ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 43 2007 Вып. 1

УДК 621.391.1.:519.2

© 2007 г. К. Ю. Горбунов

ОЦЕНКА ЧИСЛА ЭЛЕМЕНТОВ ПОКРЫТИЯ ПРОИЗВОЛЬНОГО ТЕСТА НА СЛУЧАЙНОСТЬ ЧАСТОТНЫМИ ТЕСТАМИ[1]

Улучшается известная асимптотическая оценка на число монотонных правил выбора при покрытии произвольного теста на случайность частотными тестами. Точнее, доказано, что для любого множества S (произвольный тест) двоичных последовательностей достаточно большой длины L, где $|S| \le 2^{L(1-\delta)}$, при достаточно малом δ существует полиномиальное от $1/\delta$ множество монотонных правил выбора подпоследовательности (частотные тесты), обеспечивающих выбор из любой последовательности $\boldsymbol{t} \in S$ подпоследовательности, у которой произведение ее длины на квадрат отклонения доли нулей в ней от $1/2$ имеет порядок не меньше $0{,}5 \ln 2\, L[\delta/\ln(1/\delta)](1 - 2\ln\ln(1/\delta)/\ln(1/\delta))$.

В данной статье дается ответ на вопрос, поставленный в работе [1, п. 2.4]. Напомним основные определения и постановки задач из [1], связанные с рассматриваемой проблемой. Сама проблема будет сформулирована ниже.

Произвольным тестом на случайность для двоичных последовательностей длины L считается любое множество S таких последовательностей. *Удельным дефектом* $\delta(S)$ *множества* S называется величина $(L - \operatorname{lb}|S|)/L$, где lb – двоичный логарифм, а $|S|$ – мощность множества S. Пусть $\boldsymbol{t}$ обозначает произвольный элемент множества S. *Монотонным правилом выбора* называется правило выбора (т.е. функция, определенная на всех конечных последовательностях и принимающая значение 0 или 1) подпоследовательности из $\boldsymbol{t}$, которое на каждом i-м шаге ($i = 1, \ldots, L$) решает (используя лишь начало последовательности $\boldsymbol{t}$ длины $i-1$), добавлять ли i-ю букву t_i слова $\boldsymbol{t}$ в конец строящейся подпоследовательности. *Немонотонное правило* выбора отличается от монотонного тем, что может просматривать биты из $\boldsymbol{t}$ в произвольном порядке (решение о добавлении бита в конец строящейся подпоследовательности принимается непосредственно до его просмотра на основании значений уже просмотренных битов). *Удельным дефектом правила относительно* S называется произведение числа $\dfrac{2\operatorname{lb} e}{L}$ на минимум (по всем $\boldsymbol{t} \in S$) произведения длины выбранной подпоследовательности на квадрат отклонения доли нулей в ней от $1/2$ (здесь e – основание натурального логарифма).

Кратко напомним обоснование этого определения (подробнее см. в [1]). *Нормальным правилом* называется правило, всегда выбирающее из последовательности длины L подпоследовательность одной и той же длины. Из любого правила r легко изготовить L нормальных правил $r_1, \ldots, r_L$, где каждое r_i выбирает подпоследовательность "своей" длины i и не меняет выбор правила r в тех случаях, когда оно

[1] Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 06-01-00122) и Совета поддержки научных школ при Президенте РФ (номер проекта НШ-358.2003.1).

тоже выбирало подпоследовательность длины i. Каждому же нормальному правилу r_i и заданному отклонению ε соответствует множество T (*частотный тест*) тех последовательностей длины L, из которых r_i выбирает подпоследовательность (длины i) с отклонением доли нулей от $1/2$ не меньше ε. Применяя оценку для вероятностей больших уклонений (см. [2, с. 93], иногда эту оценку называют оценкой Чернова), нетрудно показать, что удельный дефект множества T не меньше $(2n\varepsilon^2 \operatorname{lb} e)/L$, а если ε достаточно мало и n достаточно велико по сравнению с $1/\varepsilon$, то $\delta(T) \approx (2n\varepsilon^2 \operatorname{lb} e)/L$.

Скажем, что семейство R правил *δ'-покрывает* S, если существует разбиение множества S на $|R|$ подмножеств и такое взаимно однозначное соответствие между правилами и подмножествами, что удельный дефект каждого правила относительно соответствующего ему подмножества больше δ'. Отметим, что если правила нормальные, то данное определение имеет ясный смысл: произвольный тест S покрывается набором из $|R|$ частотных тестов, а число δ' отражает "качество" этих частотных тестов. Прежде чем формулировать наш основной результат, рассмотрим несколько примеров и кратко опишем известные факты и открытые проблемы.

