Об универсальных алгоритмах адаптивного прогнозирования *

В.В.Вьюгин

Аннотация

В последнее десятилетие были разработаны методы прогнозирования произвольной последовательности исходов, отличные от традиционных подходов математической статистики. В частности, “успешно” прогнозировать можно любую последовательность данных не используя никаких гипотез об источнике их порождения. Один из таких методов в модифицированной форме приводится в данной работе. Основная часть работы посвящена анализу универсальных методов прогнозирования на основе теории алгоритмов, определению границ их возможностей.

1 Введение

Рассматривается следующая задача прогнозирования: предсказатель должен выдавать вероятности p_n будущего события $\omega_n = 1$ по двоичным исходам $\omega_1, \omega_2, \dots, \omega_{n-1}$.

При вероятностном подходе мы предполагаем, что исходы генерируются с помощью вероятностной меры, т.е., $p_n = P(\omega_n = 1|\omega_1, \omega_2, \dots, \omega_{n-1})$, $n = 1, 2, \dots$, - условное вероятность события $\omega_n = 1$ при известных значениях $\omega_1, \omega_2, \dots, \omega_{n-1}$. В этом случае, предсказатель должен решать классическую задачу математической статистики - восстановление меры P по наблюдениям.

Однако на практике мы часто имеем дело с единственной исторической последовательностью исходов $\omega_1, \omega_2, \dots, \omega_{n-1}$, и меру, описывающую будущие значения величины ω_n, довольно трудно, а часто и невозможно, определить. В данной работе предположение о наличии такой меры не используется. В условиях отсутствия меры возникает естественная трудность - неизвестно, каким образом оценивать качество прогнозов. Требуются критерии качества, использующие только последовательность данных, получаемую предсказателем в режиме он-лайн. Здесь можно усмотреть сходство этой проблемы с определением индивидуальной случайной последовательности, предложенным фон Мизесом и развитием этого подхода в теории колмогоровской сложности и алгоритмической случайности [2, 3].

Тем не менее, оказалось, что можно указать метод прогнозирования произвольной последовательнсти $\omega_1, \omega_2, \dots, \omega_{n-1}$, удовлетворяющий требованию *калибруемости*. Точное определение калибруемости будет дано ниже. Идею метода калибруемости, предложенного Дейвидом [4], разъясним на следующем примере. Рассматривается задача предсказания погоды на завтра. Например, событие $\omega_n = 1$ может интерпретироваться как дождь в n-ый день, а число p_n - как его вероятность, вычисленная на основе наблюдений погоды $\omega_1, \omega_2, \dots, \omega_{n-1}$ за предыдущие $n-1$ дней. Предсказатель погоды считается хорошо калибруемым, если дождь случается также часто, как он прогнозируется предсказателем. Например, если дождь случается в 80% всех дней, для которых предсказатель давал прогноз $p_n = 0.8$ и т.д. В этом примере $p_n \in [0, 1]$.

Введем необходимые обозначения. Пусть $\Omega = \{0, 1\}^\infty$ - множество всех бесконечных двоичных последовательностей, $\Xi = \cup_{n=1}^{\infty}\{0, 1\}^n$ - множество всех конечных двоичных последовательностей, λ - пустая последовательность. Для произвольной конечной или бесконечной последовательности $\omega = \omega_1 \dots \omega_n \dots$, пишем $\omega^n = \omega_1 \dots \omega_n$ (полагаем $\omega_0 = \omega^0 = \lambda$). Пусть $l(\omega^n) = n$ - длина последовательности ω^n. Если x - конечная последовательность, ω - конечная или бесконечная последовательность, то $x\omega$ обозначает конкатенацию этих последовательностей; $x \sqsubseteq \omega$ означает, что $x = \omega^n$ для некоторого n.

*Данная работа представляет собой расширенную версию доклада [1] на конференции CSR-2008 (Computer Science Symposium in Russia), Moscow, 2008. Работа частично поддержана грантом РФФИ 06-01-00122-а.

Согласно *преквенциальному* принципу, предложенному Дейвидом [4], качество прогнозов предсказателя должно оцениваться только на основе изучаемой последовательности исходов $\omega_1, \omega_2, \dots$ и последовательности его предсказаний $p_1, p_2, \dots$.

Мы предполагаем, что предсказатель имеет некоторый метод (алгоритм) для вычисления прогноза p_n по значениям $\omega_1, \dots \omega_{n-1}$. Под *детерминированной предсказательной системой* понимается произвольная, возможно частичная, функция $f : \Xi \to [0,1]$.

Приведем точное определение калибруемости, предложенное Дейвидом [5]. Мы рассмотрим произвольные подинтервалы $I = [a,b]$, $(a,b]$, $[a,b)$, (a,b) интервала $[0,1]$ и их характеристические функции

$$I(p) = \begin{cases} 1 \text{ если } p \in I \\ 0 \text{ в противном случае,} \end{cases}$$

Предсказательная система f *калибруется* на бесконечной последовательности $\omega_1\omega_2\dots$, если для характеристической функции $I(p)$ любого подинтервала $I \subseteq [0,1]$ *калибровочная ошибка* стремится к нулю, т.е.

$$\frac{\sum_{i=1}^{n} I(p_i)(\omega_i - p_i)}{\sum_{i=1}^{n} I(p_i)} \longrightarrow 0, \qquad (1)$$

если знаменатель отношения (1) стремится к бесконечности при $n \to \infty$, где $p_i = f(\omega^{i-1})$. Характеристическая функция $I(p_i)$ определяет некоторое правило выбора, которое определяет те номера исходов i, для которых мы вычисляем отклонение прогноза p_i от соответствующего исхода ω_i.

Простые соображения, впервые приведенные в [6, 7], показывают, что никакая детерминированная предсказательная система не может хорошо калиброваться на любой бесконечной последовательности исходов. А именно, для произвольной предсказательной системы f можно определить последовательность $\omega = \omega_1\omega_2\dots$ так, что

$$\omega_i = \begin{cases} 1 \text{ если } p_i < \frac{1}{2} \\ 0 \text{ в противном случае,} \end{cases}$$

где $p_i = f(\omega_1 \dots \omega_{i-1})$, $i = 1, 2, \dots$. Легко видеть, что для $I = [0, \frac{1}{2})$ или для $I = [\frac{1}{2}, 1)$ условие (1) нарушается при $n \to \infty$.

Подобные трудности оказались оказались преодолимыми с помощью понятия рандомизированной предсказательной системы. Пусть $\mathcal{P}[0,1]$ - множество всех вероятностных мер на отрезке $[0,1]$. Под *рандомизированной* предказательной системой понимается функция $f : \Xi \to \mathcal{P}[0,1]$, значениями которой являются распределения вероятностей на $[0,1]$. Мы также обозначаем $Pr_x(\cdot) = f(x)$.

Для каждой бесконечной последовательности ω по теореме Ионеску-Тульчи (см. [8]) о продолжении меры вероятностные распределения $Pr_{\omega^{i-1}}(\cdot)$ порождают распределение вероятностей $Pr = \prod_{i=1}^{\infty} Pr_{\omega^{i-1}}$ на множестве всех бесконечных последовательностей $p_1, p_2, \dots$, где $p_i \in [0,1]$, $i = 1, 2, \dots$.

В этом случае, при фиксированной последовательности ω можно рассматривать вероятность Pr события (1). Вероятностное распределение Pr, а также теорема 1, приводимая ниже, имеют смысл и в более общей ситуации, а именно, когда исход ω_{n-1} является измеримой функцией от $p_1, \dots, p_{n-1}$, т.е., когда неизвестный источник, генерирующий последовательность исходов, адаптируется к прогнозам предсказателя.

