[В этом файле собраны математические (в основном) фрагменты писем Ан. А. Мучника. Основным адресатом является Marcus Hutter, хотя письма направлялись А. Чернову для последующего пересказа Хуттеру на английском. Хуттер также получал оригиналы на русском и переводил их с помощью babelfish. Ответные письма Хуттера отсутствуют.]

# Письмо 1

## 1st

Я очень рад, что Маркус готов к сотрудничеству. Пишет он хорошо, Вы можете подтвердить ему, что я внимательно читаю тексты, относящиеся к моей работе. Я надеюсь, что Хуттер изложит результаты, обобщив их по возможности (не только для равномерной меры, ведь задача предсказания осмысленна, если мы заранее не знаем, какую меру предсказываем). Попросите, пожалуйста, Маркуса писать мое имя Andrej, а не Andrei, и в названии Института Новых Технологий убрать окончание of education.

Теперь несколько замечаний. Когда употребляется выражение "with probability 1"или аналогичное ему, стоит указать, что речь идет о вероятности $\mu$ (а не $M$, например). Естественно, можно написать один раз, как понимается это выражение во всем тексте или в данном разделе текста.

Правильно ли я понял, что Хуттер предпочитает работать с расстоянием Хеллингера $\rho$ (квадрат разности корней), а не с квадратом разности, как у Соломонова, Ли и Витаньи?

К вопросу о скорости сходимости. Теорему Соломонова (5.2.1 у Ли и Витаньи) можно усилить. А именно, конечен не только интеграл по мере $\mu$ суммы вдоль последовательности $x$ $\rho$ - расстояний от $\mu(\cdot|x_{<n})$ до $M(\cdot|x_{<n})$, но и интеграл экспоненты от половины этой суммы. Это доказывается тем же приемом, который употреблен в статье Вовка "On a randomness criterion. Soviet Mathematics Doklady 35, 656-660 (1987)". Рассматривается мера $R$, у которой условная вероятность перехода из вершины v в вершину $va$, равна $\frac{\sqrt{\mu(a|v)\cdot M(a|v)}}{\sum_{b\in B}\sqrt{\mu(b|v)\cdot M(b|v)}}$. Сумма значений $R(y)$ по всем $y$ длины $n$ равна 1. Представим $R(y)$ как произведение $R(y_k|y_{<k})$, подставим значения этих условных вероятностей из определения R, в полученном произведении соберем произведения условных

вероятностей для $\mu$ и для M обратно в их значения в y, получится $\mu(y) \cdot \sqrt{M(y)/\mu(y)}/\prod_{k\le n}\sum_{b\in B}\sqrt{\mu(b|y_{<k})\cdot M(b|y_{<k})}$. Остается отметить, что $M(y)/\mu(y) \ge w_\mu$ и $\sum_{b\in B}\sqrt{\mu(b|v)\cdot M(b|v)} \le 1-\frac{1}{2}\rho(\mu(\cdot|v),M(\cdot|v)) \le e^{-\frac{1}{2}\rho(\mu(\cdot|v),M(\cdot|v))}$. Итак, интересующий нас интеграл не превышает $w_\mu^{-1/2}$ ($w_\mu$ — обозначение из текста Маркуса).

Мне представляется, что задача нахождения перечислимой снизу полумеры, которая является универсальным предсказателем, интересна, даже если эта полумера не априорна. Мы построим такую полумеру.

Еще я рассматривал перечислимые снизу полумеры $M'$, у которых значение в корне может быть меньше 1, но в каждой вершине v, так же, как и у меры, выполнено равенство $M'(v) = \sum_b M'(vb)$. Такие полумеры, кажется, ввел Шнор (Шень в шутку назвал их 3/4-мерами). Среди них есть полумера, случайность относительно которой совпадает со случайностью по Мартин-Лефу. Но максимальной полумеры в этом классе нет. Заметим, что в первом примере априорной полумеры, которая не является предсказателем, существенно, что для нее не выполнено $M'(v) = \sum_b M'(vb)$. Мы построим другой пример, более наглядный и очевидно распространяющийся на полумеры Шнора.

Из теоремы Соломонова следует, что каждая априорная полумера правильно в пределе предсказывает на всех $0'$-случайных последовательностях. Мы покажем, что последовательность, на которой правильно в пределе предсказывает каждая априорная полумера, может не быть $0'$-случайной. Может ли она не быть просто случайной, я не знаю, но ответ "может если не требовать скорости сходимости предсказаний, то есть требовать только $M(b|x_{<n})/\mu(b|x_{<n}) \to 1$. (Это уже третья ошибка в теореме 5.2.2 из книги Ли и Витаньи.)