Пример 1. Пусть L четно и все позиции разбиты на пары соседних. Рассмотрим множество S, характеризующееся тем, что число пар из двух нулей не больше $0{,}5L(0{,}25-\varepsilon)$, т.е. доля таких пар отклоняется от "положенной" одной четверти на ε, где $\varepsilon < 0{,}1$. Очевидно, что удельный дефект S больше некоторого положительного числа, зависящего от ε, но не от L. Рассмотрим два монотонных правила: одно выбирает в подпоследовательность все первые биты каждой пары, а другое смотрит первый бит каждой пары, и если он равен 0, то берет в подпоследовательность второй бит. Легко видеть, что либо первое правило достигнет отклонения доли нулей в выбранной подпоследовательности от $1/2$ не менее ε, либо того же достигнет второе, причем в последнем случае длина подпоследовательности будет не менее $L(0{,}5-\varepsilon)$. Следовательно, удельный дефект хотя бы одного из правил будет не менее $(0{,}8 \operatorname{lb} e)\varepsilon^2$, т.е. дефект зависит лишь от δ, но не от L. Легко понять, что такого же эффекта можно достичь одним немонотонным правилом, которое сначала выбирает все первые биты, а затем либо кончает работу (если в выбранной таким образом подпоследовательности отклонение доли нулей от $1/2$ достаточно велико), либо выбирает вторые биты там, где первые равны 0.

Пример 2. Пусть все L позиций разбиты на блоки четной длины d (последний блок может иметь длину меньше d, но при фиксированном d и больших L он не влияет на результат). Рассмотрим множество S, характеризующееся тем, что в каждом блоке число нулей равно числу единиц. Легко понять, что удельный дефект множества S зависит лишь от d, но не от L (он имеет порядок $0{,}5(\ln d)/d$). Рассмотрим два монотонных правила, одно из которых берет в подпоследовательность последний бит блока тогда, когда он равен 0, а другое – когда он равен 1 (понятно, что просмотрев все биты блока, кроме последнего, правило "понимает", чему равен последний бит). Очевидно, что одно из правил достигнет удельного дефекта примерно $(\operatorname{lb} e)/d$. Легко понять, что такого же эффекта можно достичь одним немонотонным правилом, которое сначала выясняет, чему равен последний бит каждого блока, а затем выбирает те биты, которых большинство.

Положение, когда одного немонотонного правила "хватает" там, где требуется более одного монотонного, является достаточно типичным. В частности, автору не известно ни одного случая, когда одного немонотонного правила "не хватает". Точнее говоря, существует ли некоторое $\delta > 0$, такое что для любого ε и любого сколь угодно большого L существует множество S с удельным дефектом не меньше δ, для которого не существует никакого немонотонного правила, имеющего удельный дефект относительно S не меньше ε. Известно лишь соответствующее утверждение для случая, когда мы требуем, чтобы ε было достаточно велико по сравнению с δ.

Точнее, в [1, теорема 5] показано, что если потребовать, чтобы ε было не менее $\dfrac{2\delta}{\ln \delta}$, то существует такое S, для которого не хватит даже некоторого экспоненциального от L количества немонотонных правил.

Для монотонных правил положение с нижними оценками (т.е. доказательством того, что правил "не хватает") обстоит несколько лучше. Так, легко в духе известного примера Вилля (см. [3, п. 6.2.2]) следующим образом построить пример, когда "не хватает" одного монотонного правила. Пусть, как в примере 1, все позиции разбиты на пары. Рассмотрим множество S, характеризующееся тем, что нет пар из двух нулей. Тогда по любому монотонному правилу строится следующая последовательность: каждый раз, когда правило не берет очередной бит в подпоследовательность, полагаем его равным 1, а на тех битах, которые берутся в подпоследовательность, чередуем значения 0 и 1. Очевидно, что построенная таким образом последовательность принадлежит S и что правило не достигает на ней никакого существенного удельного дефекта.

Эту нижнюю оценку можно усилить и показать, что бывают случаи, когда "не хватает" порядка $\ln(1/\delta)$ монотонных правил. Пусть, как и в примере 2, все L позиций разбиты на блоки равной длины d, где d равно 2^{2r}, а r – число правил, которые мы хотим "обмануть". Рассмотрим множество S, характеризующееся тем, что в каждом блоке число нулей отличается от числа единиц не более чем на 2^r. Из центральной предельной теоремы следует, что удельный дефект δ множества S равен c/d для некоторой константы c, следовательно, $r = 0{,}5(\operatorname{lb}(1/\delta) + \operatorname{lb} c)$. Пусть имеется r монотонных правил. Построим по ним следующую последовательность $\boldsymbol{t}$. Пусть i – номер очередного бита и M_i – множество правил, которые решают брать этот бит в подпоследовательность. Если множество M_i не встречалось раньше (т.е. $M_i \neq M_j$ ни для какого $j < i$), то ставим на i-ю позицию 1, иначе рассмотрим наибольшее j, при котором $M_i = M_j$, и поставим на i-ю позицию бит, не равный поставленному на j-ю позицию (т.е. биты, соответствующие одному множеству правил, чередуем). Очевидно, что $\boldsymbol{t} \in S$. С другой стороны, любое правило, решающее брать очередной бит, находится "в компании" одного из 2^{r-1} множеств правил, так что в выбранной подпоследовательности число нулей не более чем на 2^{r-1} отличается от числа единиц. Очевидно, что ее удельный дефект стремится к нулю при $L \to \infty$.

Известные верхние оценки также относятся к монотонным правилам (автору не известно, возможно ли их улучшение за счет допустимости немонотонных правил). Одна из первых нетривиальных верхних оценок получена Ан.А. Мучником и состоит в следующем (поскольку подробное доказательство этой теоремы не публиковалось (см. [1, с. 164; 4, п. 9.2.1]), приведем его в конце статьи).