Фостер и Вохра [9], а также Какаде и Фостер [10], определили для каждого $\Delta > 0$ рандомизированную предсказательную систему f такую, что для произвольной бесконечной последовательности $\omega = \omega_1\omega_2\dots$ вероятность Pr события

$$\left| \frac{1}{n} \sum_{i=1}^{n} I(\tilde{p}_i)(\omega_i - \tilde{p}_i) \right| \le \Delta$$

стремится к 1 при $n \to \infty$, где $\tilde{p}_i$ распределено по вероятностной мере $Pr_{\omega_{n-1}}(\cdot) = f(\omega_1, \dots, \omega_{n-1})$, $I(p)$ - характеристическая функция произвольного подинтервала $[0,1]$. Такая функция называется правилом выбора, определенным прогнозом.

Определение калибруемости (1) обладает очевидным недостатком. Например, хорошо калибруемыми прогнозами для последовательности $0,1,0,1,0,1,0,1,\ldots$ является последовательность $\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\ldots$. Однако, если рассматривать только члены последовательности исходов, имеющие четные (или нечетные) индексы, подобные прогнозы уже не будут хорошо калибруемыми на соответствующей подпоследовательности. Поэтому необходимо ввести в расмотрение дополнительные правила выбора подпоследовательностей.

В работах [11, 12] понятие калибруемости было расширено для правил выбора более общего типа $c(\omega^{i-1},p)=\delta(\omega^{i-1})I(p)$, где δ есть функция типа $\Xi\to\{0,1\}$ - правило выбора, основанное на исходной последовательности, а $I(p)$ - характеристическая функция некоторого подинтервала $[0,1]$.

Мы приводим некоторый вариант этого результата в теореме 1.

Теорема 1 *Для произвольнй счетной последовательности всюду определенных правил выбора δ_i, $i=1,2,\ldots$, можно построить рандомизированную прогнозирующую систему $Pr_n(\cdot)$ такую, что для произвольной бинарной последовательности $\omega_1\omega_2\ldots$ для любого s и для характеристической функции I произвольного подинтервала $[0,1]$ Pr-вероятность события*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}I(\tilde{p}_i)\delta_s(\omega^{i-1})(\omega_i-\tilde{p}_i)=0 \tag{2}$$

равна 1, где Pr-распределение на $[0,1]^\infty$, порожденное последовательностью распределений $Pr_{\omega^{n-1}}(\cdot)=f(\omega_1\ldots\omega_{n-1})$, $n=1,2,\ldots$, последовательность прогнозов $\tilde{p}_1,\tilde{p}_2,\ldots$ распределена по мере Pr.

Доказательство этой теоремы приведено в разделе 3. [1]

Было бы естественно распространить этот результат на все вычислимые правила выбора $\delta(x)$. [2]

Основной результат этой работы - теорема 2 (приводимая ниже в разделе 2), показывает, что это невозможно сделать: существуют последовательности исходов $\omega_1,\omega_2,\ldots$, содержащие подпоследовательности, которые не поддаются алгоритмически эффективному рандомизированному прогнозированию. Точнее, существуют такие последовательности исходов $\omega_1\omega_2\ldots$, что для каждого рандомизированного предсказательного алгоритма f существует правило выбора $\delta_f(x)\notin\{\delta_i(x):i=1,2,\ldots\}$, которое выбирает из каждой такой последовательности подпоследовательность $\omega_{i_1}\omega_{i_2}\ldots$, на которой f не калибруется.

Более того, существует вероятностная машина, которая генерирует такие последовательности с близкой к 1 вероятностью.

Этот результат прямо не противоречит теореме 1. В теореме 2 последовательность правил выбора $\delta_i(x)$ $i=1,2,\ldots$, задана априори, т.е. используется при построении универсального прогнозирующего алгоритма. Для того, чтобы метод построения универсального алгоритма был алгоритмически эффективным, правила выбора из последовательность $\delta_i(x)$ должны быть вычислимыми и всюду определенными, более того, функция $K(i,x)=\delta_i(x)$ должна быть всюду определенной и вычислимой. Один из результатов теории алгоритмов утверждает, что такой функции $K(i,x)$ не существует для последовательности $\delta_i(x)$ всех вычислимых всюду определенных правил выбора (см. [13]). Таким образом, для каждой последовательности правил выбора из теоремы 1 должны существовать всюду определенные вычислимые правила выбора $\delta(x)\notin\{\delta_i(x):i=1,2,\ldots\}$.

В доказательстве теореме 2 для каждой рандомизированной предсказательной системы f (в том числе, для каждой вычислимой) строится свое "естественное" (вычислимое) правило выбора $\delta_f(x)$. Мы также указываем способ генерации последовательности $\omega_1\omega_2\ldots$. Правило выбора $\delta_f(x)$ выбирает подпоследовательность из этой последовательности, на которой f не калибруется с вероятностью 1. Это верно и для того рандомизированного универсального алгоритма f, который строился для последовательности правил выбора $\delta_i(x)$ $i=1,2,\ldots$.

[1] Более точная оценка: в (2) множитель $\frac{1}{n}$ можно заменить на $\frac{1}{\sqrt{\alpha(n)n}}$, где $\alpha(n)$ - произвольная монотонно возрастающая неограниченная функция.

[2] Точнее, предлагается рассматривать частично рекурсивные (в смысле [13]) функции $\delta(x)$.

2 Непредсказуемые последовательности

Мы будем использовать некоторые стандартные понятия теории алгоритмов (см., например, [13]). Фиксируем некоторую алгоритмически эффективную взаимно-однозначную нумерацию всех пар (троек, и т.д.) неотрицательных целых чисел. Мы отождествляем произвольную пару (t, s) и ее номер $\langle t, s\rangle$; пусть также $p(\langle t, s\rangle) = t$ для всех s, t.

Пусть $\mathcal{R}$ - множество всех вещественных чисел. Функция $\phi\colon A \to \mathcal{R}$ называется полувычислимой (снизу), если множество $\{(r, x) : r < \phi(x)\}$ (r рационально) рекурсивно перечислимо. Функция ϕ перечислима сверху, если $-\phi$ - перечислима снизу. Используя стандартные методы теории алгоритмов, легко показать, что существуют перечислимая снизу и перечислимая сверху вещественнозначные функции $\phi^-(j, x)$ и $\phi^+(k, x)$, универсальные для всех перечислимых снизу и всех перечислимых сверху функций типа $\Xi \to \mathcal{R}$. Всюду определенная функция $\phi(x)$ типа $\Xi \to \mathcal{R}$ называется вычислимой, если она может быть представлена в виде $\phi(x) = \phi^-(j, x) = \phi^+(k, x)$ для всех x, для некоторых j и k. Легко видеть, что в этом случае, существует алгоритм, позволяющий вычислять значения $\phi(x)$ с произвольной наперед заданной степенью точности.

Пусть $\phi_s^-(j, x)$ равно максимальному рациональному числу r такому, что тройка (r, j, x) перечислена за s шагов в процессе перечсления множества $\{(r, j, x) : r < \phi(j, x),\ r \text{ рационально}\}$ и равна $-\infty$, в противном случае.

Каждая такая функция $\phi_s^-(j, x)$ принимает только конечное число рациональных значений отличных от $-\infty$. По определению $\phi_s^-(j, x) \le \phi_{s+1}^-(j, x)$ для всех j, s, x, и $\phi^-(j, x) = \lim\limits_{s\to\infty} \phi_s^-(j, x)$. Аналогичная невозрастающая последовательность функций $\phi_s^+(k, x)$ существует для всех перечислимых сверху функций.