Три основных открытых (для меня) вопроса: есть ли универсальный предсказатель среди априорных полумер, есть ли универсальный предсказатель среди полумер Шнора, влечет ли быстрая сходимость предсказаний на последовательности ее случайность.

Доказательства объявленных утверждений я собираюсь изложить в следующем письме.

## 2nd

Сначала рассмотрим новое доказательство уже известной теоремы о существовании априорной полумеры $M'$ и случайной (относительно равномерной меры $\mu$) последовательности x, на которой $M'$ плохо предсказывает $\mu$. Пусть M — какая-нибудь априорная полумера и x — самая левая последовательность, на которой при всех n $M(x_{<n}) \leq 2\mu(x_{<n})$.

Построим вспомогательную перечислимую снизу полумеру $\nu$ (она будет полумерой Шнора). Когда про какую-нибудь вершину u обнаруживается, что она строго левее x, мы делаем полумеру $\nu$ равной $\mu$ в конусе над u. С каждой вершиной v свяжем полумеру $\nu_v$, которая с конусе над v совпадает с $\mu$, а в стороне от v равна 0 (таким образом, в корне $\nu$ равно $\mu(v)$). Ясно, что $\nu$ может быть представлена как сумма $\nu_v$ для некоторой вычислимой последовательности вершин v, конуса над которыми попарно не пересекаются и в объединении дают $\{y : y \text{ строго левее } x\}$. Значение $\nu$ в корне равно равномерной мере последнего множества. Поскольку все $\nu_v$ являются полумерами Шнора, то и их сумма тоже полумера Шнора. Искомая априорная полумера $M'$ определяется как $\frac{9}{10}\nu + \frac{1}{10}M$. Так как последовательность x случайна, то в нее бесконечно много раз входит сочетание $< l, r >$ (налево, направо). Пусть такое сочетание следует за $x_{<n}$.

Оценим снизу $M'(l|x_{<n})$. Так как $x_n = l$, а $x_{n+1} = r$, то число $\alpha$, равное $\nu(x_{<n}) = \nu(x_{\leq n})$, лежит в $(2^{-n-1}, 2^{-n})$. Далее $\frac{\frac{9}{10}\nu(x_{\leq n}) + \frac{1}{10}M(x_{\leq n})}{\frac{9}{10}\nu(x_{<n}) + \frac{1}{10}M(x_{<n})}$ больше $9\alpha/(9\alpha + 2\mu(x_{\leq n})) = 1/(1 + 2^{-n+1}/9\alpha) > 1/(1 + 2^{-n+1}/(9 \cdot 2^{-n-1})) = 9/13 > 1/2$. С другой стороны, $\mu(l|x_{<n}) = 1/2$. Понятно, что у последовательности $M'(x_n|x_{<n})$ вообще нет предела. Если бы предел был строго больше $1/2$, то отношение $M'(x_{<n}/\mu(x_{<n})$, представленное как произведение отношений условных вероятностей, стремилось бы к бесконечности вопреки случайности x. Осталось отметить, что если в качестве M взять полумеру Шнора, то и $M'$ станет полумерой Шнора.

Последовательность x из предыдущего доказательства не только не $0'$-случайна, но даже $0'$-вычислима. Теперь построим другую случайную $0'$-вычислимую последовательность y, на которой априорная полумера M (произвольная, но фиксированная) быстро сходится к равномерной мере. Если $y_{<n}$ уже построено, то мы хотели бы выбрать $y_n$ так, чтобы отношение $M(y_n|y_{<n})/\mu(y_n|y_{<n})$ было минимальным. С помощью оракула $0'$ можно найти $y_n$, для которого $M(y_n|y_{<n})/\mu(y_n|y_{<n})$ меньше минимума, умноженного на $1 + 2^{-n}$. Так как минимум не больше 1, то отношение $M(y_{\leq n})/\mu(y_{\leq n})$ меньше произведения $1 + 2^{-n}$ и ограничено, а последовательность y случайна. Рассмотрим расстояние Кульбака $\sum_b \mu(b|y_{<n}) \cdot \log(\frac{\mu(b|y_{<n})}{M(b|y_{<n})})$. Оно не больше максимума $\log(\frac{\mu(b|y_{<n})}{M(b|y_{<n})})$, который

меньше $\log(\frac{\mu(y_n|y_{<n})}{M(y_n|y_{<n})} \cdot (1+2^{-n}))$. Тем самым, сумма первых n расстояний Кульбака меньше $\log(\mu(y_{\le n})/M(y_{\le n})) + \sum_k \log(1+2^{-k}) < \infty$.