Теорема 1. *Для произвольного множества S с удельным дефектом не менее δ существует $c_1 \dfrac{1}{\delta}$ монотонных правил, $c_2\delta^2$-покрывающих S (здесь c_1 и c_2 – некоторые константы).*

Перєйдем к изложению нашего основного результата. В [1, теорема 4] доказано, что для любого достаточно малого $\delta > 0$, любого $L \geq (1/\delta)^5$ и любого множества S последовательностей длины L с удельным дефектом не менее δ существует семейство R монотонных правил, δ'-покрывающее S, где $\delta' = \dfrac{\delta}{\ln(1/\delta)}(1-\beta)$, и в качестве β можно взять $\dfrac{2\ln\ln(1/\delta)}{\ln(1/\delta)}$. Мощность семейства R имеет порядок экспоненты от $1/\delta$ (мы имеем в виду не обязательно нормальные правила). Следующая теорема улучшает оценку на мощность семейства R. Если $\delta' = \dfrac{\delta}{\ln(1/\delta)}(1-\beta)$, то при фикси-

рованном β эта мощность становится полиномиальной от $1/\delta$, а при $\beta = \dfrac{2\ln\ln(1/\delta)}{\ln(1/\delta)}$ она субэкспоненциальна от $1/\delta$.

Теорема 2. *Пусть* $\delta \in (0; \exp_e(-e^{50}))$, *и фиксировано натуральное* $L \geq (1/\delta)^5$. *Будем рассматривать множества двоичных последовательностей длины* L. *Для произвольного множества* S *с удельным дефектом не менее* δ *существует семейство не более чем из*

$$\left(\frac{1}{\delta}\right)^{\frac{0{,}6\ln 2\ln(1/\delta)}{\beta\ln(1/\delta)-1{,}3\ln\ln(1/\delta)}+6{,}8}$$

монотонных правил, δ'*-покрывающее* S, *где* $\delta' = \dfrac{\delta}{\ln(1/\delta)}(1-\beta)$. *В качестве* β *можно взять* $\dfrac{2\ln\ln(1/\delta)}{\ln(1/\delta)}$.

Доказательство. В [1, начало доказательства теоремы 4] рассматривается игра, где Математик и Природа по очереди делают L ходов: на i-м ходу Математик указывает ставку $x_i \in [0; 1]$ на 0 или 1, а Природа указывает элемент $t_i \in \{0, 1\}$, причем так, чтобы после L-го хода построенная последовательность $\boldsymbol{t}$ принадлежала S. Вначале 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 противном случае (капитал может быть и отрицательным). Там же доказывается, что для любого множества S существует стратегия Математика, позволяющая выиграть не менее $L\delta(S)\ln 2$. Величина ставки на i-м ходу равна $2\dfrac{|S_1(i)|}{|S_1(i)|+|S_2(i)|} - 1$, где $S_1(i)$ и $S_2(i)$ – два множества возможных продолжений известного (перед i-м ходом) начала последовательности $\boldsymbol{t}$: одно с $t_i = 0$, другое с $t_i = 1$, причем $|S_1(i)| \geq |S_2(i)|$ (Математик делает ставку на бит, соответствующий $S_1(i)$). Далее, говоря об игре, мы всегда будем предполагать, что ставки рассчитываются по этому правилу.

В дальнейшем мы будем говорить о вероятностях, имея в виду вероятности событий, порожденных равномерным распределением искомого элемента $\boldsymbol{t}$ по множеству S. Оценивать эти вероятности нам поможет следующий случайный процесс, соответствующий упомянутой игре. А именно, играя с Природой, Математик на i-м ходу выигрывает с вероятностью $\dfrac{|S_1(i)|}{|S_1(i)|+|S_2(i)|}$ и проигрывает, соответственно, с вероятностью $\dfrac{|S_2(i)|}{|S_1(i)|+|S_2(i)|}$. Очевидно, что любая последовательность из S с равной вероятностью может оказаться построенной в конце игры последовательностью $\boldsymbol{t}$.

Лемма 1. *Пусть* M – *подмножество отрезка* $[0; 1]$, a *и* b – *нижняя и верхняя его грани соответственно. Предположим, что* n – *число ставок, попавших в* M, d – *разность числа выигрышей и проигрышей (из рассматриваемых* n *ходов). Тогда для любого* $\varepsilon \in (0; 0{,}1]$, *если* $L > 20/\varepsilon^4$, $n \geq L\varepsilon$, *то вероятность того, что* $d < n(a-\varepsilon)$, *не превышает* $\exp_e(-0{,}4L\varepsilon^3)$. *Такова же и оценка на вероятность того, что* $d > n(b+\varepsilon)$.

Доказательство. Рассмотрим бинарное дерево T, соответствующее описанному случайному процессу (будем представлять его "растущим вверх"). Его вершины (кроме листьев) пометим размерами ставок в них, однозначно определяющими вероятности сдвигов по ребрам. Вершины, где ставки принадлежат M, будем называть *активными*. *Активной высотой* вершины v назовем число активных вершин на пу-

ти из корня в v без учета самой v. Чтобы оценить вероятность первого события из формулировки леммы, рассмотрим для произвольного фиксированного n событие $A(n,\varepsilon)$, состоящее в том, что на пути из корня в лист встретилось ровно n активных вершин, и $d < n(a-\varepsilon)$. Изменим дерево T следующим образом. Каждый его лист, имеющий активную высоту меньше n, формально дополним сверху поддеревом так, чтобы для новых листьев их активная высота равнялась n. Каждую же вершину, у которой активная высота равна n, сделаем листом, удалив все, что было выше. Очевидно, что для нового дерева T' вероятность события $A(n,\varepsilon)$ не уменьшилась.