Пусть $i = \langle t, k\rangle$. Функция $\phi_i(x)$ типа $\Xi \to \mathcal{R}$ *определена* на x, если для произвольного положительного рационального числа $\kappa > 0$ выполнено $|\phi_s^+(t, x) - \phi_s^-(k, x)| \le \kappa$ для некоторого s; $\phi_i(x)$ неопределена на x, в противном случае. Для минимального такого s, если такое s существует, значение $\phi_{i,\kappa}(x) = \phi_s^-(k, x)$ называется рациональным приближением (снизу) значения $\phi_i(x)$ с точностью до κ; $\phi_{i,\kappa}(x)$ неопределено, в противном случае.

Для того, чтобы определить вероятностную меру P на Ω, определим $P(z) = P(\Gamma_z)$ для всех интервалов $\Gamma_z = \{\omega \in \Omega : z \sqsubseteq \omega\}$, где $z \in \Xi$, и далее, продолжим эту функцию на все борелевские подмножества Ω по теореме Колмогорова о продолжении меры [8].

Нам также потребуется понятие вычислимого оператора на $\Xi \bigcup \Omega$ (см. [14]). Пусть $\hat{F}$ - некоторое рекурсивно перечислимое множество упорядоченных пар конечных последовательностей такое, что (i) $(x, \lambda) \in \hat{F}$ для всех x; (ii) если $(x, y) \in \hat{F}$, $(x', y') \in \hat{F}$ и $x \sqsubseteq x'$, то $y \sqsubseteq y'$ или $y' \sqsubseteq y$ для всех $x, x', y, y' \in \Xi$. Вычислимый оператор F определяется

$$F(\omega) = \sup\{y \mid x \sqsubseteq \omega \text{ и } (x, y) \in \hat{F} \text{ для некоторого } x\},$$

где $\omega \in \Omega \bigcup \Xi$ и sup относятся к частичному порядку $\sqsubseteq$ на Ξ.

Вероятностный алгоритм - это пара (L, F), где $L(x) = L(\Gamma_x) = 2^{-l(x)}$ - равномерная мера на Ω и F - вычислимый оператор. Для произвольного вероятностного алгоритма (L, F) и множества $A \subseteq \Omega$ рассмотрим вероятность $L\{\omega : F(\omega) \in A\}$ генерации последовательности из A с помощью F.

Мы будем рассматривать частично определенные рандомизированные предсказательные системы $Pr_x(\cdot) = f(x)$, где $x \in \Xi$. С каждой такой системой свяжем функцию $\varphi(x) = Pr_x([0, \frac{1}{2}))$. Предсказательная система f называется слабо вычислимой, если $\varphi(x) = \phi_i(x)$ для некоторого i.

Будем считать, что такая система определена на $x \in \Xi$, если функция $\varphi(x) = \phi_i(x)$ определена на x в указанном выше смысле.

Основной результат представлен в следующей теореме.

Теорема 2 *Для произвольного рационального $\epsilon > 0$ можно построить вероятностный алгоритм (L, F), который с вероятностью $\ge 1 - \epsilon$ выдает бесконечную двоичную последовательность $\omega = \omega_1\omega_2\ldots$ такую, что для любой частичной слабо вычислимой рандомизированной предсказательной системы f, определенной на всех начальных фрагментах ω, существует вычислимое правило выбора δ_f, определенное на всех этих фрагментах, и такое, что для $\nu = 0$ или для $\nu = 1$*

Pr-вероятность события

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\delta_f(\omega^{i-1})I_\nu(\tilde{p}_i)(\omega_i-\tilde{p}_i)\right|\geq 1/16 \tag{3}$$

равна 1, где I_0 и I_1 - характеристические функции отрезков $[0,\frac{1}{2})$ и $[\frac{1}{2},1]$, $\tilde{p}_1,\tilde{p}_1,\dots$ - траектории прогнозов, распределенные по мере Pr, а вероятностная мера Pr порождается распределениями $Pr_{\omega^{i-1}}(\cdot)=f(\omega^{i-1})$, $i=1,2,\dots$.

Правило выбора $\delta_f(x)$ является всюду определенным, если всюду определенной является предсказательная система f.

Доказательство. Для произвольного вероятностного алгоритма (L,F) рассмотрим функцию

$$Q(x)=L\{\omega: x\sqsubseteq F(\omega)\}. \tag{4}$$

Легко видеть, что эта функция удовлетворяет условиям: (i) $Q(\lambda)\leq 1$; (ii) $Q(x0)+Q(x1)\leq Q(x)$ для всех x; (iii) Q перечислима снизу.

Произвольная функция, удовлетворяющая условиям (i)-(iii), называется полувычислимой полумерой. Для любой полувычислимой полумеры Q может быть определен вероятностный алгоритм (L,F) такой, что имеет место (4).

В общем случае, полумера Q не является мерой. Рассмотрим связанную с ней меру, которая задается соотношениями

$$\bar{Q}(\Gamma_x)=\inf_n\sum_{l(y)=n,x\sqsubseteq y}Q(y).$$

В дальнейшем, мы определим полувычислимую полумеру Q как поток по некоторой сети. Сеть будет определена на основе бесконечного дерева конечных двоичных последовательностей.

Первоначально, каждая последовательность $x\in\Xi$ определяет два направленных ребра $(x,x0)$ и $(x,x1)$ единичной длины. В процессе конструкции мы будем добавлять к сети дополнительные направленные ребра (x,y) длины >1, где $x,y\in\Xi$, $x\sqsubseteq y$ и $y\neq x0,x1$. Под длиной ребра (x,y) понимается число $l(y)-l(x)$. Для произвольного направленного ребра $\sigma=(x,y)$ посредством $\sigma_1=x$ обозначаем его начальную вершину, а посредством $\sigma_2=y$ обозначаем его конечную вершину. Вычислимая функция $q(\sigma)$, определенная на всех ребрах и принимающая рациональные значения, называется *сетью*, если для всех $x\in\Xi$

$$\sum_{\sigma:\sigma_1=x}q(\sigma)\leq 1.$$

Пусть G - множество всех дополнительных ребер. Под *q-потоком* понимаем минимальную полумеру P такую, что $P\geq R$, где функция R определена следующим рекурсивными соотношениями:

$$R(\lambda)=1,$$
$$R(y)=\sum_{\sigma:\sigma_2=y}q(\sigma)R(\sigma_1) \tag{5}$$

для $y\neq\lambda$. Легко заметить, что такая полумера существует и единственная. Кроме этого, нетрудно построить алгоритм для перечисления снизу ее значений.

Сеть q называется *элементарной*, если множество всех дополнительных ребер конечно и $q(\sigma)=1/2$ для почти всех ребер единичной длины. Ясно, что элементарная сеть задается конечным объектом.

Для любой сети q определим *функцию задержки* q-потока (функция q-задержки)

$$d(x)=1-q(x,x0)-q(x,x1).$$

Приводимая ниже конструкция будет использовать все вычислимые функции $\phi_t(x)$, $x\in\Xi$, $t=1,2,\dots$.

Мы предполагаем, что для любой такой вычислимой функции ϕ существует бесконечно много программ t таких, что $\phi_t = \phi$. [3] Произвольная упорядоченная пара $i = \langle t, s\rangle$ может рассматриваться как программа для вычисления рационального приближения $\phi_{t,\kappa_s}(\omega^{n-1})$ снизу для ϕ_t с точностью до $\kappa_s = 1/s$.

В конструкции, приводимой ниже, каждая функция ϕ_{t,κ_s} будет просматриваться на бесконечном числе шагов n. Для этого будем использовать функцию $p(n)$, определенную выше. Для произвольного неотрицательного числа i имеем $p(n) = i = \langle t, s\rangle$ для бесконечно многих n.