Как обеспечить, чтобы на последовательности *y любая* априорная полумера быстро сходилась к равномерной мере, и как связаны расстояния Кульбака и Хеллингера, обсудим в следующий раз.

# Письмо 2

Я много думал над вопросом о скорости сходимости предсказаний, который мы выделили как начальный. Я пришел к выводу, что это может быть достаточно сложный вопрос. В настоящем письме я собираюсь рассказать Вам (и нашим швейцарским коллегам) о том, что нам удалось придумать. Буду рад, если у Вас возникнут новые идеи.

Теперь краткий обзор. Я посмотрел в книгу Ширяева. Кажется, прием из статьи Вовка, дающий коэффициент $1/2$, как мы и предполагали, взят из классической теории вероятностей. Оптимален ли этот коэффициент? При прогнозировании пары мер: равномерной $\mu_0$ и бернуллиевской $\mu_\epsilon$, немного сдвинутой по отношению к равномерной, — расходимость интеграла одной из двух величин $e^{\alpha \cdot \sum \rho(\mu_0,\nu)}$ и $e^{\alpha \cdot \sum \rho(\mu_\epsilon,\nu)}$ можно гарантировать, когда $\alpha > 2$ (где $\nu$ — мера, используемая для прогнозирования, а $\rho$ — расстояние Хеллингера между распределениями условных вероятностей в данной вершине). Коэффициет 2 меньше, чем 10 из статьи Вовка, но в четыре раза больше коэффициента $1/2$, при котором можно обеспечить сходимость обоих интегралов. С другой стороны, если мы хотим прогнозировать любые две меры, то сходимость можно обеспечить, когда $\alpha < 1$. Поэтому точность $1/2$ нельзя доказать (если она вообще имеет место), прогнозируя только две меры. Еще стоит отметить, что способ прогнозирования двух мер, который я собираюсь изложить, всегда дает распределение условных вероятностей, совпадающее с тем, которое у первой или у второй меры (но в разных вершинах, естественно, по-разному). Для получения $1/2$ берется, как Вы помните, $\nu$, равное среднему арифметическому прогнозируемых мер.

Перейдем к математическим подробностям. Мы начнем с соображений, которые формально не требуются, но помогают понять происхождение метода, как нашего, так и метода Ширяева. Семенов предложил начать как бы с конца. То, что для какого-то $\alpha$ есть способ прогнозирования, можно интерпретировать как то, что есть стратегия, которая, получив от противника распределения условных вероятностей прогнозируемых мер в текущей вершине, выдает распределение условных вероятностей прогнозирующей меры. Победа стратегии означает, что оба интеграла ограничены сверху константой, не зависящей от поведения противника. Удобно рассмотреть множество тех пар констант, для которых есть стратегия, обеспечивающая, что первый интеграл меньше первой константы, а второй интеграл меньше второй константы. Затем удобно для каждой первой константы рассмотреть инфимум тех вторых констант, для которых в паре с первой есть стратегия. Интересующие нас константы принимают значения от 1 до $\infty$ (поскольку $\rho \geq 0$ и $e$ в неотрицательной степени не меньше 1). Когда первая константа растет, то вторая, очевидно, убывает. Если вторую константу назвать первой, то в качестве второй ей будет соответствовать прежняя первая константа (так как наша игра полностью симметрична относительно номера прогнозируемой меры). Равенство первой константы 1 означает, что прогнозирующая мера всегда совпадает с первой прогнозируемой; после чего просто подобрать вторую прогнозируемую меру так, чтобы второй интеграл стал бесконечен. Итак, мы предполагаем, что для $\alpha$ существует монотонно убывающая функция $f \geq 0$, для которой первой константе $1+x$ соответствует вторая константа $1+f(x)$. Причем $f \to 0$ при $x \to \infty$ и $\forall x\ f(f(x)) = x$. Этим условиям удовлетворяют функции $f(x) = C/x$. Конечно, есть и другие, но вероятно, они сложнее. Понятно, что выигрывающая стратегия будет построена, если мы сможем поддержать инвариант (связанный с константами) на одном шаге игры. Пусть мы находимся в некоторой вершине, в которой первая мера имеет распределение вероятностей $p_i$, а вторая мера — $q_i$, и дано положительное $x$. Наша цель каждому направлению $i$ сопоставить условную вероятность $r_i$ для прогнозирующей меры и положительное $y_i$. Теперь предполагая по индукции, что для каждого $i$ после прохода в $i$-м направлении мы имеем стратегию, выигрывающую припаре констант $1+y_i$, $1+C/y_i$, мы можем оценить сверху интегралы в исходной вершине.

