $\mathbf{t} \in \{0,1\}^{\leqslant L}$, $z_i(\mathbf{t})$ — количество выигрышей на ставках размера i/M, сделанных на цифру 0, когда Природа играет в соответствии с последовательностью $\mathbf{t}$, а Математик играет в соответствии со своей (модифицированной) стратегией. Отметим, что $z_i(\mathbf{t}) = (n_i(\mathbf{t}) + d_i(\mathbf{t}))/2$. Предполагая, что $S_\mathbf{s} \neq \varnothing$, докажем индукцией по длине $\mathbf{s}$ (спуском от L к 0) следующее соотношение[9]:

$$\forall\nu\forall\sigma \quad \Pr_S\{n_i(\mathbf{t}) = \nu + n_i(\mathbf{s}), z_i(\mathbf{t}) \leqslant \sigma + z_i(\mathbf{s}) \mid \mathbf{t} \sqsupset \mathbf{s}\} \leqslant \Pr(X_{\nu,\sigma})\,.$$

Если $\nu < 0$ или $\sigma < 0$, то вероятность слева равна 0. Далее считаем, что $\nu \geqslant 0$, $\sigma \geqslant 0$.

Предположим $\ell(\mathbf{s}) = L$, тогда

если $\nu = 0$, то вероятность справа равна 1;
если $\nu \neq 0$, то вероятность слева равна 0.

Базис индукции доказан.

Делая индуктивный переход, разберём два случая. Для краткости обозначим $P(\mathbf{s},\nu,\sigma) = \Pr_S\{n_i(\mathbf{t}) = \nu + n_i(\mathbf{s}), z_i(\mathbf{t}) \leqslant \sigma + z_i(\mathbf{s}) \mid \mathbf{t} \sqsupset \mathbf{s}\}$, $\Pr_S\{\mathbf{t} \sqsupset \mathbf{s}0 \mid \mathbf{t} \sqsupset \mathbf{s}\} = \frac{|S_{\mathbf{s}0}|}{|S_\mathbf{s}|} = \frac{1+x}{2}$.

1) В позиции $\mathbf{s}$ стратегия не делает на нуль ставку размера i/M. Тогда

$$P(\mathbf{s},\nu,\sigma) = \frac{1+x}{2}P(\mathbf{s}0,\nu,\sigma) + \frac{1-x}{2}P(\mathbf{s}1,\nu,\sigma) \leqslant \left(\frac{1+x}{2}+\frac{1-x}{2}\right)\Pr(X_{\nu,\sigma}) = \Pr(X_{\nu,\sigma})\,.$$

[9] $\Pr_S\{U \mid V\}$ обозначает вероятность выполнения U при условии выполнения V, то есть $\Pr_S(\{\mathbf{t} \mid U, V\})/\Pr_S(\{\mathbf{t} \mid V\})$.

2) В позиции $\mathbf{s}$ стратегия делает на нуль ставку размера i/M. Напомним, что в этом случае $i/M \leqslant x$. Тогда

$$P(\mathbf{s},\nu,\sigma)=\frac{1+x}{2}P(\mathbf{s}0,\nu-1,\sigma-1)+\frac{1-x}{2}P(\mathbf{s}1,\nu-1,\sigma)\leqslant \frac{1+x}{2}\Pr(X_{\nu-1,\sigma-1})+\frac{1-x}{2}\Pr(X_{\nu-1,\sigma})\,.$$

С другой стороны,

$$\Pr(X_{\nu,\sigma})=\frac{1+i/M}{2}\Pr(X_{\nu-1,\sigma-1})+\frac{1-i/M}{2}\Pr(X_{\nu-1,\sigma})\,.$$

Очевидно, что $\Pr(X_{\nu-1,\sigma})\geqslant\Pr(X_{\nu-1,\sigma-1})$. Поэтому, уменьшая вес меньшего слагаемого (с $(1+x)/2$ до $(1+i/M)/2$) и соответственно увеличивая вес большего, мы только увеличим сумму. Отсюда получаем требуемое неравенство $P(\mathbf{s},\nu,\sigma)\leqslant\Pr(X_{\nu,\sigma})$.

Таким образом, $|S(\mathbf{t}_1,i)|/|S|=\Pr_S\{n_i(\mathbf{t})=\nu,\ z_i(\mathbf{t})\leqslant\sigma\}\leqslant\Pr(X_{\nu,\sigma})$, где $\nu=n_i(\mathbf{t}_1)$, $\sigma=(\frac{1}{2}+i/2M)\nu-i\nu/2BM$.

Из оценки Чернова имеем

$$\Pr(X_{\nu,\sigma})\leqslant \mathrm{e}^{-2(\frac{i}{2BM})^2\nu}\,,$$

и используя то, что $i\geqslant 1$, а $n_i(\mathbf{t}_1)>\frac{L}{M^2}$, получаем окончательную оценку

$$|S(\mathbf{t}_1,i)|<|S|\mathrm{e}^{-\frac{L}{2B^2M^4}}\,.$$

□

Замечание. Можно доказать, что в теореме 4 (в отличие от теоремы 3) недостаточно рассматривать только семейства правильных множеств с одним и тем же отклонением ε.

Замечание. В доказательстве теоремы 4 были использованы правильные множества, порождённые только монотонными правилами. В теореме 6 будет показано, что нельзя ограничиться правильными множествами, порождёнными неадаптивными правилами.

Следующая теорема показывает, что оценка теоремы 4 на удельный дефект правильных множеств точна.

Теорема 5. *Пусть* $\delta \in (0, \mathrm{e}^{-\mathrm{e}^8})$ *и фиксировано натуральное* $L \geqslant (1/\delta)^5$. *Будем рассматривать множества двоичных последовательностей длины* L. *Существует множество* S *с удельным дефектом больше* δ, *которое нельзя покрыть менее чем* $\mathrm{e}^{L\delta^4/70}$ *правильными множествами с удельным дефектом не меньше*

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

Доказательство. Положим

$$m = \left\lfloor \left(2 - 4\sqrt{\delta \ln(1/\delta)}\right) \Big/ \delta \ln 2 \right\rfloor\,,$$
$$k = \lfloor L/2m \rfloor\,.$$

Каждую последовательность длины L мы будем представлять в виде конкатенации m пар последовательностей длины k и „остатка" длины $L - 2km < 2m$. В каждой паре одна последовательность сопоставлена (в уточняемом далее смысле) цифре 0, другая — цифре 1; мы будем называть эти последовательности i-ым 0- и 1-отрезками исходной последовательности $\mathbf{t}$ длины L и обозначать $\mathbf{t}_{i,0}$ и $\mathbf{t}_{i,1}$ соответственно.

Определим сначала для каждого i от 1 до m множества $S_{i,0}$ и $S_{i,1}$ последовательностей длины k. Пусть $p_i = \frac{1}{2} + \frac{1}{\sqrt{i \ln m}}$. Тогда $S_{i,0}$ состоит из последовательностей, в которых не менее $k(p_i - \frac{1}{m})$ нулей, а $S_{i,1}$ состоит из последовательностей, в которых не менее $k(p_i - \frac{1}{m})$ единиц. К множеству S отнесём все последовательности $\mathbf{t}$, у которых $\mathbf{t}_{i,0} \in S_{i,0}$ и $\mathbf{t}_{i,1} \in S_{i,1}$ для каждого i.