Пусть β - конечная последовательность и $1 \le k < l(\beta)$. Бит β_k последовательности β называется *трудно предсказуемым* посредством программы $i = \langle t, s\rangle$, если $\phi_{t,\kappa_s}(\beta^{k-1})$ определено и

$$\beta_k = \begin{cases} 0, \text{ если } \phi_{t,\kappa_s}(\beta^{k-1}) \ge \frac{1}{2} \\ 1, \text{ в противном случае.} \end{cases}$$

Лемма 1 *Пусть $i = \langle t, s\rangle$ - произвольная программа и μ - произвольное достаточно малое положительное рациональное число.*

Тогда для произвольной двоичной последовательности x длины n доля всех двоичных последовательностей γ длины $K = \lceil(2+\mu)i\rceil n$ (в множестве $\{0,1\}^K$) таких, что

1) $\phi_{t,\kappa_s}(x\gamma^k)$ определено для всех $0 \le k < K$,

2) число всех трудно предсказуемых битов γ посредством программы i меньше чем in, не превосходит $2^{-2\mu^2 in + O(\log(in))}$ для всех достаточно больших n.

Доказательство. Функция $\sigma(x)$, где $x \in \Xi$ и $\sigma(x) \in \{A, B\}$, назывется *разметкой*, если $\sigma(x0) \ne \sigma(x1)$ для всех $x \in \Xi$. Для произвольного γ длины K и для любого k такого, что $1 \le k < K$, определим $\sigma(\gamma^{k+1}) = A$ и $\sigma(\gamma^k\bar{\gamma}_{k+1}) = B$, если бит γ_{k+1} последовательности $x\gamma$ является трудно предсказуемым, где $\bar{\theta} = 1 - \theta$ для двоичного бита θ. Определим $\sigma(\gamma^{k+1}) = B$ и $\sigma(\gamma^k\bar{\gamma}_{k+1}) = A$ в противном случае. По определению, один из битов γ_{k+1} или $\bar{\gamma}_{k+1}$ трудно предсказуем.

Пусть γ удовлетворяет условиям 1) и 2). Так как $\phi_{t,\kappa_s}(x\gamma^k)$ определено для всех $0 \le k < K$, значение $\sigma(\gamma^{k+1})$ также определено для всех этих k. Эту частичную разметку σ легко расширить до разметки всего множества $\{0,1\}^K$, причем, это можно сделать многими способами.

Фиксируем некоторое такое расширение. Тогда общее число всех γ, удовлетворяющих 1)-2), не превосходит общего числа всех двоичных последовательностей длины K, размеченых не более чем in метками A. Поэтому для всех достаточно больших n доля таких γ не превосходит

$$\sum_{j \le in} \binom{K}{j} 2^{-K} \le 2^{-(1-H(1/2-\mu))2in + O(\log(in))} \le 2^{-2\mu^2 in + O(\log(in))},$$

где $H(r) = -r\log r - (1-r)\log(1-r)$. $\triangle$

Полагаем $\mu = 1/\log(i+1)$.

Определим вспомогательную отношение $B(i, q^{n-1}, \sigma, n)$ и функцию $\beta(x, q^{n-1}, n)$. Пусть $x, \beta \in \Xi$. Отношение $B(i, q^{n-1}, (x,\beta), n)$ имеет значение *истина*, если выполнено

- $n \ge (1 + \lceil(2 + \log^{-1}(i+1))i\rceil)l(x)$;
- $l(\beta) = n$ и $x \sqsubseteq \beta$;
- $d^{n-1}(\beta^j) < 1$ для всех j таких, что $1 \le j < n$;
- для всех j, $l(x) < j \le (1 + \lceil(2+\log^{-1}(i+1))i\rceil)l(x)$, значение $\phi_{t,\kappa_s}(\beta^{j-1})$ может быть вычислено за $\le n$ шагов и, по крайней мере, для $il(x)$ таких j бит β_j трудно предсказуем посредством программы $i = \langle t, s\rangle$.

Значение $B(i, q^{n-1}, (x,\beta), n)$ *ложно*, в противном случае.

Определим

$$\beta(x, q^{n-1}, n) = \min\{y : p(l(y)) = p(l(x)), B(p(l(x)), q^{n-1}, (x,y), n)\},$$

[3] Для того, чтобы это свойство выполнялось, заменим последовательность $\phi_t(x)$, определенную в начале этого раздела, на последовательность $\phi'_t(x)$, определенную соотношением $\phi'_{\langle t,s\rangle}(x) = \phi_t(x)$ для всех s, t.

min рассматривается относительно лексикорафического порядка на множестве всех двоичных строк; считаем, что $\min\emptyset$ не определен.

Конструкция. Пусть $\rho(n) = (n + n_0)^2$ для некоторого достаточно большого n_0 (значение n_0 будет определено позже в доказательстве леммы 4).

Определим последовательность элементарных сетей q^n математической индукцией по n. Полагаем $q^0(\sigma) = 1/2$ для всех ребер σ длины 1.

Пусть $n > 0$ и сеть q^{n-1} определена, d^{n-1} - функция q^{n-1}-задержки, и пусть G^{n-1} - множество всех дополнительных ребер. Предположим, что $l(\sigma_2) < n$ для всех $\sigma \in G^{n-1}$.

Определим сеть q^n. Для этого, предварительно определим функцию задержки потока d^n и множество G^n.

Конструкция распадается на три случая.

Пусть $w(i, q^{n-1})$ равно минимальному m такому, что $p(m) = i$ и $m > l(\sigma_2)$ для всех дополнительных ребер $\sigma \in G^{n-1}$ таких, что $p(l(\sigma_1))) < i$.

Неравенство $w(i, q^m) \neq w(i, q^{m-1})$ может быть вызвано выполнением некоторого *задания* $j < i$, которое добавляет дополнительное ребро $\sigma = (x, y)$ такое, что $l(x) > w(i, q^{m-1})$ и $p(l(x)) = p(l(y)) = j$. Лемма 2 (ниже) показывает, что это может случиться только на конечном числе шагов конструкции.

Случай 1. $w(p(n), q^{n-1}) = n$ (цель этой части конструкции заключается в назначении и начале выполнения нового задания $i = p(n)$ или в восстановлении ранее выполнявшегося задания $i = p(n)$, если оно было разрушено некоторым заданием $j < i$ (с более высоким приоритетом) на некотором предшествующем шаге).

Полагаем $d^n(y) = 1/\rho(n)$ для всех y, $l(y) = n$, и определим $d^n(y) = d^{n-1}(y)$ для всех других y. Полагаем также $G^n = G^{n-1}$.

Случай 2. $w(p(n), q^{n-1}) < n$ (цель этой части провести очередной этап обработки задания $i = p(n)$). Пусть C_n - множество всех таких x, что $w(i, q^{n-1}) \leq l(x) < n$, $0 < d^{n-1}(x) < 1$, функция $\beta(x, q^{n-1}, n)$ определена [4] и не существует ни одного дополнительного ребра $\sigma \in G^{n-1}$ такого, что $\sigma_1 = x$.

В этом случае, для каждого $x \in C_n$ определим $d^n(\beta(x, q^{n-1}, n)) = 0$, а для всех остальных y длины n таких, что $x \sqsubset y$ определим

$$d^n(y) = \frac{d^{n-1}(x)}{1 - d^{n-1}(x)}.$$

Определим $d^n(y) = d^{n-1}(y)$ для всех остальных y.

Добавим новое дополнительное ребро к G^{n-1}, а именно, определим

$$G^n = G^{n-1} \cup \{(x, \beta(x, q^{n-1}, n)) : x \in C_n\}.$$

Говорим, что задание $i = p(n)$ *добавляет* дополнительное ребро $(x, \beta(x, q^{n-1}, n))$ к сети, а все ранее установленные задания $j > i$ разрушены заданием i.