Покажем, что вероятность этого события не уменьшится и при следующем преобразовании: ставки во всех активных вершинах положим равными a (эти вершины продолжаем считать активными, даже если $a \notin M$). Обозначим через $p(k,m)$ вероятность набрать не более k выигрышей за m ходов, если вероятность выигрыша на каждом ходу равна $(a+1)/2$. Очевидно, достаточно доказать следующее утверждение: для любого k вероятность набрать не более k выигрышей при движении в лист из вершины v, имеющей активную высоту h, до преобразования была не больше $p(k,n-h)$, а после преобразования стала равна $p(k,n-h)$. Доказываем его индукцией по убыванию расстояния вершины v от корня. Индуктивный переход очевиден, так как преобразование не увеличивает вероятность выигрыша.

Поскольку в получившемся дереве число выигрышей имеет бернуллиевское распределение с вероятностью успеха $(a+1)/2$, можно применить оценку вероятности больших уклонений. Получим, что вероятность события $A(n,\varepsilon)$ не превышает $\exp_e(-2n(\varepsilon/2)^2) = \exp_e(-0{,}5n\varepsilon^2)$. Суммируя по всем n от $L\varepsilon$ до L, получаем требуемую оценку на вероятность первого события в формулировке леммы. Оценка на вероятность второго события доказывается аналогично. Отметим, что другой путь доказательства подобных оценок предложен в [1, предложение 1]. Лемма 1 доказана.

Капиталом $K(M)$, набранным на M, назовем капитал, который был бы набран в описанной игре, если бы учитывались лишь ставки, попадающие в M.

Лемма 2. *Пусть M, a, b, n, d обозначают то же, что и в формулировке леммы* 1. *Тогда для любого $\varepsilon \in (0;\,0{,}1]$, если $n \geq L\varepsilon$, $L > 20/\varepsilon^8$ и $a \geq \varepsilon$, то вероятность того, что $K(M) > db + 3L\varepsilon$, не превышает* $\exp_e(-0{,}3L\varepsilon^6)$.

Доказательство. Разобьем отрезок $[a;b]$ на части: полуинтервалы $[a_1;b_1)$, $[a_2;b_2)$, …, $[a_{m-1};b_{m-1})$ длины ε и отрезок $[a_m;b_m]$ длины не более ε, где $a_1 = a$, $a_2 = b_1$, …, $a_m = b_{m-1}$, $b_m = b$, $m \leq 1/\varepsilon$. Под ставками будем иметь в виду лишь ставки из M. Те части, куда попало меньше $L\varepsilon^2$ ставок, исключим из рассмотрения (заметим, что суммарное число ставок на них меньше $L\varepsilon$, а суммарный капитал не превышает числа ставок). Остальные части занумеруем слева направо, и будем считать обозначения a_i, b_i, n_i и d_i, относящимися к i-й части.

По лемме 1 с вероятностью не меньшей, чем $1 - \exp_e(-0{,}4L\varepsilon^6)$, каждое d_i неотрицательно, а значит, с вероятностью не меньшей, чем $1 - \exp_e(-0{,}3L\varepsilon^6)$, все d_i неотрицательны. Рассматривая наихудший для Математика случай (т.е. когда все выигрыши приходятся на ставку b_i, а все проигрыши – на ставку a_i), заключаем, что с той же вероятностью капитал, набранный на i-й части, не превосходит $b_i\left(d_i + \dfrac{n_i - d_i}{2}\right) - a_i\dfrac{n_i - d_i}{2} \leq d_i b_i + n_i\varepsilon$. Суммируя по i и вспоминая про исключенные части, получаем, что с упомянутой вероятностью

$$K(M) \leq b\sum_i d_i + \varepsilon\sum_i n_i + L\varepsilon \leq b(d + L\varepsilon) + L\varepsilon + L\varepsilon \leq bd + 3L\varepsilon.$$

Лемма 2 доказана.

Напомним, что гарантированный капитал на $[0;1]$ равен $L\delta\ln 2$. Поскольку правило выбора стремится достичь большого отклонения доли нулей от $1/2$, мы (имея в виду возможность преобразования стратегии в монотонное правило) расщепляем

стратегию на две: одна делает ставки только на $t_i = 0$, другая – только на $t_i = 1$. По крайней мере одна из этих стратегий гарантирует на $[0;1]$ капитал $(L\delta \ln 2)/2$. Обозначим ее через R.

Лемма 3. *Пусть Математик использует стратегию R,*

$$\alpha = \sqrt{0{,}5 \ln 2 (1 - \delta^{0{,}05}) \frac{\delta}{\ln(1/\delta)}}.$$

Тогда вероятность набрать на отрезке $[0;\alpha]$ капитал больший, чем $\dfrac{L\delta \ln 2}{2\ln(1/\delta)}$, не превосходит $\exp_e(-L\delta^{6{,}7})$.