Оценим сначала удельный дефект S. Мощность S равна произведению $|S_{i,0}| \cdot |S_{i,1}|$ по всем i, умноженному на 2^{L-2km} — количество возможных „остатков", поэтому из оценки Чернова (для равномерного распределения на всех двоичных последовательностях) получаем:

$$|S| \leqslant 2^{L-2km} \prod_{i=1}^{m} \left(2^k \mathrm{e}^{-2k\left(\frac{1}{\sqrt{i \ln m}} - \frac{1}{m}\right)^2}\right)^2 = 2^{L-2km} \cdot 2^{2km} \mathrm{e}^{-4k \sum\limits_{i=1}^{m} \left(\frac{1}{\sqrt{i \ln m}} - \frac{1}{m}\right)^2}\,.$$

Поскольку $1 - \frac{\sqrt{i \ln m}}{m} \geqslant 1 - \sqrt{\frac{\ln m}{m}}$ и $\sum\limits_{i=1}^{m} \frac{1}{i} > \ln m$, имеем

$$\sum_{i=1}^{m} \left(\frac{1}{\sqrt{i \ln m}} - \frac{1}{m}\right)^2 = \sum_{i=1}^{m} \frac{1}{i \ln m} \left(1 - \frac{\sqrt{i \ln m}}{m}\right)^2 > \frac{\ln m}{\ln m} \left(1 - \sqrt{\frac{\ln m}{m}}\right)^2\,.$$

Откуда $|S| < 2^L \mathrm{e}^{-4k\left(1-2\sqrt{\frac{\ln m}{m}}\right)}$, и таким образом удельный дефект S больше $\frac{4k}{L\ln 2}\left(1-2\sqrt{\frac{\ln m}{m}}\right) > \left(\frac{2}{m\ln 2} - \frac{4}{L\ln 2}\right)\left(1-2\sqrt{\frac{\ln m}{m}}\right) > \delta$.

Возьмём какое-нибудь покрытие S правильными множествами, в котором меньше $\mathrm{e}^{L\delta^4/70}$ элементов. Докажем, что хотя бы одно множество в этом покрытии имеет удельный дефект меньше $\frac{2\delta}{\ln(1/\delta)}$.

Схема доказательства следующая. Рассмотрим множество правил, соответствующих участвующим в покрытии правильным множествам. Для каждого правила r назовём последовательность r-типичной, если в выбранной подпоследовательности отклонение частоты от $\frac{1}{2}$ невелико по отношению к длине выборки. Затем мы докажем, что во множестве S найдётся последовательность, типичная для всех рассматриваемых правил (показав, что вероятность противоположного события относительно некоторого распределения вероятностей строго меньше единицы). Поэтому одно из множеств покрытия имеет малый дефект.

Пусть $\mathbf{t}$ — последовательность длины L, а r — нормальное правило выбора. Мы дадим определение r-типичности в предположении, что в выборке $r[\mathbf{t}]$ количество единиц больше количества нулей; если это не так, всюду в определении надо заменить 0 на 1 и наоборот. Обозначим через n длину выбираемой подпоследовательности $r[\mathbf{t}]$, через n_i — количество цифр, выбранных правилом r из i-ого 1-отрезка последовательности, а через n' — общее количество цифр, выбранных правилом r из 0-отрезков последовательности (и таким образом $n \geqslant \sum_{i=1}^{m} n_i + n'$ и $n_i \leqslant k$). Последовательность $\mathbf{t}$ назовём r-типичной, если выполнены следующие условия:

- количество единиц, выбранных из i-го 1-отрезка, меньше $n_i p_i + \sqrt{\frac{kn_i}{m^3}}$;
- либо $n' < k/m$, либо среди цифр, выбранных из 0-отрезков, количество единиц меньше количества нулей.

Определим вспомогательную меру на последовательностях длины L: цифры появляются независимо, цифры из i-го 0-отрезка (1-отрезка) с вероятностью p_i равны нулю (соответственно, единице), цифры из „остатка“ равны нулю с вероятностью $\frac{1}{2}$.

Оценим вероятность (по введённой мере) того, что наугад выбранная последовательность $\mathbf{t}$ длины L не принадлежит множеству S:

$$\Pr\{\mathbf{t} \notin S\} \leqslant \sum_{i=1}^{m} (\Pr\{\mathbf{t}_{i,0} \notin S_{i,0}\} + \Pr\{\mathbf{t}_{i,1} \notin S_{i,1}\}) \leqslant$$

$$2m \cdot \mathrm{e}^{-2k\left(\frac{1}{m}\right)^2} = \mathrm{e}^{-\frac{2k}{m^2}\left(1-\frac{m^2 \ln 2m}{2k}\right)} .$$

Теперь зафиксируем какое-нибудь правило выбора и оценим вероятность того, что наугад выбранная последовательность $\mathbf{t}$ длины L не r-типична (как и в определении типичности, мы считаем, что в выборке $r[\mathbf{t}]$ количество единиц больше количества нулей; противоположный случай аналогичен).

Сначала оценим вероятность того, что хотя бы для одного i из i-ого 1-отрезка выбрано по крайней мере $n_i p_i + \sqrt{\frac{kn_i}{m^3}}$ единиц. Она не превышает суммы вероятностей соответствующих событий для каждого i. При фиксированном $n_i > 0$, пользуясь оценкой Чернова, получаем

$$\mathrm{e}^{-2\frac{1}{n_i}\left(\sqrt{\frac{kn_i}{m^3}}\right)^2} .$$

Поскольку n_i нам неизвестно, мы просуммируем вероятности по всем возможным значениям $n_i \leqslant k$. Окончательная верхняя оценка:

$$m \cdot k \cdot \mathrm{e}^{-\frac{2k}{m^3}} .$$

Оценим теперь вероятность того, что из 0-отрезков выбрано $n' \geqslant \frac{k}{m}$ цифр и среди этих цифр количество единиц не меньше количества нулей. При заданном n' она не превышает

$$\mathrm{e}^{-2n'\left(p_m - \frac{1}{2}\right)^2} \leqslant \mathrm{e}^{-2\frac{k}{m}\frac{1}{m \ln m}} .$$

Суммируя по всем возможным значениям $n' \leqslant mk$, окончательно получаем оценку

$$mk\,\mathrm{e}^{-\frac{2k}{m^2 \ln m}} .$$

Мы видим, что оценка вероятности нарушения первого условия r-типичности гораздо больше оценки вероятности нарушения второго условия. Учитывая симметричный случай (замена 0 на 1), получим, что вероятность не r-типичности наугад взятой последовательности меньше $2mk\,\mathrm{e}^{-\frac{L}{m^4}+1} \leqslant \mathrm{e}^{-\frac{L}{m^4}+\ln L+1}$. Умножая на количество правильных множеств, покрывающих S, найдём, что вероятность „не быть типичной последовательностью хотя бы для одного правила“ значительно меньше 1 при $\delta < \mathrm{e}^{-\mathrm{e}^8}$, $L \geqslant (1/\delta)^5$. После добавления малой вероятности не принадлежать S эта оценка по-прежнему будет меньше единицы, поэтому существует последовательность $\mathbf{t} \in S$, которая типична для всех правил из покрытия.

Рассмотрим то множество A из покрытия S, которое содержит $\mathbf{t}$, и соответствующее правило r. Оценим сверху возможную величину удельного дефекта A, используя r-типичность $\mathbf{t}$. Мы рассмотрим отдельно два случая.