В конце случая 1 и случая 2 определим для каждого ребра σ единичной длины

$$q^n(\sigma) = \frac{1}{2}(1 - d^n(\sigma_1))$$

и $q^n(\sigma) = d^n(\sigma_1)$ для всех дополнительных ребер $\sigma \in G^n$.

Случай 3. Случаи 1 и 2 не имеют места. Определим $d^n = d^{n-1}$, $q^n = q^{n-1}$, $G^n = G^{n-1}$.

В результате выполнения конструкции будет определена сеть $q = \lim\limits_{n\to\infty} q^n$, функцию задержки потока $d = \lim\limits_{n\to\infty} d^n$ и множество дополнительных ребер $G = \cup_n G^n$.

Функции q и d вычислимые, множество G рекурсивно перечислимо. Пусть Q обозначае q-поток.

Следующая лемма показывает, что каждое задание может добавлять дополнительные ребра только на конечном числе шагов.

Пусть $G(i)$ - множество всех дополнительных ребер, добавленных заданием i, $w(i, q) = \lim_{n\to\infty} w(i, q^n)$.

---

[4] В частности, $p(l(x)) = i$ и $l(\beta(x, q^{n-1}, n)) = n$.

Лемма 2 *Множество $G(i)$ конечное, $w(i,q)$ существует и $w(i,q) < \infty$ для всех i.*

Доказательство. Заметим, что если $G(j)$ конечное множество для всех $j < i$, то $w(i,q) < \infty$. Таким образом, мы должны доказать, что множество $G(i)$ конечное для всех i.

Допустим противное. Пусть i - минимальное число такое, что $G(i)$ бесконечное множество. По выбору i, множество $G(j)$ конечно для всех $j < i$. Тогда $w(i,q) < \infty$.

По определению, если $d(\omega^m) \neq 0$, то $p_m = 1/d(\omega^m)$ - положительное целое число. Кроме того, если $(\omega^n, y), (\omega^m, y') \in G(i)$, где $n < m$ и $l(y) = m$, то $p_n > p_m$. Следовательно, для каждого $\omega \in \Omega$ существует максимальное m такое, что $(\omega^m, y) \in G(i)$ для некоторого y или же не существует ни одного такого дополнительного ребра. В последнем случае, полагаем $m = w(I,q)$. Определим $u(\omega) = 1/d(\omega^m)$.

По конструкции, целочисленная функция $u(\omega)$ постоянна на интервале Γ_{ω^m}. Следовательно, она непрерывна в топологии, образованной такими интервалами. Так как Ω - компакт в этой топологии, $u(\omega)$ ограничена сверху. Тогда для некоторого m', $u(\omega) = u(\omega^{m'})$ для всех ω. По конструкции, если некоторое дополнительное ребро i-того типа будет добавлено в $G(i)$ на некотором шаге, то $d(y) > d(x)$ выполнено для некоторой новой пары (x,y) такой, что $x \subseteq y$. Это противоречит предположению о том, что $G(i)$ бесконечное множество. $\triangle$

Бесконечная последовательность $\alpha \in \Omega$ называется *i-расширением* конечной последовательности x, если $x \sqsubseteq \alpha$ и $B(i, q^{n-1}, x, \alpha^n, n)$ истинно для почти всех n.

Последовательность $\alpha \in \Omega$ называется *i-запрещенной*, если $d(\alpha^n) = 1$ для некоторого n такого, что $p(n) = i$, где d - функция q-задержки. По конструкции, если $\sigma \in G(i)$, то $B(i, q^{n-1}, \sigma, n)$ истинно, где $n = l(\sigma_2)$.

Лемма 3 *Допустим, что для любого начального фрагмента ω^n бесконечной последовательности ω существует некоторое i-расширение. Тогда последовательность ω будет i-запрещенной или ω содержит (как подслово) некоторое дополнительное ребро i-го типа (т.е., $\sigma_2 \sqsubseteq \omega$ для некоторого $\sigma \in G(i)$).*

Доказательство. Допустим, что последовательность ω не является i-запрещенной. По лемме 2 существует максимальное m такое, что $p(m) = i$ и $d(\omega^m) > 0$. Так как последовательность ω^m имеет i-расширение и $d(\omega^m) < 1$, по случаю 2 некоторое новое дополнительное ребро i-го типа (ω^m, y) должно быть добавлено к бинарному дереву. По конструкции $d(y) = 0$ и $d(z) \neq 0$ для всех z таких, что $\omega^m \sqsubseteq z$, $l(z) = l(y)$, и $z \neq y$. По выбору m имеем $y \sqsubseteq \omega$. $\triangle$

Из определения потока следует, что $Q(y) = 0$ тогда и только тогда, когда $q(\sigma) = 0$ для некоторого ребра σ единичной длины, расположенного на y (т.е., такого, что $\sigma_2 \sqsubseteq y$).

Для произвольной полумеры P определим

$$E_P = \{\omega \in \Omega : \forall n (P(\omega^n) \neq 0)\}$$

- носитель P. Легко видеть, что $\bar{P}(E_P) = \bar{P}(\Omega)$, а также

$$E_Q = \Omega \setminus \cup_{d(x)=1} \Gamma_x.$$

Лемма 4 *Имеет место неравенство $\bar{Q}(E_Q) > 1 - \frac{1}{2}\epsilon$.*

Доказательство. Ограничим $\bar{Q}(\Omega)$ снизу. Пусть R определено по (5). По определению функции задержки потока имеем

$$\sum_{u:l(u)=n+1} R(u) = \sum_{u:l(u)=n} (1 - d(u))R(u) + \sum_{\sigma:\sigma \in G, l(\sigma_2)=n+1} q(\sigma)R(\sigma_1). \tag{6}$$

Определим вспомогательную функцию

$$S_n = \sum_{u:l(u)=n} R(u) - \sum_{\sigma:\sigma \in G, l(\sigma_2)=n} q(\sigma)R(\sigma_1).$$

Сначала рассмотрим случай $w(p(n), q^{n-1}) < n$. Если существует ребро $\sigma \in G$ такое, что $l(\sigma_2) = n$, то $S_{n+1} \geq S_n$. Допустим, что такое ребро существует. Определим

$$P(u, \sigma) \Longleftrightarrow l(u) = l(\sigma_2) \& \sigma_1 \sqsubseteq u \& u \neq \sigma_2 \& \sigma \in G.$$

По определению функции задержки потока имеем

$$\sum_{u: l(u)=n} d(u)R(u) = \sum_{\sigma: \sigma \in G, l(\sigma_2)=n} d(\sigma_2) \sum_{u: P(u,\sigma)} R(u) =$$
$$= \sum_{\sigma: \sigma \in G, l(\sigma_2)=n} \frac{d(\sigma_1)}{1 - d(\sigma_1)} \sum_{u: P(u,\sigma)} R(u) \leq \sum_{\sigma: \sigma \in G, l(\sigma_2)=n} d(\sigma_1)R(\sigma_1) =$$
$$= \sum_{\sigma: \sigma \in G, l(\sigma_2)=n} q(\sigma)R(\sigma_1). \tag{7}$$

Здесь мы использовали неравенство $\sum\limits_{u: P(u,\sigma)} R(u) \leq R(\sigma_1) - d(\sigma_1)R(\sigma_1)$ для всех $\sigma \in G$ таких, что $l(\sigma_2) = n$. Комбинируя эту оценку с (6), получим $S_{n+1} \geq S_n$.