Оценка на первый интеграл: $[\sum_i p_i(1+y_i)] \cdot e^{\alpha\rho(p,r)}$. Оценка на второй интеграл: $[\sum_i q_i(1+C/y_i)] \cdot e^{\alpha\rho(q,r)}$.

Мы хотим, чтобы первая оценка оказалась меньше $1+x$, а вторая меньше $1+C/x$. Без ограничения общности предположим, что $x \leq C/x$. В этом случае положим $r = p$ (иначе $r = q$). Теперь наша цель доказать,что можно так подобрать значения $y_i$, что $1 + \sum p_i y_i = 1 + x$ (то есть $x = \sum p_i y_i$), и $[1 + \sum q_i(1 + C/y_i)] \cdot e^{\alpha\rho(p,q)} \leq 1 + C/x$. Другими словами, мы хотим минимизировать величину $[1+\sum q_i(1+C/y_i)]\cdot e^{\alpha\rho(p,q)}$ при условии,что $y_i$ дают решение уравнения $x=\sum p_i y_i$. По известному из анализа правилу в точке минимума вектор частных производных по $y_i$ минимизируемой величины должен быть пропорционален вектору частных производных величины, нули которой задают условия. То есть пропорциональными должны быть векторы $p_i$ и $q_i/y_i^2$. Отсюда мы видим, что в точке минимума вектор $y_i$ должен быть пропорционален $\sqrt{q_i/p_i}$. Теперь применим уравнение, котороезадавало условие, и получим, что $y_i=\sqrt{q_i/p_i}\cdot x/\sum\sqrt{q_i\cdot p_i}$. Теперь вспомним, что $\sqrt{q_i\cdot p_i}$ есть $1-(1/2)\rho(p,q)$. Подставим найденные $y_i$ в минимизируемую величину. Получим $[1+(C/x)(\sum\sqrt{q_i\cdot p_i})^2]\cdot e^{\alpha\rho(p,q)}$. Теперь можно проверить, что когда $\alpha<1$, а $C$ достаточно велико (тем больше, чем ближе $\alpha$ к 1), минимизируемая величина окажется меньше $1+C/x$.

Еще раз напоминаю, что приведенные выше рассуждения формально не нужны, но они мотивируют то, как мы действовали. Формально же последовательность действий такова: берется $\alpha<1$, по нему подбирается достаточно большое $C$. Строится некоторая стратегия, памятью этой стратегии является число $x$. В начале оно равно $\sqrt{C}$. При переходе в $i$-м направлении $x$ заменяется на $x\cdot\sqrt{q_i/p_i}$, если $x\geq\sqrt{C}$ и заменяется на $x\cdot\sqrt{q_i/p_i}/(\sum\sqrt{q_i\cdot p_i})^2$, если $x\leq\sqrt{C}$. Условные вероятности $r_i$ (для прогнозируюещй меры) равны $p_i$, если $x\leq\sqrt{C}$, и равны $q_i$ в противном случае. Описание стратегии закончено. Мы хотим доказать, что в каждом конусе интеграл для первой меры будет не превышать $1+x$, а интеграл для второй меры будет не превышать $1+C/x$ (где $x$ — память стратегии в основании конуса). Так как все, что нужно, положительно, достаточно проверить интересующее нас свойство для конечных деревьев, а для них можно провестииндукцию по высоте. То, что при больших $C$ индуктивный переход получается, показывают элементарные вычисления.

С добрыми пожеланиями, Андрей Мучник.

# Письмо 3

Построение перечислимой снизу полумеры (назовем ее Г), хорошо предсказывающей каждую вычислимую меру $\mu$ на любой случайной относительно $\mu$ последовательности.

Конструкция Г.

Во-первых, каждая перечислимая снизу полумера эффективно преобразуется в перечислимую снизу полумеру, которая совпадает с исходной, если исходная была мерой. Если же исходная полумера не была мерой, то новая полумера не превышает исходной и для некоторого n новая полумера равна 0 на всех последовательностях длины $> n$ и очень близка к мере на последовательностях длины $\leq n$. Степень близости к мере эффективно зависит от $n$ и от номера исходной полумеры в обычной нумерации (конкретное значение извлекается из доказательства).