1) $n \leqslant 8k/(\ln m \ln 2)$. Дефект A не больше n, поэтому удельный дефект не больше
$$\frac{n}{L} \leqslant \frac{4}{m \ln m \ln 2}\,.$$

2) $n > 8k/(\ln m \ln 2)$. Оценим сверху разность между количеством единиц в $r[\mathbf{t}]$ и половиной длины выборки (для нулей рассуждение симметрично). Она складывается из соответствующих разностей для 1-отрезков, для 0-отрезков и для „остатка“. Для „остатка“ разность оценим просто половиной его длины, которая меньше m. В силу типичности на 0-отрезках либо общая длина выборки (а значит и удвоенная разность) меньше k/m, либо в ней нулей больше чем единиц. На 1-отрезках также в силу типичности разность меньше
$$\sum_{i=1}^{m}\left(\frac{n_i}{\sqrt{i \ln m}} + \sqrt{\frac{kn_i}{m^3}}\right) = \frac{1}{\sqrt{\ln m}}\sum_{i=1}^{m}\left(\frac{n_i}{\sqrt{i}}\right) + \sqrt{\frac{k}{m^3}}\sum_{i=1}^{m}\sqrt{n_i}\,.$$
Так как корень — выпуклая функция,
$$\sum_{i=1}^{m}\sqrt{n_i} \leqslant m\sqrt{\sum_{i=1}^{m} n_i/m} \leqslant \sqrt{nm}\,.$$
С одной стороны, $\sum\limits_{i=1}^{m}\left(n_i/\sqrt{i}\right) \leqslant n$. С другой стороны, чтобы оценить $\sum\limits_{i=1}^{m}\left(n_i/\sqrt{i}\right)$ при $n > 4k$, заметим, что если n_i увеличить на

единицу, а n_{i+1} уменьшить на единицу, то сумма увеличится. Мы знаем, что $n_i \leqslant k$, $\sum n_i \leqslant n$, поэтому увеличивая первые n_i до k, получаем:

$$\sum_{i=1}^{m} \left(\frac{n_i}{\sqrt{i}}\right) \leqslant \sum_{i=1}^{\lceil n/k \rceil} \left(\frac{k}{\sqrt{i}}\right) < k \cdot 2\sqrt{n/k} = 2\sqrt{nk}\,.$$

Окончательно получаем, что разность между количеством единиц в $r[\mathbf{t}]$ и $n/2$ меньше

$$m + \frac{k}{2m} + \frac{\sqrt{nk}}{m} + \frac{n}{\sqrt{\ln m}} < \frac{n}{\sqrt{\ln m}} \left(1 + \frac{\sqrt{\ln^3 m}}{m}\right) \quad \text{при } n \leqslant 4k\,;$$
$$m + \frac{k}{2m} + \frac{\sqrt{nk}}{m} + \frac{2\sqrt{nk}}{\sqrt{\ln m}} < \frac{2\sqrt{nk}}{\sqrt{\ln m}} \left(1 + \frac{\sqrt{\ln m}}{m}\right) \quad \text{при } n > 4k\,.$$

Поскольку правильное множество A содержит последовательность $\mathbf{t}$, его удельный дефект не больше

$$\frac{1}{L} \cdot \frac{2}{n \ln 2} \cdot \frac{n^2}{\ln m}(1 + o(1)) \leqslant \frac{4}{m \ln m \ln 2}(1 + o(1)) \quad \text{при } n \leqslant 4k\,;$$
$$\frac{1}{L} \cdot \frac{2}{n \ln 2} \cdot \frac{4nk}{\ln m}(1 + o(1)) \leqslant \frac{4}{m \ln m \ln 2}(1 + o(1)) \quad \text{при } n > 4k\,.$$

Нижние оценки на биномиальный коэффициент

$$\binom{n}{j} \geqslant \frac{\mathrm{e}^{n \cdot h(j/n)}}{\sqrt{8j(n-j)/n}}$$

и на шенноновскую функцию энтропии (при $\varepsilon \leqslant \frac{1}{\sqrt{12}}$)

$$h\left(\frac{1}{2} + \varepsilon\right) \geqslant \ln 2 - 2\varepsilon^2(1 + \varepsilon^2)$$

показывают, что величины $o(1)$, входящие в предыдущие выражения, меньше $1/\ln(1/\delta)$. С другой стороны, $\ln m > \ln(1/\delta)(1 + 1.05/\ln(1/\delta))$.

Во всех случаях хотя бы у одного правильного множества из покрытия S удельный дефект меньше

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

$\square$

Теорема 6. *Для любого $\sigma > 0$ и для любого $L \geqslant 12 + 6\,\mathrm{lb}(1/\sigma)$ существует множество двоичных последовательностей длины L с удельным дефектом $\geqslant 1/3$, которое нельзя покрыть менее чем $2^{\sigma L/2}$ неадаптивными правильными множествами с удельным дефектом $\geqslant \sigma$.*

Доказательство. Рассмотрим упорядоченный набор из $2^{\lfloor 2L/3 \rfloor}$ последовательностей длины L (возможно, с повторениями). Множество входящих в него последовательностей имеет удельный дефект не меньше $1/3$. Поэтому для доказательства теоремы достаточно найти такой набор, который нельзя покрыть никаким семейством, состоящим из менее чем $2^{\sigma L/2}$ неадаптивных правильных множеств с удельным дефектом не меньше σ (можно считать, что $\sigma \leqslant 1$).

Количество наборов из $2^{\lfloor 2L/3 \rfloor}$ последовательностей равно

$$\left(2^L\right)^{2^{\lfloor 2L/3 \rfloor}}\,.$$

Оценим количество наборов, которые можно покрыть семействами указанного типа.

Зафиксируем какое-нибудь такое семейство $\mathcal{A}_L$. Количество покрываемых им наборов равно

$$|\cup\mathcal{A}_L|^{2^{\lfloor 2L/3 \rfloor}}\,.$$

Из условия на $\mathcal{A}_L$ имеем:

$$|\cup\mathcal{A}_L| \leqslant |\mathcal{A}_L| \max\{|A| \colon A \in \mathcal{A}_L\} < 2^{\sigma L/2} \cdot 2^{L-\sigma L} = 2^{L-\sigma L/2}\,.$$

Таким образом, количество наборов, покрываемых одним семейством правильных множеств, меньше

$$2^{(L-\sigma L/2)\cdot 2^{\lfloor 2L/3 \rfloor}}\,.$$

Семейству $\mathcal{A}_L$ соответствует некоторое семейство неадаптивных правил $\mathcal{R}_L$. Выбирая различные отклонения, мы получаем, вообще говоря, различные семейства правильных множеств (большого дефекта), соответствующие одному и тому же множеству правил. Однако с ростом отклонения класс покрываемых наборов может только уменьшаться, поэтому одно семейство правил мощности $< 2^{\sigma L/2}$ обеспечивает покрытие также менее

$$2^{(L-\sigma L/2)\cdot 2^{\lfloor 2L/3 \rfloor}}$$

наборов.