Доказательство. Положим $K = \dfrac{L\delta \ln 2}{2\ln(1/\delta)}$. Очевидно, что на отрезке $[0;\delta^{1,1}]$ можно набрать капитал не больше $L\delta^{1,1}$, поэтому остается оценить вероятность набрать капитал больший, чем $K - L\delta^{1,1}$, на отрезке $\Delta = [\delta^{1,1};\alpha]$, к которому будем относить наши обычные обозначения n и d. Поскольку такой капитал можно набрать лишь при $n > L\delta^{1,1}$, то по лемме 2 (где $\varepsilon = \delta^{1,1}$) получаем, что вероятность того, что набранный капитал больше $d\alpha + 3L\delta^{1,1}$, не превышает $\exp_e(-0{,}3L\delta^{6,6})$. Кроме того, по лемме 1 (где $\varepsilon = \delta^{1,1}$) с вероятностью не меньшей, чем $1 - \exp_e(-0{,}4L\delta^{3,3})$, выполняется неравенство $d \le n(\alpha + \delta^{1,1})$. Таким образом, с вероятностью не меньшей, чем $1 - \exp_e(-L\delta^{6,7})$, набранный на Δ капитал не превосходит $n(\alpha + \delta^{1,1})\alpha + 3L\delta^{1,1}$. Легко видеть, что это меньше $K - L\delta^{1,1}$ при малых δ. Лемма 3 доказана.

Доказанная лемма позволяет нам сосредоточиться на ставках, не меньших α. В частности, лемма 3 с учетом леммы 1 гарантирует, что наши правила выбора будут достаточно эффективны, по крайней мере, с точки зрения отклонения доли нулей от $1/2$ в построенных подпоследовательностях: при длине не меньше $0{,}5L\delta \ln 2 \left(1 - \dfrac{1}{\ln(1/\delta)}\right)$ это отклонение с большой вероятностью будет не намного меньше $(1+\alpha)/2$.

Разобьем отрезок $[\alpha;1]$ на части: полуинтервалы $[a_1;b_1)$, $[a_2;b_2)$, $\ldots$, $[a_{m-1};b_{m-1})$ и отрезок $[a_m;b_m]$, где $a_1 = \alpha$, $a_2 = b_1$, $\ldots$, $a_m = b_{m-1}$, $b_m = 1$ и $b_i/a_i = k$ для всех $i = 1,\ldots,m-1$, $b_m/a_m \le k$,

$$k = \frac{1}{1-\beta}\left(1 - \frac{1{,}2\ln\ln(1/\delta)}{\ln(1/\delta)}\right).$$

Отметим, что $k > 1$ при $\beta = \dfrac{2\ln\ln(1/\delta)}{\ln(1/\delta)}$. Найдем верхнюю оценку на число частей m. Имеем $\alpha k^{m-1} \le 1/k$, откуда

$$m \le 1 + \frac{\ln(1/\alpha)}{\ln k} \le 1 + \frac{0{,}5(\ln(2\,\mathrm{lb}\,e) - \ln(1-\delta^{0{,}05}) + \ln\ln(1/\delta) + \ln(1/\delta))}{\ln k} \le$$
$$\le \frac{0{,}6\ln(1/\delta)}{\ln k}.$$

Как обычно, относим обозначения n_i и d_i к i-й части. Набранный на ней капитал обозначаем через K_i. Любое подмножество M множества частей определяет правило выбора: отбирать очередную букву слова $\boldsymbol{t}$ тогда и только тогда, когда стратегия R делает ставку из части, принадлежащей M. Согласно определению удельный дефект

этого правила относительно S равен

$$\frac{2\operatorname{lb} e}{L}\left(\frac{\sum d_i}{2\sum n_i}\right)^2\sum n_i=\frac{\operatorname{lb} e}{2L}\frac{(\sum d_i)^2}{\sum n_i},$$

где все суммы берутся по $i\in M$. Таким образом, нам надо доказать, что с большой вероятностью существует M, у которого $\dfrac{(\sum d_i)^2}{\sum n_i}$ не меньше $Z=2L\ln 2(1-\beta)\dfrac{\delta}{\ln(1/\delta)}$. Будем составлять M лишь из таких частей, где $n_i\geq L\delta^{1,1}$, поскольку на других частях суммарный капитал меньше $Lm\delta^{1,1}$ и мы исключаем их из рассмотрения. Множество оставшихся частей обозначим через M^*. Следующая лемма устанавливает нижнюю оценку на сумму $\sum\limits_i \dfrac{d_i^2}{n_i}$, где i "пробегает" по *всем* частям из M^*.

Лемма 4. *С вероятностью не менее* $1-\exp_e(-L\delta^{6,8})$ *выполняется неравенство*

$$\sum_{i\in M^*}\frac{d_i^2}{n_i}>\frac{L\delta\ln 2}{2k}\left(1-\frac{2}{\ln(1/\delta)}\right).$$