Рассмотрим теперь случай $w(p(n), q^{n-1}) = n$. Имеем $\sum\limits_{u: l(u)=n} d(u)R(u) \leq \rho(n) = (n + n_0)^{-2}$. Комбинируя (6) и (7), получим $S_{n+1} \geq S_n - (n + n_0)^{-2}$ для всех n. Так как $S_0 = 1$, получим $S_n \geq 1 - \sum\limits_{i=1}^{\infty} (i + n_0)^{-2} \geq 1 - \frac{1}{2}\epsilon$ для некоторого достаточно большого числа n_0. Так как $Q \geq R$, получим

$$\bar{Q}(\Omega) = \inf_n \sum_{l(u)=n} Q(u) \geq \inf_n S_n \geq 1 - \frac{1}{2}\epsilon.$$

Лемма доказана. $\triangle$

Лемма 5 *Для любого $\epsilon > 0$ существует множество U бесконечных двоичных последовательностей такое, что $\bar{Q}(U) \leq \epsilon/2$ и для любой последовательности $\omega \in E_Q \setminus U$ и для любой частичной слабо вычислимой рандомизированной предсказательной системы выполнено условие (3) теоремы 2.*

Доказательство. Пусть ω - бесконечная двоичная последовательность и f - произвольная частичная слабо вычислимая рандомизированная предсказательная система такая, что для $Pr_x(\cdot) = f(x)$ соответствующая функция $\phi_t(\omega^{n-1}) = Pr_x([0, \frac{1}{2}))$ определена для всех n. Пусть $i = \langle t, s \rangle$ - программа, вычисляющая рациональные приближения ϕ_{t,κ_s} снизу, с точностью до $\kappa_s = 1/s$.

Если $d(\omega^m) = 1$ для некоторого m такого, что $p(m) = i$, то для каждого β длины $(1 + \lceil (2 + \log^{-1}(i+1) \rceil i)m$ такого, что $\omega^m \sqsubseteq \beta$, существует менее im битов, трудно предсказуемых посредством программы i.

Мы покажем, что $\bar{Q}$-мера объединения всех интервалов, порожденных такими β, будет как угодно мала для достаточно больших i. Так как не существует дополнительного ребра σ такого, что $\omega^m \sqsubseteq \sigma_1$, ограничение меры $\bar{Q}$ на интервале Γ_{ω^m} будет пропорционально равномерной мере. Тогда по лемме 1, где $\mu = \log^{-1}(i + 1)$, $\bar{Q}$-мера всех таких β убывает экспоненциально по im. Следовательно, для каждого j существует число m_j такое, что $\bar{Q}(U_j) \leq 2^{-(j+1)}$, где U_j - объединение всех интервалов Γ_β, определенных теми β длины $(1 + \lceil (2 + \log^{-1}(i + 1))i \rceil)m$, для которых $m \geq m_j$, и содержащих менее im битов, трудно предсказуемых программой $i = p(m)$. Определим $U = \cup_{j>k} U_j$, где $k = \lceil -\log_2 \epsilon - 1 \rceil$. Имеем $\bar{Q}(U) < \epsilon/2$.

Определим правило выбора γ следующим образом:

- $\gamma(\omega^{j-1}) = 1$, если $\sigma_1 \sqsubseteq \omega^{j-1} \sqsubseteq \sigma_2$ для некоторого $\sigma \in G(i)$ и j-ый бит σ_2 трудно предсказуем посредством программы i;
- $\gamma(\omega^{j-1}) = 0$ в противном случае.

Также определим два правила выбора J_ν, где $\nu = 0, 1$,

$$J_\nu(\omega^{j-1}) = \begin{cases} 1 - \nu \text{ если } \phi_{t,\kappa_s}(\omega^{j-1}) < \frac{1}{2} \\ \nu \text{ если } \phi_{t,\kappa_s}(\omega^{j-1}) \geq \frac{1}{2} \end{cases}$$

Допустим, что $\omega \notin U$ и $\phi_t(\omega^n)$ определено для всех n. Тогда ω является i-расширением ω^n для каждого n. Так как для каждого n последовательность ω^n не является i-запрещенной, по лемме 3 существует дополнительное ребро $\sigma \in G(i)$ такое, что $\sigma_2 \sqsubseteq \omega$. Далее, пусть $m = l(\sigma_1)$, $n = (1 + \lceil (2 + \log^{-1}(i+1))i \rceil)m$.

По конструкции одно из правил правил выбора $\delta_\nu(\omega^{j-1}) = \gamma(\omega^{j-1})J_\nu(\omega^{j-1})$, где $\nu = 0$ или $\nu = 1$, выбирает из фрагмента ω длины n подпоследовательность $\omega_{t_1}, \dots, \omega_{t_l}$ длины $l \geq im/2$. Так как все выбранные биты являются трудно предсказуемыми, имеем $\omega_{t_j} = 1$ для всех j таких, что $1 \leq j \leq l$, если они выбирались по правилу выбора δ_0, и $\omega_{t_j} = 0$ для всех j, если они выбирались по правилу выбора δ_1.

Напомним, что $Pr_x(\cdot) = f(x)$, $j = 1, 2, \dots$, - произвольная слабо вычислимая рандомизированная предсказательная система, такая, что функция $\phi(x) = Pr_x([0, \frac{1}{2}))$ определена на всех начальных фрагментах последовательности $\omega = \omega_1\omega_2\dots$. Кроме того, $\phi(x)$ - вычислимая функция. По определению $\phi = \phi_t$ для бесконечно многих t и

$$\phi_{t,\kappa_s}(\omega^{j-1}) \leq \phi_t(\omega^{j-1}) \leq \phi_{t,\kappa_s}(\omega^{j-1}) + \kappa_s. \tag{8}$$

для всех s и j. Рассмотрим две случайные величины, для $\nu = 0$ и $\nu = 1$,

$$\vartheta_{n,\nu}(\tilde{p}_j) = \sum_{j=1}^{n} \delta_\nu(\omega^{j-1}) I_\nu(\tilde{p}_j)(\omega_j - \tilde{p}_j).$$

Допустим, что $l \geq im/2$ для $\nu = 0$. Тогда, используя (8), получим для математического ожидания E по мере Pr

$$E(\vartheta_{n,0}) \geq \sum_{j=m+1}^{n} \delta_0(\omega^{j-1}) Pr\{\tilde{p}_j < \frac{1}{2}\}\frac{1}{2} - m \geq$$
$$\geq \frac{im}{4}(\frac{1}{2} - \kappa_s) - m \tag{9}$$

Аналогичное неравенство получим при $\nu = 1$. Так как $n = (1 + \lceil (2 + \log^{-1}(i+1))i \rceil)m$, i можно выбрать как угодно большим, и мы просматриваем в процессе конструкции каждую пару $i = \langle t, s \rangle$ бесконечное число раз, (9) или аналогичное нервенство для случая $\nu = 1$ будет иметь место для бесконечно многих m и i. Поэтому

$$\limsup_{n\to\infty} \frac{1}{n} E(\vartheta_{n,0}) \geq 1/16. \tag{10}$$

или

$$\liminf_{n\to\infty} \frac{1}{n} E(\vartheta_{n,1}) \leq -1/16. \tag{11}$$

По мартингальному строгому закону больших чисел, при $\nu = 0, 1$, Pr-вероятность события

$$\frac{1}{n} \sum_{j=1}^{n} \delta_\nu(\omega^{j-1}) I_\nu(\tilde{p}_j)(\omega_j - \tilde{p}_j) - \frac{1}{n} E(\vartheta_{n,\nu}) \to 0 \tag{12}$$

стремится к 1 при $n \to \infty$. Комбинируя (10), (11) и (12), получим (3).