Заметим, что неадаптивное правило по сути является подмножеством $\{1, \ldots, L\}$, поэтому количество неадаптивных правил равно 2^L. Количество семейств, состоящих из менее чем $2^{\sigma L/2}$ неадаптивных правил, меньше

$$2^{L\cdot 2^{\sigma L/2}}.$$

Поэтому с помощью таких семейств можно покрыть менее

$$2^{L\cdot 2^{\sigma L/2}} \cdot 2^{(L-\sigma L/2)\cdot 2^{\lfloor 2L/3 \rfloor}} = \left(2^L\right)^{2^{\sigma L/2}+\left(1-\frac{\sigma}{2}\right)2^{\lfloor 2L/3 \rfloor}}$$

наборов. Чтобы убедиться в существовании не покрытых наборов, достаточно проверить, что

$$\left(2^L\right)^{2^{\sigma L/2}+\left(1-\frac{\sigma}{2}\right)2^{\lfloor 2L/3 \rfloor}} \leqslant \left(2^L\right)^{2^{\lfloor 2L/3 \rfloor}}.$$

Последнее следует из $L(2/3 - \sigma/2) \geqslant 2 - \operatorname{lb}\sigma$. Это неравенство вытекает из $L \geqslant 12 + 6\operatorname{lb}(1/\sigma)$ и $\sigma \leqslant 1$. □

2.3. Алгоритмический подход

Определение 5. *Дефектом* непустой двоичной последовательности $\mathbf{t}$ длины L называется величина $L - K(\mathbf{t}|L)$, где $K(\mathbf{t}|L)$ — энтропия $\mathbf{t}$ при известном L. *Удельным дефектом* называется дефект, делённый на длину. (Аддитивная константа из определения энтропии, делённая на длину последовательности, стремится к нулю, когда длина стремится к бесконечности. Так что в пределе понятие удельного дефекта инвариантно.)

Так как в доказательствах для оценки дефекта последовательностей мы будем использовать дефект вспомогательных правильных множеств, введём следующее обозначение. Если нормальное правило r выбирает подпоследовательности длины n, то дефект правильного множества $A_{r,\varepsilon}$ будет обозначаться через $D(n,\varepsilon)$. Важным свойством дефекта правильного множества является то, что он не зависит от L (длины последовательностей, к которым применяется правило выбора). Если n фиксировано, то D становится монотонно возрастающей ступенчатой функцией от $\varepsilon \in [0, \frac{1}{2}]$. При этом границы между ступенями — рациональные числа со знаменателем $2n$, значения функции — это двоичные логарифмы рациональных чисел, и все эти рациональные числа вычислимо зависят от n.

Когда ε достаточно мало, а n достаточно велико по сравнению с $1/\varepsilon$, то $D(n,\varepsilon) \sim d(n,\varepsilon)$.

На первый взгляд естественным аналогом теоремы 4 была бы такая

Формулировка 1. *Пусть $\delta > 0$ достаточно мало и натуральное L достаточно велико. Если $\mathbf{t}$ — двоичная последовательность длины L с удельным дефектом $\geqslant \delta$, то существует правило r, для которого*

$$K(r|L) \leqslant \alpha(\delta) \quad \textit{и} \quad D(n,\varepsilon)/L \geqslant \delta'(\delta)\,,$$

где n — длина $r[\mathbf{t}]$, ε — отклонение доли нулей в $r[\mathbf{t}]$ от $\frac{1}{2}$, а α и δ' — некоторые положительные функции.

Однако следующая теорема очевидно влечёт отрицание формулировки 1.

Теорема 4$''$. *Пусть α и δ' — положительные числа. Для любого достаточно большого натурального L найдётся такая двоичная последовательность $\mathbf{t}$ длины L с удельным дефектом больше $\frac{1}{2}$, что для каждого правила r*

$$K(r|L) \leqslant \alpha \quad \Rightarrow \quad D(n,\varepsilon)/L < \delta' \vee n = 0\,,$$

где n — длина $r[\mathbf{t}]$, ε — отклонение доли нулей в $r[\mathbf{t}]$ от $\frac{1}{2}$.

Доказательство. Обозначим через $\mathcal{R}_L$ семейство всех правил r, для которых $K(r|L) \leqslant \alpha$. Зная числа $\lfloor\alpha\rfloor$, L и $|\mathcal{R}_L|$, можно эффективно найти список элементов $\mathcal{R}_L$. Обозначим n-нормализацию правила r через r_n.

Рассмотрим на двоичных последовательностях длины L равномерное распределение. Из определения функции D следует, что для каждого правила r вероятность множества $A_{r_n,\varepsilon}$ равна $2^{-D(n,\varepsilon)}$.

Для каждого n возьмём наименьшую границу ступени ε, для которой $D(n,\varepsilon) \geqslant \delta' L$. Если такое ε существует, оно будет рациональным числом из $[0, \frac{1}{2}]$ со знаменателем $2n$. Полученный набор пар $\langle n, \varepsilon \rangle$ обозначим через E. Поскольку $|\mathcal{R}_L| < 2^{\alpha+1}$, а $|E| \leqslant L$, то вероятность объединения множеств $A_{r_n,\varepsilon}$ для $r \in \mathcal{R}_L$ и $\langle n, \varepsilon \rangle \in E$ не превышает

$$\sum_{\langle n,\varepsilon \rangle \in E} 2^{-D(n,\varepsilon)+\alpha+1} \leqslant 2^{-\delta' L+\alpha+1+\mathrm{lb}\, L}\,.$$

При достаточно больших L последнее выражение строго меньше 1. Зная списки элементов $\mathcal{R}_L$ и E, можно перебором найти последовательность $\mathbf{t}$ длины L, которая не принадлежит $A_{r_n,\varepsilon}$ при всех $r \in \mathcal{R}_L$ и $\langle n, \varepsilon \rangle \in E$.

Теперь постараемся экономно закодировать информацию, достаточную для построения списка элементов E. Понятно, что набор E можно построить, если известно L и та пара, на которой функция D достигает своего минимума на E. Таким образом, энтропия списка элементов E при известном L не превышает $2\,\mathrm{lb}\, L + O(1)$.

Так как последовательность $\mathbf{t}$ построена по числам L, $\lfloor \alpha \rfloor$, $|\mathcal{R}_L|$ и набору E, то $K(\mathbf{t}|L) \leqslant 2\,\mathrm{lb}\, L + C(\alpha)$. Удельный дефект $\mathbf{t}$ равен $1 - K(\mathbf{t}|L)/L$, что больше $\frac{1}{2}$ при достаточно больших L.

Пусть $r \in \mathcal{R}_L$, $n = \ell(r[\mathbf{t}])$ и ε — отклонение в $r[\mathbf{t}]$. Предположим, что $D(n,\varepsilon)/L \geqslant \delta'$. Тогда по построению $\mathbf{t}$ не принадлежит $A_{r_n,\varepsilon}$, и поэтому отклонение в $r[\mathbf{t}]$ должно быть строго меньше ε. Противоречие, теорема доказана. $\square$

Для восстановления аналогии с комбинаторным подходом при алгоритмическом подходе правила выбора надо понимать не как явно заданные функции, а как программы, которые, получив в качестве дополнительного входа длину последовательности L, вычисляют функцию r из определения 1, причём теперь функция r может быть *не всюду определённой*.

Определение 6. *Частотным α-дефектом* двоичной последовательности $\mathbf{t}$ длины L называется максимум величины $D(n,\varepsilon)/L$ по правилам r, заданным такими программами, что $r[\mathbf{t}]$ определено и энтропия программы при известном L меньше α (где n — длина $r[\mathbf{t}]$, а ε — отклонение доли нулей в $r[\mathbf{t}]$ от $\frac{1}{2}$).

Отметим, что с ростом α частотный α-дефект может только увеличиваться.

Теорема 4′. *Пусть $\delta_1 > 0$ достаточно мало. Тогда по каждому рациональному $\delta_0 > 0$ можно эффективно указать такое L_0, что для любых $\delta \in [\delta_0, \delta_1]$ и натурального $L \geqslant L_0$ выполнено следующее. Если $\mathbf{t}$ — двоичная последовательность длины L с удельным дефектом δ, то частотный α-дефект $\mathbf{t}$ больше*

$$\delta' = \frac{\delta}{\ln(1/\delta)}(1 - B(\delta)),$$

где $\alpha = \frac{4\ln(1/\delta)}{\delta}$ и $B(\delta) = O(\frac{\ln\ln(1/\delta)}{\ln(1/\delta)})$.