Доказательство. По лемме 1 (где $\varepsilon=\delta^{1,1}$) для $i\in M^*$ с вероятностью не меньшей, чем $1-\exp_e(-0{,}4L\delta^{3,3})$, выполняется неравенство $d_i\geq n_i(a_i-\delta^{1,1})$. По лемме 2 (где $\varepsilon=\delta^{1,1}$) с вероятностью не меньшей, чем $1-\exp_e(-0{,}3L\delta^{6,6})$, выполняется неравенство $K_i\leq d_ib_i+3L\delta^{1,1}$, т.е. $d_ib_i\geq K_i-3L\delta^{1,1}$. Следовательно, с вероятностью не меньше $1-\exp_e(-L\delta^{6,7})$ имеем

$$\sum_{i\in M^*}\frac{d_i^2}{n_i}\geq\sum_{i\in M^*}d_i\frac{n_i(a_i-\delta^{1,1})}{n_i}=\sum_{i\in M^*}d_ia_i-\delta^{1,1}\sum_{i\in M^*}d_i\geq$$
$$\geq\sum_{i\in M^*}\frac{d_ib_i}{k}-L\delta^{1,1}\geq\frac{1}{k}\left(\sum_{i\in M^*}K_i-3Lm\delta^{1,1}\right)-L\delta^{1,1}.$$

В соответствии с леммой 3 и определением множества M^* с вероятностью не меньшей, чем $1-\exp_e(-L\delta^{6,7})$, выполняется неравенство

$$\sum_{i\in M^*}K_i\geq\frac{L\delta\ln 2}{2}\left(1-\frac{1}{\ln(1/\delta)}\right)-Lm\delta^{1,1}.$$

Учитывая соотношение между β и δ, оценим снизу $\ln k$. Имеем

$$\ln k=\ln\frac{1}{1-\beta}+\ln\left(1-\frac{1{,}2\ln\ln(1/\delta)}{\ln(1/\delta)}\right)\geq\beta-\frac{1{,}3\ln\ln(1/\delta)}{\ln(1/\delta)}\geq\frac{0{,}7\ln\ln(1/\delta)}{\ln(1/\delta)}.$$

Отсюда получаем, что с вероятностью не меньшей, чем $1-\exp_e(-L\delta^{6,8})$, при малых δ имеем

$$\sum_{i\in M^*}\frac{d_i^2}{n_i}\geq\frac{L\delta\ln 2}{2k}\left(1-\frac{1}{\ln(1/\delta)}\right)-\frac{Lm\delta^{1,1}}{k}-\frac{3Lm\delta^{1,1}}{k}-L\delta^{1,1}\geq$$
$$\geq\frac{L\delta\ln 2}{2k}\left(1-\frac{1}{\ln(1/\delta)}\right)-\frac{4L\delta^{1,1}\ln^2(1/\delta)}{\ln\ln(1/\delta)}-L\delta^{1,1}>\frac{L\delta\ln 2}{2k}\left(1-\frac{2}{\ln(1/\delta)}\right).$$

Лемма 4 доказана.

Лемма 5. *С вероятностью не меньшей, чем* $1 - \exp_e(-L\delta^{6,8})$, *существует такое подмножество* M *множества* M^*, *что*

$$\frac{\left(\sum\limits_{i\in M} d_i\right)^2}{\sum\limits_{i\in M} n_i} \geq Z.$$

Доказательство. По лемме 4 с вероятностью не меньшей, чем $1-\exp_e(-L\delta^{6,8})$, выполняется неравенство

$$\sum_{i\in M^*} \frac{d_i^2}{n_i} > \frac{L\delta \ln 2}{2k}\left(1 - \frac{2}{\ln(1/\delta)}\right).$$

Пусть требуемого M не существует. Тогда для любого $J \subseteq M^*$ выполняется неравенство

$$\left(\sum_{i\in J} \frac{d_i}{n_i} n_i\right)^2 \leq Z\left(\sum_{i\in J} n_i\right).$$

Повторяя рассуждение из доказательства предложения 2 работы [1], получаем неравенство

$$\sum_{i\in M^*} \frac{d_i^2}{n_i} < Z + \frac{Z}{4}\left(\ln\left(\sum_{i\in M^*} n_i\right) - \ln Z\right),$$

доказательство которого проходит дословно, незначительное отличие возникает лишь при обосновании неравенства $\sum\limits_{i\in M^*} n_i > Z$, которое в нашем случае выглядит следующим образом:

$$\sum n_i = \sum \frac{n_i^2}{n_i} \geq \sum \frac{d_i^2}{n_i} > \frac{L\delta \ln 2}{2k}\left(1 - \frac{2}{\ln(1/\delta)}\right) > Z.$$

Следовательно,

$$\frac{L\delta \ln 2}{2k}\left(1 - \frac{2}{\ln(1/\delta)}\right) < (1-\beta)\frac{2L\delta \ln 2}{\ln(1/\delta)}\left(1 + \frac{1}{4}\ln\frac{L}{Z}\right),$$

т.е.

$$\frac{1}{k}\left(1 - \frac{2}{\ln(1/\delta)}\right) < (1-\beta)\frac{4 + \ln\ln(1/\delta) - \ln 2 - \ln\ln 2 + \ln(1/\delta) - \ln(1-\beta)}{\ln(1/\delta)},$$

что неверно при

$$k \leq \frac{\ln(1/\delta) - 2}{(1-\beta)(4 + \ln\ln(1/\delta) - \ln\ln 4 + \ln(1/\delta) - \ln(1-\beta))},$$

в частности, при

$$k(1-\beta) \leq \frac{\ln(1/\delta) - 2}{\ln(1/\delta) + 1{,}1\ln\ln(1/\delta)}.$$

Непосредственной подстановкой легко проверить, что при наших β и k (если $\delta \leq \leq \exp_e(-e^{50})$) это неравенство выполняется. Лемма 5 доказана.