Лемма 5 и теорема 2 доказаны. $\triangle$

Следующая теорема обобщает результат из [15] для частичных вычислимых детерминированных предсказательных систем.

Теорема 3 *Для произвольного $\epsilon > 0$ можно построить вероятностный алгоритм (L, F), который с вероятностью $\geq 1-\epsilon$ выдает бесконечную двоичную последовательность $\omega = \omega_1\omega_2\ldots$ такую, что для каждого предсказательного алгоритма f, определенного на всех начальных фрагментах последовательности ω, существует вычислимое правило выбора δ, определенное на всех начальных фрагментах последовательности ω, такое, что*

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\delta(\omega^{i-1})(\omega_i - f(\omega^{i-1}))\right| \geq 1/16 \tag{13}$$

Докаательство этой теоремы основано на конструкции теоремы 2, в котором $\phi_i(x)$ - все вычислимые детерминированные предсказательные системы $f(x)$.

3 Доказательство теоремы 1: алгоритм вычисления хорошо калибруемых прогнозов

Предварительно приведем конструкцию алгоритма, который удовлетворяет более слабым требованиям, а именно, для произвольного $\Delta > 0$ и счетной последовательность всюду определенных правил выбора δ_i, $i = 1, 2, \ldots$, построим рандомизированную прогнозирующую систему $Pr_n(\cdot)$ такую, что для произвольной бинарной последовательности $\omega_1\omega_2\ldots$ для любого s и для характеристической функции I произвольного подинтервала $[0, 1]$ Pr-вероятность события

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}I(\tilde{p}_i)\delta_s(\omega^{i-1})(\omega_i - \tilde{p}_i)\right| \leq \Delta \tag{14}$$

равна 1, где Pr-распределение на $[0, 1]^\infty$, порожденное последовательностью распределений $Pr_{\omega^{n-1}}(\cdot) = f(\omega_1\ldots\omega_{n-1})$, $n = 1, 2, \ldots$, последовательность прогнозов $\tilde{p}_1, \tilde{p}_2, \ldots$ распределена по мере Pr.

Приводимый ниже вариант универсального рандомизированного алгоритма случайного округления основан на идеях из [10] и [16].

Пусть $\omega_1\omega_1\ldots$ - произвольная последовательность элементов $\{0, 1\}$, поступающая в режиме онлайн. Построим алгоритм для вычисления случайной величины, выдающей прогноз $p_n \in [0, 1]$ будущего значения $\omega_n \in \{0, 1\}$ по начальному фрагменту $\omega_1\ldots,\omega_{n-1}$. Основное требование к таким прогнозам - они должны с вероятностью 1 удовлетворять условию калибруемости (14).

Соответствующее распределение вероятностей является внутренним по отношению к алгоритму и строится в процессе конструкции.

Предварительно разобъем интервал $[0, 1]$ на равные части длины $\Delta = 1/K$ с помощью рациональных точек $v_i = i\Delta$, $i = 0, 1, \ldots, K$. Пусть V обозначает множество всех этих точек. Любое число $p \in [0, 1]$ представляется в виде линейной комбинации соседних точек подинтервала разбиения, содержащего p,

$$p = \sum_{v\in V} w_v(p)v = w_{v_{i-1}}(p)v_{i-1} + w_{v_i}(p)v_i,$$

где $p \in [v_{i-1}, v_i]$, $i = \lfloor p/\Delta + 1\rfloor$, и

$$w_{v_{i-1}}(p) = 1 - \frac{p - v_{i-1}}{\Delta},\ w_{v_i}(p) = 1 - \frac{v_i - p}{\Delta}.$$

Полагаем $w_v(p) = 0$ для всех остальных значений $v \in V$.

В дальнейшем, детерминированный прогноз p, выдаваемый алгоритмом приведенным далее, будет округляться до v_{i-1} с вероятностью $w_{v_{i-1}}(p)$ и до v_i с вероятностью $w_{v_i}(p)$.

Построим детерминированный алгоритм, выдающий прогноз p. Пусть прогнозы $p_1, \ldots, p_{n-1}$ уже определены (полагаем $p_1 = 0$). Построим прогноз p_n.

Рассмотрим вспомогательную величину

$$\mu_{s,n-1}(v) = \sum_{i=1}^{n-1} w_v(p_i)\delta_s(\omega^{i-1})(\omega_i - p_i), \tag{15}$$

где $n, s = 1, 2, \ldots$. Имеем

$$(\mu_{s,n}(v))^2 = (\mu_{s,n-1}(v))^2 + 2w_v(p_n)\mu_{s,n-1}(v)\delta_s(\omega^{n-1})(\omega_n - p_n) +$$
$$(w_v(p_n))^2(\delta_s(\omega^{n-1}))^2(\omega_n - p_n)^2. \tag{16}$$

Рассмотрим взвешенную сумму величин (16)

$$\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\sum_{v\in V}(\mu_{s,n}(v))^2 = \sum_{s=1}^{\infty}\frac{1}{s(s+1)}\sum_{v\in V}(\mu_{s,n-1}(v))^2 +$$
$$2(\omega_n - p_n)\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\delta_s(\omega^{n-1})\sum_{v\in V}w_v(p_n)\mu_{s,n-1}(v) + \tag{17}$$
$$\sum_{v\in V}(w_v(p_n))^2\sum_{s=1}^{\infty}\frac{1}{s(s+1)}(\delta_s(\omega^{n-1}))^2(\omega_n - p_n)^2. \tag{18}$$

Используя определение (15), пребразуем часть (17)

$$\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\delta_s(\omega^{n-1})\sum_{v\in V}w_v(p)\mu_{s,n-1}(v) =$$
$$\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\sum_{v\in V}w_v(p)\sum_{i=1}^{n-1}w_v(p_i)\delta_s(\omega^{n-1})\delta_s(\omega^{i-1})(\omega_i - p_i) =$$
$$\sum_{i=1}^{n-1}(\sum_{v\in V}w_v(p)w_v(p_i))\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\delta_s(\omega^{n-1})\delta_s(\omega^{i-1})(\omega_i - p_i) =$$
$$\sum_{i=1}^{n-1}(\bar{w}(p), \bar{w}(p_i))\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\delta_s(\omega^{n-1})\delta_s(\omega^{i-1})(\omega_i - p_i) =$$
$$\sum_{i=1}^{n-1}K(p, p_i)\left(\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\delta_s(\omega^{n-1})\delta_s(\omega^{i-1})\right)(\omega_i - p_i),$$

где $\bar{w}(p) = (w_1, \ldots, w_{v_K}) = (0, \ldots, w_{v_{i-1}}(p), w_{v_i}(p), \ldots, 0)$ - вектор вероятностей округления, $p \in [v_{i-1}, v_i]$, и скалярное произведение

$$K(p, p_i) = (\bar{w}(p), \bar{w}(p_i)) \tag{19}$$

соответствующих векторов (ядро). По определению - $K(p, p_i)$ - непрерывная функция (точнее, кусочно-линейная).

Второй член правой части суммы (18) при подходящем значении p_n всегда можно сделать меньшим или равным нулю. Действительно, в качестве p_n берем какой-нибудь корень $p_n = p$ уравнения

$$\sum_{v\in V}w_v(p)\mu_{s,n-1}(v) = \sum_{i=1}^{n-1}K(p, p_i)\left(\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\delta_s(\omega^{n-1})\delta_s(\omega^{i-1})\right)(\omega_i - p_i) = 0, \tag{20}$$

если он существует. В противном случае, если левая часть уравнения (20) (которая является непрерывной по p функцией) больше нуля для всех значений p_n, то полагаем $p_n = 1$, если она меньше нуля, то полагаем $p_n = 0$. Таким образом определенное значение p_n выдаем в качестве детерминированного прогноза.