Замечание. Мы видим, что данная теорема соответствует формулировке 1.

Доказательство. Строим правило, выбирающее из $\mathbf{t}$ подпоследовательность с большим отклонением. Разделим последовательность $\mathbf{t}$ на две части примерно равной длины (то есть их длины равны $\lfloor L/2 \rfloor$ и $\lceil L/2 \rceil$). Каждую часть снова разделим на две части примерно равной длины и так далее. Перед $(i+1)$-м этапом рекурсивного построения нашего правила будет указана одна из частей $\mathbf{t}$, получившаяся после i делений (перед первым этапом указанная часть — это сама последовательность $\mathbf{t}$). Обозначим эту часть через $\mathbf{u}_i$, а её дополнение[10] через $\mathbf{v}_i$. Предполагается, что перед $(i+1)$-м этапом перевёрнуты были в точности карты из $\mathbf{v}_i$, причём ни одна не была включена в подпоследовательность. Предполагается также, что условный удельный дефект $\mathbf{u}_i$ при известном $\mathbf{v}_i$ больше $\delta + \lambda(i-1)$, где λ — такое число из интервала $(\frac{\delta}{\ln(1/\delta)}, \frac{2\delta}{\ln(1/\delta)})$, что число $1 - \delta + 3\lambda$ рационально. При больших L последнее предположение, очевидно, выполнено для $i = 0$.

Теперь опишем $(i+1)$-ый этап. Разобьём последовательность $\mathbf{u}_i$ на две примерно равные по длине части, обозначим их $\mathbf{w}_1$ и $\mathbf{w}_2$. Пусть $\gamma = \lfloor (1 - \delta + 3\lambda)\ell(\mathbf{w}_1) \rfloor$. Рассмотрим два возможных случая.

1) Пусть $K(\mathbf{w}_1|\mathbf{v}_i) < \gamma$ и $K(\mathbf{w}_2|\mathbf{v}_i) < \gamma$.

   Поскольку функция энтропии перечислима сверху, при известных $\mathbf{v}_i$ и γ эти факты можно обнаружить, скажем, за T_1 и T_2 шагов перечисления. Числа T_1 и T_2 нам неизвестны. Наша цель — узнать

[10] Под дополнением мы понимаем двоичный код результата замены в последовательности $\mathbf{t}$ цифр из $\mathbf{u}_i$ на некоторый специальный знак.

большее из них. Пусть, например, $T_1 \leqslant T_2$. Тогда правило должно перевернуть все карты из $\mathbf{w}_2$, ничего не включая в подпоследовательность. Зная $\mathbf{w}_2$, можно дождаться его появления в упомянутом перечислении и тем самым узнать T_2. Теперь рассмотрим S — множество последовательностей $\mathbf{s}$ длины $\ell(\mathbf{w}_1)$, для которых выполнение неравенства $K(\mathbf{s}|\mathbf{v}_i) < \gamma$ обнаруживается не более чем за T_2 шагов. Очевидно, что $\mathbf{w}_1 \in S$. Множество S содержит менее 2^γ элементов, поэтому его удельный дефект больше $1-\gamma/\ell(\mathbf{w}_1) \geqslant \delta - 3\lambda$. (Если бы оказалось, что $T_1 \geqslant T_2$, то получилось бы аналогичное неравенство, поскольку $\ell(\mathbf{w}_1) \leqslant \ell(\mathbf{w}_2)$.)

Теперь применим теорему 4 к множеству S с указанной оценкой на удельный дефект. Из доказательства теоремы 4 вытекает существование такого семейства $\mathcal{R}$ правил, что

$$|\mathcal{R}| \leqslant \left(\frac{1}{\delta - 3\lambda}\right)^{1/(\delta-3\lambda)+O(1)}$$

и правильные множества, задаваемые нормализациями правил из $\mathcal{R}$, покрывают S. Одно из этих правильных множеств (назовём его A) содержит $\mathbf{w}_1$ и имеет удельный дефект $\geqslant \frac{\delta-3\lambda}{\ln(1/(\delta-3\lambda))}(1-\beta)$, где $\beta = O\left(\frac{\ln\ln(1/(\delta-3\lambda))}{\ln(1/(\delta-3\lambda))}\right) = O\left(\frac{\ln\ln(1/\delta)}{\ln(1/\delta)}\right)$. Пусть нормальное правило, задающее множество A, выбирает из $\mathbf{w}_1$ подпоследовательность длины n с отклонением ε. Поскольку дефект A не больше $D(n,\varepsilon)$, то

$$\begin{aligned}\frac{D(n,\varepsilon)}{L} \geqslant \frac{\delta - 3\lambda}{\ln(1/(\delta-3\lambda))}(1-\beta) = \\ \frac{\delta}{\ln(1/\delta)}\left(1 - O\left(\frac{1}{\ln(1/\delta)}\right)\right)(1-\beta) = \\ \frac{\delta}{\ln(1/\delta)}\left(1 - O\left(\frac{\ln\ln(1/\delta)}{\ln(1/\delta)}\right)\right).\end{aligned}$$

После реализации случая 1) правило выбора завершает работу.

2) Пусть $K(\mathbf{w}_1|\mathbf{v}_i) \geqslant \gamma$ или $K(\mathbf{w}_2|\mathbf{v}_i) \geqslant \gamma$.

Пусть для определённости $K(\mathbf{w}_1|\mathbf{v}_i) \geqslant \gamma$. По теореме об энтропии пары $K(\mathbf{u}_i|\mathbf{v}_i) = K(\mathbf{w}_1|\mathbf{v}_i) + K(\mathbf{w}_2|\mathbf{w}_1,\mathbf{v}_i) - O(\log K(\mathbf{u}_i|\mathbf{v}_i))$ (понятно, что можно отождествлять пару $\langle \mathbf{w}_1, \mathbf{v}_i \rangle$ с дополнением до $\mathbf{w}_2$). Следовательно, $K(\mathbf{w}_2|\mathbf{w}_1, \mathbf{v}_i) = K(\mathbf{u}_i|\mathbf{v}_i) - K(\mathbf{w}_1|\mathbf{v}_i) + O(\log K(\mathbf{u}_i|\mathbf{v}_i)) \leqslant (1-\delta-\lambda(i-1))\ell(\mathbf{u}_i) - (1-\delta+3\lambda)\ell(\mathbf{w}_1) + O(\log \ell(\mathbf{u}_i)) \leqslant (1-\delta-2\lambda i-\lambda)\ell(\mathbf{w}_2) + O(\log \ell(\mathbf{u}_i))$. Отсюда при большом $\ell(\mathbf{u}_i)$ просто вывести следующее неравенство на условный удельный дефект $\mathbf{w}_2$ при известных $\mathbf{w}_1$ и $\mathbf{v}_i$:

$$1 - \frac{K(\mathbf{w}_2|\mathbf{w}_1, \mathbf{v}_i)}{\ell(\mathbf{w}_2)} > \delta + \lambda i\,.$$

После этого правило выбора переходит к $(i+2)$-ому этапу с $\mathbf{u}_{i+1} = \mathbf{w}_2$.

Так как удельный дефект не бывает больше 1, то второй случай реализуется менее $(1-\delta)/\lambda$ раз. Поэтому количество этапов в построенном правиле выбора не превышает некоторого числа, не зависящего от L. Отсюда следует, что при достаточно большом L предположение о том, что $\ell(\mathbf{u}_i)$ велико, выполнено для всех i.