Доказательство теоремы 2 проводится по той же схеме, что и в [1]. А именно, то малое подмножество множества S (называемое *исключительным*), на котором не действует ни одно из описанных правил, объявляется новым S, и вся конструкция итерируется на нем заново. Возникает еще меньшее исключительное множество, и т.д. Осталось оценить число правил. Пользуясь полученной ранее оценкой на $\ln k$, имеем

$$\operatorname{lb} k = \operatorname{lb} e \ln k \geq \operatorname{lb} e \left(\beta - \frac{1{,}3\ln\ln(1/\delta)}{\ln(1/\delta)}\right) = \frac{\beta\ln(1/\delta) - 1{,}3\ln\ln(1/\delta)}{\ln 2 \ln(1/\delta)}.$$

Поэтому число правил на каждой итерации не превышает

$$2^m \leq \exp_2\left(\frac{0{,}6\operatorname{lb}(1/\delta)}{\operatorname{lb} k}\right) = \left(\frac{1}{\delta}\right)^{\frac{0{,}6\ln 2\ln(1/\delta)}{\beta\ln(1/\delta) - 1{,}3\ln\ln(1/\delta)}}.$$

Из леммы 5 следует, что количество самих итераций не больше $(1/\delta)^{6{,}8}$. Теорема 2 доказана.

Особый интерес могут представлять правила выбора, у которых множество учитываемых ставок связно и просто описываемо. Поэтому может быть интересен вопрос о соответствующих оценках на удельный дефект и число правил, у которых это множество состоит из одного отрезка (назовем такие правила *отрезочными относительно* S). Результат Ан.А. Мучника, сформулированный в теореме 1, в действительности утверждает, что при δ' порядка δ^2 для δ'-покрытия множества S (где $\delta(S) = \delta$) достаточно линейного от $1/\delta$ числа монотонных отрезочных относительно S правил (точнее, пороговых правил, т.е. таких, у которых отрезок учитываемых ставок имеет вид $[p; 1]$).

Наша методика позволяет получить оценку для отрезочных правил, которые покрывают хотя и не все множество S, но почти все. Точнее, верна следующая

Теорема 3. *Пусть фиксировано число $\delta \in (0; \exp_e(-e^{50}))$ и натуральное число $L \geq (1/\delta)^5$. Будем рассматривать множества двоичных последовательностей длины L. Для произвольного множества S с удельным дефектом не менее δ существует семейство из не более $0{,}6\ln(1/\delta)$ монотонных отрезочных относительно S правил, δ'-покрывающее почти все S, кроме, быть может, подмножества мощности не более $|S|\exp_e(-L\delta^{6{,}8})$, где $\delta' = \dfrac{1}{2{,}5e}\dfrac{\delta}{\ln(1/\delta)}$.*

Доказательство. Строим множество частей так же, как в доказательстве теоремы 2, но с $k = e$. Будем рассматривать лишь правила, для которых множество ставок – некоторый отрезок вида $[a_i; b_i]$. Из леммы 3 следует, что с вероятностью не меньшей, чем $1 - \exp_e(-L\delta^{6{,}7})$, на некотором таком отрезке наберется капитал $K_i \geq \dfrac{L\delta\ln 2}{2m}\left(1 - \dfrac{1}{\ln(1/\delta)}\right)$. Соответствующее ему правило имеет удельный дефект $D = \dfrac{\operatorname{lb} e}{2L}\dfrac{d_i^2}{n_i}$. По лемме 1 ($\varepsilon = \delta^{1,1}$) с вероятностью не меньшей, чем $1 - \exp_e(-0{,}4L\delta^{3,3})$, имеем $d_i \geq n_i(a_i - \delta^{1,1})$. Следовательно, с той же вероятностью $D \geq \dfrac{\operatorname{lb} e}{2L} n_i(a_i - \delta^{1,1})^2$. С другой стороны, по лемме 2 ($\varepsilon = \delta^{1,1}$) с вероятностью не меньшей, чем $1 - \exp_e(-0{,}3L\delta^{6,6})$, имеем $K_i \leq d_i b_i + 3L\delta^{1,1}$, т.е. $d_i \geq \dfrac{K_i - 3L\delta^{1,1}}{b_i}$. Следовательно, с той же вероятностью $D \geq \dfrac{\operatorname{lb} e}{2L}\dfrac{(K_i - 3L\delta^{1,1})^2}{b_i^2 n_i}$. Поскольку первая оценка на D возрастает с ростом n_i, а вторая убывает, то наименьшее возможное зна-

чение D достигается, если $n_i(a_i - \delta^{1,1})^2 = \dfrac{(K_i - 3L\delta^{1,1})^2}{b_i^2 n_i}$, т.е. при $n_i = \dfrac{K_i - 3L\delta^{1,1}}{b_i(a_i - \delta^{1,1})}$. При таком n_i с учетом оценок на K_i, m и неравенств $b_i/a_i \le k$, $b_i > \alpha$, получаем, что с вероятностью не меньшей, чем $1 - \exp_e(-L\delta^{6,8})$, для малых δ выполняется неравенство

$$D \ge \frac{\mathrm{lb}\, e}{2L} \frac{(a_i - \delta^{1,1})(K_i - 3L\delta^{1,1})}{b_i} \ge \frac{\mathrm{lb}\, e(a_i - \delta^{1,1})}{2Lb_i} \frac{L\delta \ln 2}{1{,}2 \ln(1/\delta)} \left(1 - \frac{2}{\ln(1/\delta)}\right) \ge$$

$$\ge \frac{(1/e) - \sqrt{\delta}}{2{,}4} \frac{\delta}{\ln(1/\delta)} \left(1 - \frac{2}{\ln(1/\delta)}\right) \ge \frac{1}{2{,}5e} \frac{\delta}{\ln(1/\delta)}.$$

Число правил не превышает $m \le 0{,}6 \ln(1/\delta)$. Теорема 3 доказана.