Третий член (18) ограничен числом 1. Действительно, так как $|\omega_i - p_i| \leq 1$ для всех i и $\sum\limits_{s=1}^{\infty}\frac{1}{s(s+1)}\delta_s(\omega^{n-1}) \leq 1$, имеем

$$\sum_{v\in V}(w_v(p_n))^2\sum_{s=1}^{\infty}\frac{1}{s(s+1)}(\delta_s(\omega^{n-1}))^2(\omega_n - p_n)^2 \leq \sum_{v\in V}w_v(p_n) = 1.$$

Отсюда и по (18), если последовательно выбирать прогнозы p_i согласно указанному правилу, получим для произвольного j

$$\sum_{v\in V}(\mu_{j,n}(v))^2 \le j(j+1)\sum_{s=1}^{\infty}\frac{1}{s(s+1)}\sum_{v\in V}(\mu_{s,n}(v))^2 \le j(j+1)n.$$

Так как сумма квадратов не превосходит n, каждый член этой суммы также не превосходит n. Поэтому для каждого $v \in V$ имеет место неравенство

$$\left|\sum_{i=1}^{n} w_v(p_i)\delta_j(\omega^{i-1})(\omega_i - p_i)\right| \le \sqrt{j(j+1)n}. \tag{21}$$

Фиксируем правило выбора δ_j. Пусть теперь $\tilde{p}_i$ - случайная величина, принимающая значения $v \in V$ с вероятностями $w_v(p_i)$ (на самом деле, для каждого p ненулевыми являются только значения $w_v(p)$ для двух соседних границ подинтервала разбиения, содержащего число p). Пусть также $I(p)$ - характеристическая функция произвольного интервала. Для любого i математическое ожидание случайной величины $I(\tilde{p}_i)\delta_j(\omega^{i-1})(\omega_i - \tilde{p}_i)$ равно

$$E(I(\tilde{p}_i)\delta_j(\omega^{i-1})(\omega_i - \tilde{p}_i)) = \sum_{v\in V} w_v(p_i)I(v)\delta_j(\omega^{i-1})(\omega_i - v). \tag{22}$$

Согласно усиленному мартингальному закону больших чисел, с Pr-вероятностью 1

$$\left|\frac{1}{n}\sum_{i=1}^{n} I(\tilde{p}_i)\delta_j(\omega^{i-1})(\omega_i - \tilde{p}_i) - \frac{1}{n}\sum_{i=1}^{n} E(I(\tilde{p}_i)\delta_j(\omega^{i-1})(\omega_i - \tilde{p}_i))\right| \to 0. \tag{23}$$

при $n \to \infty$.

По определению детерминированного прогноза p_i и функции $w_v(p)$

$$\left|\sum_{v\in V} w_v(p_i)I(v)\delta_j(\omega^{i-1})(\omega_i - v) - \sum_{v\in V} w_v(p_i)I(v)\delta_j(\omega^{i-1})(\omega_i - p_i)\right| < \Delta \tag{24}$$

для каждого i.

Так как (21) имеет место для каждого $v \in V$, суммируя (24) по $i = 1,\dots,n$, получаем верхнюю оценку для абсолютной величины суммы математических ожиданий (22)

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i)\delta_j(\omega^{i-1})(\omega_i - \tilde{p}_i))\right| =$$

$$\left|\sum_{i=1}^{n}\sum_{v\in V} w_v(p_i)I(v)\delta_j(\omega^{i-1})(\omega_i - v)\right| < \Delta n + K\sqrt{j(j+1)n} \tag{25}$$

для всех n, где $K = 1/\Delta$ - число подинтервалов разбиения.

Из (25) и (23) получаем, что с Pr-вероятностью 1

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n} I(\tilde{p}_i)\delta_j(\omega^{i-1})(\omega_i - \tilde{p}_i)\right| \le \Delta. \tag{26}$$

Для доказательства (1), в приведенном выше алгоритме нужно закон больших чисел (23) заменить на неравенство Хёфдинга - Азумы [17]

$$Pr\{|S_n| > t\} \le 2e^{-\frac{2t^2}{n}},$$

где $S_n = \sum\limits_{i=1}^{n} V_i$,

$$V_i = I(\tilde{p}_i)\delta_j(\omega^{i-1})(\omega_i - \tilde{p}_i) - E(I(\tilde{p}_i)\delta_j(\omega^{i-1})(\omega_i - \tilde{p}_i))$$

- последовательность мартингал-разностей. Это неравенство дает экспоненциальную оценку скорости сходимости в законе (23). После этого, можно выбрать последовательность $\Delta_i \to 0$ и менять время от времени в приведенном выше алгоритме Δ_i на Δ_{i+1} для подходящих достаточно больших n. Так как ряд полученных экспоненциальных оценок сходится, лемма Бореля - Кантелли позволит нам утверждать о сходимости (1) почти всюду.

Список литературы

[1] Vladimir V'yugin. On Sequences with Non-learnable Subsequences // CSR 2008 (E.. Hirsh,, A.Razborov, and A.Semenov, A.Slissenko (Eds.)), LNCS, V.5010, P.302-313. Springer Verlag, Berlin - Heildeberg 2008.

[2] Успенский В.А., Семенов А.Л., Шень А.Х. Может ли (индивидуальная) последовательность нулей и единиц быть случайной? // Успехи математических наук. 1990. Т.45. Вып.1. С.105-162.

[3] Li M., Vitányi P. An Introduction to Kolmogorov Complexity and Its Applications, 2nd ed. New York: Springer–Verlag, 1997.

[4] Dawid, A.P.: The Well-Calibrated Bayesian [with discussion] // J. Am. Statist. Assoc. 1982. V.77. P.605-613.

[5] Dawid, A.P.: Calibration-Based Empirical Probability [with discussion] // Ann. Statist. 1985. V.13. P.1251-1285.

[6] Oakes, D.: Self-Calibrating Priors Do not Exists [with discussion] // J. Am. Statist. Assoc. 1985. V.80. P.339-342.

[7] Schervish, V. Comment [to Oakes, 1985] // J. Am. Statist. Assoc. 1985. V.80. P.341-342.

[8] Ширяев А.Н. Вероятность. М.: Наука, 1980.

[9] Foster, D.P.,Vohra, R. Asymptotic Calibration // Biometrika 1998. V.85. P.379-390.

[10] Kakade, S.M., Foster, D.P. Deterministic Calibration and Nash Equilibrium // In John Shawe Taylor and Yoram Singer, editors Proceedings of the Seventeenth Annual Conference on Learning Theory Volume 3120 of Lecture Notes in Computer Science. 2004. P.33-48. Heidelberg, Springer.

[11] Lehrer, E. Any Inspection Rule is Manipulable // Econometrica. 2001. V.69. P.1333-1347.

[12] Sandroni, A., Smorodinsky R., Vohra, R. // Calibration with Many Checking Rules // Mathematics of Operations Research 2003. V.28. P.141-153.

[13] Роджерс Х. Теория рекурсивных функций и эффективная вычислимость. М.: Мир., 1972.

[14] Звонкин А.К., Левин Л.А. Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов // Успехи математических наук. 1970. Т.25. Вып.6. С.85-127.

[15] V'yugin, V.V. Non-Stochastic Infinite and Finite Sequences // Theor. Comp. Science. 1998. V.207. P.363-382.

[16] Vovk, V. Defensive Forecasting for Optimal Prediction with Expert Advice // arXiv:0708.1503v1. 2007.

[17] Cesa-Bianchi, N, Lugosi G. Prediction, Learning and Games. Cambridge University Press, 2006.