Осталось оценить энтропию (при известном L) построенного правила. При его построении мы пользовались следующей информацией:

- чему равно $1-\delta+3\lambda$;
- для каждого этапа — выполнен первый или второй случай;
- для второго случая (на каждом этапе): какое из неравенств верно: $K(\mathbf{w}_1|\mathbf{v}_i) \geqslant \lfloor (1-\delta+3\lambda)\ell(\mathbf{w}_1) \rfloor$ или $K(\mathbf{w}_2|\mathbf{v}_i) \geqslant \lfloor (1-\delta+3\lambda)\ell(\mathbf{w}_1) \rfloor$;
- для первого случая (на одном этапе — последнем): какое из неравенств верно — $T_1 \leqslant T_2$ или $T_1 \geqslant T_2$, а также номер множества A в семействе $\mathcal{R}$.

Для нахождения рационального числа $1-\delta+3\lambda$ достаточно найти число $\delta-3\lambda$, а для этого достаточно найти натуральное число из интервала $((\delta-\frac{3\delta}{\ln(1/\delta)})^{-1}, (\delta-\frac{6\delta}{\ln(1/\delta)})^{-1})$. Поскольку длина указанного интервала больше 1, в нём есть натуральное число, и энтропия этого числа не больше $\mathrm{lb}(1/\delta) + O(1)$.

Итак, верхняя оценка на энтропию нашего правила получится, если к следующей сумме добавить несколько её логарифмов (нужных для кодирования кортежа):

$$\operatorname{lb}\frac{1}{\delta}+O(1)+2\cdot\left(\frac{1-\delta}{\lambda}+O(1)\right)+\operatorname{lb}|\mathcal{R}|+O(1)\,.$$

При малых δ найденная оценка меньше $\frac{4\ln(1/\delta)}{\delta}$.

Заметим наконец, что когда мы говорили „L достаточно велико“, соответствующая нижняя оценка на L могла быть эффективно указана по δ_0. Значение L_0 есть максимум этих оценок. □

Замечание. Если в теореме 4 достаточно ограничиться правильными множествами, порождёнными монотонными правилами (а неадаптивных правил не хватает), то для теоремы $4'$ необходимы немонотонные правила, как показывает теорема $6'$.

Теорема $5'$. *Пусть $\delta_0>0$ достаточно мало и натуральное $L\geqslant(1/\delta_0)^5$. Существует двоичная последовательность $\mathbf{t}$ длины L с удельным дефектом больше δ_0, у которой частотный α-дефект меньше*

$$\frac{2\delta_0}{\ln(1/\delta_0)}(1+3\delta_0)$$

при $\alpha=L\delta_0^4/70$.

Доказательство. Возьмём множество правил, у которых энтропия при известном L меньше α. Некоторые из этих правил не всюду определены. Доопределим их произвольным образом и произведём n-нормализацию для всех $n=1,\ldots,L$. Получившихся нормальных правил меньше $L\cdot 2^\alpha<\mathrm{e}^{L\delta_0^4/70}$. Рассмотрим множество последовательностей S, построенное (эффективно) в теореме 5 по числам $\delta=\frac{1}{\lfloor 1/\delta_0\rfloor-1}$ и L. Поскольку $\delta>\delta_0$, то $L>(1/\delta)^5$, $\mathrm{e}^{L\delta^4/70}>\mathrm{e}^{L\delta_0^4/70}$ и по теореме 5 найдётся $\mathbf{t}\in S$, для которого

$$\max\{D(n,\varepsilon)/L\mid n=\ell(r[\mathbf{t}]),K(r|L)<\alpha\}<\frac{2\delta}{\ln(1/\delta)}<\frac{2\delta_0}{\ln(1/\delta_0)}(1+3\delta_0)\,.$$

Энтропия любого $\mathbf{t}\in S$ при известном L не превышает $2\operatorname{lb}(1/\delta)+\operatorname{lb}|S|$. Поэтому удельный дефект $\mathbf{t}$ не меньше $1-\frac{\operatorname{lb}|S|+2\operatorname{lb}(1/\delta)}{L}$. По теореме 5 удельный дефекет S (равный $1-\operatorname{lb}|S|/L$) больше δ. Следовательно, удельный дефект $\mathbf{t}$ больше $\delta-2\operatorname{lb}(1/\delta)/L>\delta_0$. □

Определение 7. Назовём *монотонным α-частотным дефектом* двоичной последовательности $\mathbf{t}$ длины L максимум величины $D(n,\varepsilon)/L$ по монотонным правилам r, заданным программами энтропии меньше α, где n — длина $r[\mathbf{t}]$, ε — отклонение в $r[\mathbf{t}]$, в энтропии предполагается условие L.

Теорема 6′. *Пусть $\delta > 0$ достаточно мало и натуральное $L \geqslant (1/\delta)^2$. Существует двоичная последовательность $\mathbf{t}$ длины L с удельным дефектом больше $\frac{1}{2}$, для которой монотонный $(\delta L/4)$-частотный дефект меньше δ.*

Доказательство. Для построения последовательности $\mathbf{t}$ мы рассмотрим игру, похожую на ту, которая была использована в доказательстве теоремы 4. Отличие заключается в том, что величина ставки может принимать любые значения, не превышающие текущего капитала (поэтому игра называется „игрой на наличные“). Начальный капитал равен 1. Удобно называть ставкой долю σ текущего капитала; тогда текущий капитал умножается на $(1+\sigma)$ при выигрыше (если следующая цифра угадана) и на $(1-\sigma)$ при проигрыше.

Каждому монотонному правилу r, двоичной цифре b и числу $\sigma \in [0,1]$ сопоставим следующую стратегию в игре: на шаге i правило r применяется к отрезку $\mathbf{t}_{1:i-1}$; если следующая цифра выбрана в подпоследовательность, делается ставка σ на цифру b, если же следующая цифра не выбрана, ставка не делается. Если r на некотором $\mathbf{t}_{1:i-1}$ не определено (программа не останавливается), то на всех последующих шагах стратегия не определена и текущий капитал тоже становится неопределённым. Очевидно, что если в выбранной подпоследовательности $r[\mathbf{t}]$ длины n цифра b встречается n_b раз, то конечный капитал стратегии равен $(1+\sigma)^{n_b}(1-\sigma)^{n-n_b}$ (а отклонение равно $\left|\frac{1}{2} - \frac{n_b}{n}\right|$, поэтому из оценки на капитал можно получить оценку на отклонение).

Сначала предположим, что $1/\delta$ — натуральное число. В конце объясним, как перейти к произвольному δ. Пусть $\mathcal{R}_L$ — множество всех монотонно выбирающих программ энтропии меньше $\delta L/4$, а R — их количество. Без ограничения общности можно считать, что каждому кодовому слову сопоставлена монотонно выбирающая программа (возможно, нигде не определённая) и различным кодовым словам сопоставлены различные программы. Так что $R = 2^{\lceil \delta L/4 \rceil} - 1$. Для $n = \left\lceil \frac{\ln(4LR)}{\delta} \right\rceil, \ldots, L$ определим ставки $\sigma_n = \sqrt{\frac{\ln(4LR)}{n}}$. Будем рассматривать множество стратегий, сопоставленных всем возможным σ_n, $r \in \mathcal{R}_L$ и $b \in \{0, 1\}$.

Для построения последовательности $\mathbf{t}$ нам понадобятся некоторые вспомогательные величины, зависящие от $\mathbf{t}$. Пусть S_i — суммарный капитал всех ещё определённых стратегий на шаге i (он зависит только от $\mathbf{t}_{1:i-1}$). Начальный капитал S_0 равен количеству стратегий, то есть меньше $2LR$.