Замечание. Легко видеть, что за счет уменьшения области возможных значений δ константу 0,6 в формулировке теоремы 2 можно сделать сколь угодно близкой к 0,5, константу 6,8 в формулировках теорем 2 и 3 – сколь угодно близкой к 6, а константу 2,5 в формулировке теоремы 3 – сколь угодно близкой к 2.

Приведем теперь

Доказательство теоремы 1. Как уже упоминалось, все наши правила будут пороговыми, т.е. решение о включении очередного бита в подпоследовательность принимается, если ставка (рассчитываемая по правилу, описанному в начале доказательства теоремы 2) не меньше некоторого порога. Назовем упорядоченную пару из двух чисел $\langle r_1, r_2 \rangle$ хорошей, если $r_2 - r_1 \le \varepsilon\delta$, где ε – число, определяющее соотношение между c_1 и c_2 (например, $\varepsilon = 0{,}1$). Как мы знаем, существует стратегия R, которая, во-первых, соответствует правилу выбора (в том смысле, что правило "знает", каких именно битов оно стремится набрать побольше), а во-вторых, набирает капитал не меньше $0{,}5L\delta \ln 2$. Таким образом, если мы не будем учитывать некоторое множество хороших пар $\langle r_1, r_2 \rangle$ ставок, где ставка r_1 проиграна, то подсчитанный капитал все еще будет составлять как минимум $L\delta(0{,}5 \ln 2 - \varepsilon)$.

Пусть игра в соответствии со стратегией R завершилась, и все ставки отмечены на действительной оси с указанием выигрышей и проигрышей. Произведем удаление пар ставок по следующему правилу: перебираем ставки в порядке возрастания и для каждого (неудаленного к данному моменту) проигрыша r_1 смотрим, существует ли (неудаленный) выигрыш r_2, образующий хорошую пару $\langle r_1, r_2 \rangle$, и если существует, берем наибольший такой r_2 и удаляем пару $\langle r_1, r_2 \rangle$. Понятно, что после всех удалений все оставшиеся проигрыши будут меньше всех оставшихся выигрышей, и интервал между этими двумя множествами (обозначим их, соответственно, P и V) будет больше $\varepsilon\delta$. Будем устанавливать пороги, начиная с нуля, с шагом $\varepsilon\delta$, тогда число правил составит примерно $(\varepsilon\delta)^{-1}$. Если P пусто, рассмотрим порог $T = 0$, иначе рассмотрим наименьший порог T из интервала между P и V. Покажем, что для любой удаленной пары $\langle r_1, r_2 \rangle$ не может быть $r_2 < T \le r_1$. Действительно, если предположить такую возможность, то в соответствии со способом удаления пар выигрыш r_2 еще до удаления спарился бы с наибольшим элементом из P, и этот элемент был бы удален.

Итак, в любом случае разность d между числом реальных (т.е. лежащих не ниже порога T) выигрышей и проигрышей будет не меньше $|V|$. Учитывая, что $|V| \ge L\delta(0{,}5 \ln 2 - \varepsilon)$, получаем следующую оценку на удельный дефект D правила (здесь n – длина выбранной подпоследовательности):

$$D = \frac{2\,\mathrm{lb}\, e}{L} \left(\frac{d}{2n}\right)^2 n = \frac{d^2\, \mathrm{lb}\, e}{2Ln} \ge \frac{L^2\delta^2(0{,}5 \ln 2 - \varepsilon)^2}{2L^2} = 0{,}5\, \mathrm{lb}\, e(0{,}5 \ln 2 - \varepsilon)^2\delta^2,$$

что и доказывает теорему.

Автор глубоко признателен Ан.А. Мучнику, который привлек его внимание к данной теме и высказал много полезных замечаний, способствовавших существенному улучшению текста статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Мучник Ан.А., Семёнов А.Л.* О роли закона больших чисел в теории случайности // Пробл. передачи информ. 2003. Т. 39. № 1. С. 134–165.
2. *Ширяев А.Н.* Вероятность—1. Элементарная теория вероятностей. Математические основания. Предельные теоремы. М.: МЦНМО, 2004.
3. *Успенский В.А., Семёнов А.Л., Шень А.Х.* Может ли (индивидуальная) последовательность нулей и единиц быть случайной? // Успехи мат. наук. 1990. Т. 45. № 1. С. 105–162.
4. *Muchnik An.A, Semenov A.L., Uspensky V.A.* Mathematical Metaphysics of Randomness // Theoret. Computer Science. 1998. V. 207. № 1–2. P. 263–317.

Горбунов Константин Юрьевич
Институт проблем передачи информации
им. А.А. Харкевича РАН .
gorbunov@iitp.ru

Поступила в редакцию
17.10.2006