Идея построения $\mathbf{t}$ следующая: на каждом шаге выбираем t_i так, чтобы суммарный капитал не возрастал; тогда конечный капитал любой стратегии не превышает суммарного начального капитала S_0, что даёт оценку на отклонение в подпоследовательности. Однако непосредственно этот подход неосуществим, поскольку мы не можем алгоритмически проверить, какие правила определены на данном начале, а какие — нет, а значит, не можем вычислить S_{i+1} при каждом из двух возможных значений t_i. Поэтому на каждом шаге алгоритма необходима некоторая дополнительная информация.

Пусть Q_i — суммарный капитал на шаге i тех стратегий, которые будут также определены и на шаге $i+1$ (то есть на начале $\mathbf{t}_{1:i}$). Очевидно, что $Q_i \leqslant S_i$.

Определим последовательности P_i и m_i следующими соотношениями:

$$P_0 = S_0\,,$$
$$m_i = \left\lfloor \frac{P_i - Q_i}{2R}\sqrt{\delta} \right\rfloor\,,$$
$$P_{i+1} = P_i + 2R - \frac{2Rm_i}{\sqrt{\delta}}\,.$$

Из определения ясно, что

$$P_i - \frac{2Rm_i}{\sqrt{\delta}} - \frac{2R}{\sqrt{\delta}} < Q_i \leqslant P_i - \frac{2Rm_i}{\sqrt{\delta}}\,.$$

Опишем теперь, как по началу $\mathbf{t}_{1:i}$ и заданным P_i и m_i строится следующая цифра t_{i+1}. Построение состоит из двух этапов:

1) К началу $\mathbf{t}_{1:i}$ параллельно применяются все стратегии; вычисляется текущий капитал тех, которые уже определились, до тех пор, пока их суммарный капитал не превысит $(P_i - \frac{2Rm_i}{\sqrt{\delta}} - \frac{2R}{\sqrt{\delta}})$.

2) Выбирается то значение t_{i+1}, при котором суммарный выигрыш стратегий, определившихся к концу этапа 1, неположителен.

Сначала докажем осуществимость алгоритма. Если начало $\mathbf{t}_{1:i}$ построено, то значения S_i, Q_i, P_i и m_i определены. Пусть алгоритм получил на вход корректные P_i и m_i. В процессе выполнения этапа 1 наступит момент, когда уже определились все стратегии, которые вообще определены на данном начале. Тогда их суммарный капитал равен $Q_i > P_i - \frac{2Rm_i}{\sqrt{\delta}} - \frac{2R}{\sqrt{\delta}}$, поэтому в некоторый момент этап 1 завершится. Теперь заметим, что для каждой (определённой) стратегии сумма выигрышей при $t_{i+1} = 0$ и $t_{i+1} = 1$ равна нулю. Равна нулю и сумма при $t_{i+1} = 0$ и $t_{i+1} = 1$ суммарных выигрышей любого множества (определённых) стратегий, поэтому суммарный выигрыш на этапе 2 не может быть положительным для обоих значений t_{i+1}.

Нам понадобятся следующие три свойства введённых величин:

1) $S_i \leqslant P_i$.

   $S_0 = P_0$ по определению. Оценим S_{i+1}. Для тех стратегий, которые определились в течение этапа 1, суммарный капитал (обозначим его Q_i') не возрос. Капитал остальных стратегий равен $Q_i - Q_i'$, он увеличился не более чем в $(1+\sqrt{\delta})$ раз, поскольку все ставки не превышают $\sqrt{\delta}$. Очевидно, что $Q_i - Q_i' < Q_i - (P_i - \frac{2Rm_i}{\sqrt{\delta}} - \frac{2R}{\sqrt{\delta}}) \leqslant \frac{2R}{\sqrt{\delta}}$. Таким образом,

$$S_{i+1} \leqslant Q_i' + (Q_i - Q_i')(1+\sqrt{\delta}) = Q_i + (Q_i - Q_i')\sqrt{\delta} < (P_i - \frac{2Rm_i}{\sqrt{\delta}}) + \frac{2R}{\sqrt{\delta}}\sqrt{\delta} = P_{i+1}\,.$$

2) $m_i \geqslant 0$.

   Неравенство следует из того, что $Q_i \leqslant S_i \leqslant P_i$.

3) $\sum m_i < 2L\sqrt{\delta}$.

   Просуммируем равенства $P_{i+1} = P_i + 2R - \frac{2Rm_i}{\sqrt{\delta}}$ по всем i. Получим $P_L = P_0 + 2RL - 2R\sum m_i/\sqrt{\delta}$, откуда с учётом того, что $P_0 < 2RL$ и $P_L > S_L \geqslant 0$, имеем

$$\sum m_i = (2RL + P_0 - P_L)\frac{\sqrt{\delta}}{2R} < 2L\sqrt{\delta}\,.$$

Теперь мы можем оценить энтропию построенной последовательности $\mathbf{t}$. Алгоритм построения $\mathbf{t}$ использовал знание множества стратегий и последовательностей P_i, m_i. Множество стратегий строится по L и δ. P_i строится по m_i, L и δ. Поэтому $K(\mathbf{t}) \leqslant K(\langle m_0, \ldots, m_{L-1}\rangle | L, 1/\delta) + 2\,\mathrm{lb}\, L + 2\,\mathrm{lb}(1/\delta) = K(\langle m_0, \ldots, m_{L-1}\rangle | L, 1/\delta) + o(L)$.

Энтропию $\langle m_0, \ldots, m_{L-1}\rangle$ при известных L, $1/\delta$ оценим через двоичный логарифм количества таких последовательностей из L натуральных чисел, что сумма этих чисел меньше $N = \left\lceil 2L\sqrt{\delta}\right\rceil$. Легко показать, что последовательность $\langle m_0, \ldots, m_{L-1}\rangle$ однозначно определяется мультимножеством (с учётом кратностей) своих частичных сумм $\{m_0, m_0 + m_1, \ldots, m_0 + \ldots + m_{L-1}\}$. Частичные суммы могут принимать только значения $\{0, \ldots, N-1\}$. Таким образом, количество возможных последовательностей m_i не превосходит количества неупорядоченных выборок с повторениями из N по L, которое равно $\binom{N+L}{L} < \frac{(N+L)^N}{N!} < \left(\frac{(N+L)\mathrm{e}}{N}\right)^N$. В результате,

$$K(\langle m_0, \ldots, m_{L-1}\rangle | L, 1/\delta) \leqslant \mathrm{lb}\binom{N+L}{L} + O(1) \leqslant N\,\mathrm{lb}\,\frac{(N+L)\mathrm{e}}{N} + O(1) \leqslant \left\lceil 2L\sqrt{\delta}\right\rceil \cdot \mathrm{lb}\left(1 + \frac{1}{2\sqrt{\delta}}\right)\mathrm{e} + O(1) < L/3\,.$$

Из того, что $K(\mathbf{t}) < L/2$, следует, что удельный дефект $\mathbf{t}$ больше $\frac{1}{2}$.

Осталось оценить монотонный частотный дефект построенной последовательности $\mathbf{t}$. Для этого зафиксируем правило r и рассмотрим выборку $r[\mathbf{t}]$ длины n с отклонением ε. Оценим сверху $D(n, \varepsilon)$. Рассмотрим два возможных случая.

1) $n \leqslant 7\delta L/8$. Так как $D(n, \varepsilon) \leqslant n$, то $D(n, \varepsilon)/L \leqslant \frac{7}{8}\delta$.

2) $n > 7\delta L/8$. Пусть отклонение в $r[\mathbf{t}]$ в сторону нулей. Возьмём стратегию, соответствующую правилу r, ставящую долю капитала σ_n на 0. Конечный капитал этой стратегии не превышает суммарного капитала $S_L < P_L \leqslant P_0 + 2RL < 4RL$ (мы воспользовались уже упомянутым равенством $P_L = P_0 + 2RL - 2R\sum m_i/\sqrt{\delta}$). С другой стороны, капитал стратегии равен $(1+\sigma_n)^{n\left(\frac{1}{2}+\varepsilon\right)}(1-\sigma_n)^{n\left(\frac{1}{2}-\varepsilon\right)}$. Получаем неравенство

$$\frac{n}{2}\left((1+2\varepsilon)\ln(1+\sigma_n) + (1-2\varepsilon)\ln(1-\sigma_n)\right) < \ln(4RL)\,,$$

откуда

$$2\varepsilon \ln \frac{1+\sigma_n}{1-\sigma_n} + \ln(1-\sigma_n^2) < \frac{2}{n}\ln(4RL)\,.$$

Используя то, что $\ln\frac{1+x}{1-x} \geqslant 2x$ при $0 \leqslant x < 1$, а также то, что $\ln(1-x) \geqslant -2x$ при $0 \leqslant x \leqslant \frac{1}{2}$, найдём, что

$$\varepsilon < \frac{\frac{2}{n}\ln(4LR) + 2\sigma_n^2}{4\sigma_n} = \frac{\ln(4LR)}{2n\sigma_n} + \frac{\sigma_n}{2} = \sqrt{\frac{\ln(4LR)}{n}}\,.$$

Ниже мы используем оценку

$$\binom{n}{j} \geqslant \frac{\mathrm{e}^{n\cdot h(j/n)}}{\sqrt{8j(n-j)/n}}\,,$$

вытекающую из формулы Стирлинга, и оценку (при $\varepsilon \leqslant \frac{1}{\sqrt{5}}$)

$$h\left(\frac{1}{2}+\varepsilon\right) \geqslant \ln 2 - 2\varepsilon^2(1+10\varepsilon^2/3)\,,$$

получающуюся двукратным дифференцированием. Так как

$$\begin{aligned}D(n,\varepsilon) &< 2n\varepsilon^2 \operatorname{lb}\mathrm{e}(1+10\varepsilon^2/3) + \ln n < \\ &2\ln(4LR)\operatorname{lb}\mathrm{e}\left(1+\frac{10\ln(4LR)}{3n}\right) + \ln n < \\ &2\operatorname{lb}(4LR)\left(1+\frac{10\ln(4LR)}{3(7\delta L/8)}\right) + \ln n \leqslant \frac{5}{6}\delta L + O(\ln L)\,,\end{aligned}$$

то $D(n,\varepsilon)/L < \frac{7}{8}\delta$ (мы учли, что δ мало, $L \geqslant (1/\delta)^2$ и $\varepsilon \leqslant \frac{1}{\sqrt{5}}$).

Наконец, если число $1/\delta$ не было натуральным, то можно провести всё рассуждение для числа $\delta_0 = \frac{1}{[1/\delta]} \geqslant \delta$. При этом $\frac{7}{8}\delta_0 < \delta$.

□

2.4. Об оптимальности оценок

Все оценки, доказанные в настоящей работе, представляются близкими к оптимальным, за исключением одного места, которое мы сейчас обсудим. Пусть δ и δ' пробегают действительные числа из $(0,1)$, L пробегает натуральные числа, и S пробегает множества двоичных последовательностей длины L. Обозначим через $\Phi_{L,\delta}(\delta')$ максимум по множествам S дефекта $\geqslant \delta$ наименьшего количества правильных множеств дефекта $\geqslant \delta'$, объединение которых покрывает S. По существу, в теоремах 4, 5 оценивается эта функция Φ. А именно, для некоторых функций ρ_1, ρ_2

$$\delta' \leqslant c_1 \cdot \frac{\delta}{\ln(1/\delta)} \quad \Rightarrow \quad \Phi_{L,\delta}(\delta') < L \cdot \rho_1(1/\delta)\,,$$
$$\delta' \geqslant c_2 \cdot \frac{\delta}{\ln(1/\delta)} \quad \Rightarrow \quad \Phi_{L,\delta}(\delta') > \mathrm{e}^{L/\rho_2(1/\delta)}\,,$$

где $c_1 \approx 1$, $c_2 = 2$ и $L \geqslant (1/\delta)^5$. То есть, когда δ' меняется от $c_1 \cdot \frac{\delta}{\ln(1/\delta)}$ до $c_2 \cdot \frac{\delta}{\ln(1/\delta)}$, функция Φ делает скачок от линейного по L выражения до экспоненциального по L выражения. Если нас интересует поведение Φ при фиксированном δ и $L \to \infty$, то оценка выглядит оптимальной (разве что можно сближать константы c_1 и c_2). Теперь рассмотрим поведение Φ при $\delta \to 0$. Тогда в полученных оценках ρ_2 растёт полиномиально, а ρ_1 растёт экспоненциально.

Открытая проблема: нельзя ли усилить верхнюю оценку на Φ, сделав ρ_1 полиномиальным?

Ан. Мучник доказал (см. [7]), что если уменьшить верхнюю оценку на δ', то в качестве ρ_1 годится линейная функция[11]. Более точно,

$$\delta' \leqslant c_3 \cdot \delta^2 \quad \Rightarrow \quad \Phi_{L,\delta}(\delta') < L \cdot (c_4/\delta)\,.$$

Поэтому положительное решение сформулированной проблемы кажется достаточно вероятным.

Благодарности

Авторы признательны Николаю Константиновичу Верещагину за интересные обсуждения вопросов колмогоровской теории, которые послужили для нас одним из стимулов начать настоящую работу. Большую помощь оказал Алексей Вячеславович Чернов при подготовке текста к

[11]Алгоритмический аналог этого результата недавно получили Н. Верещагин и Б. Дюран, см. [8]. Некоторые элементы их рассуждения мы использовали в доказательстве теоремы 4′.

публикации, за что авторы ему очень благодарны. Основное содержание статьи было доложено на Колмогоровском семинаре Московского государственного университета весной 2002 года. Мы признательны его участникам за внимание.

Литература

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

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

[3] А. Н. Колмогоров. О таблицах случайных чисел. *Семиотика и информатика*, вып. 18 (1982), М.: ВИНИТИ, с. 3–13. (Перепечатано в сборнике: А. Н. Колмогоров. Теория информации и теория алгоритмов. М.: Наука, 1987, с. 204–213.)

[4] А. М. Яглом, И. М. Яглом. Вероятность и информация. 2-е издание. М.: Физматгиз, 1960.

[5] M. van Lambalgen. Von Mises' definition of random sequences reconsidered. *Journal of Symbolic Logic*, v. 52 (1987), pp. 725–755.

[6] V. A. Uspensky, A. Kh. Shen. Relations between varieties of Kolmogorov complexities. *Math. Systems Theory*, 29(1996), pp. 271–292.

[7] An. A. Muchnik, A. L. Semenov, V. A. Uspensky. Mathematical metaphysics of randomness. *Theoretical Computer Science*, v. 207 (1998), pp. 263–317.

[8] B. Durand, N. Vereshchagin. Kolmogorov–Loveland stochasticity for finite strings. `http://markov.math.msu.ru/~ver/papers/kolm-love.ps` 2002.