Во-вторых, преобразованные полумеры складываются с коэффициентами, очень быстро убывающими с ростом их номера в обычной нумерации (конкретное значение эффективно извлекается из доказательства).

Доказательство.

Лемма 1. Если $P$ - мера, а $Q$ - полумера и $Q > P(1-\epsilon)$, то сумма расстояний Хеллингера (между условными вероятностями относительно $P$ и относительно $Q$) вдоль последовательности после усреднения по $P$ оказывается меньше $2\epsilon$.

Доказательство леммы 1.

Как мы делали раньше, рассмотрим вспомогательную меру $R$. Условная вероятность относительно $R$ перехода из вершины $x$ в вершину $xa$ равна $\sqrt{(}(P(xa)/P(x))\cdot(Q(xa)/Q(x)))/\sum_b\sqrt{(}(P(xb)/P(x))\cdot(Q(xb)/Q(x)))$. Сумма значений $R$ по листьям конечного дерева равна 1. Значение $R$ в листе $y$ равно произведению условных вероятностей $R$, ведущих в этот лист, то есть $P(y)\sqrt{(}Q(y)/P(y))/\prod_{x\prec y}\sum_b\sqrt{(}(P(xb)/P(x))\cdot(Q(xb)/Q(x)))$. Из условия леммы $Q(y)/P(y) > 1-\epsilon$. Таким образом,

$$\sum_y P(y)/\prod_{x\prec y}\sum_b\sqrt{(}(P(xb)/P(x))\cdot(Q(xb)/Q(x))) < 1+\epsilon\,.$$

Поскольку $\sum_y P(y) = 1$, то

$$\sum_y P(y)(1/\prod_{x\prec y}\sum_b\sqrt{(}(P(xb)/P(x))\cdot(Q(xb)/Q(x)))-1) < \epsilon\,.$$

(Именно в последнем переходе мы использовали то, что $P$ — мера; именно для этого нам было нужно, чтобы предсказывающая полумера была суммой мер или полумер, близких к мерам.) Как Вы помните,

$$1/\prod_{x\prec y}\sum_b\sqrt{(}(P(xb)/P(x))\cdot(Q(xb)/Q(x))) > e^{(1/2)\sum_{x\prec y}\rho_x(P,Q)}\,,$$

где $\rho_x(P,Q)$ — расстояние Хеллингера между условными вероятностями по $P$ и по $Q$ в вершине $x$. В свою очередь $(e^{(1/2)\sum_{x\prec y}\rho_x(P,Q)}-1)$ больше $(1/2)\sum_{x\prec y}\rho_x(P,Q)$. Получаем $\sum_y P(y)(\sum_{x\prec y}\rho_x(P,Q)) < 2\epsilon$.

Лемма 2. Расстояние Хеллингера между распределениями $p$ и $r$ не превышает удвоенной суммы расстояний Хеллингера между $p$ и $q$ и между $q$ и $r$. Доказательство очевидно. То же самое неравенство выполнено для суммы расстояний Хеллингера вдоль последовательности в дереве.

Лемма 3. Пусть $\mu$ - вычислимая мера, а $\nu$ - полумера из суммы, образующей Г. Для случайной относительно $\mu$ последовательности $Y$ отношение $\nu/\mu$ во всех началах $Y$ очень быстро убывает с ростом номера $\nu$ ($\mu$ и $Y$ фиксированы, а от начала $Y$ оценка не зависит).

Доказательство леммы 3.

Пусть $\nu'$ — полумера, из которой получилась $\nu$ (то есть, какой она была до преобразования в нашей конструкции и до умножения на коэффициент). Обозначим номер $\nu$ через $k$. Тогда $\nu'$ входит в сумму для обычной априорной полумеры $M$ с коэффициентом (например) $1/k^2$. Так как $Y$ случайна относительно $\mu$, то во всех началах $Y$ имеем $\mu > c\cdot M > c\nu'/k^2$. Следовательно, $\nu' < \mu\cdot k^2/c$ и $\nu < \epsilon_k\cdot\mu\cdot k^2/c$, где $\epsilon_k$ — $k$-й коэффициент в сумме, образующей Г. Остается вспомнить, что $\epsilon_k$ убывает очень быстро, гораздо быстрее $1/k^2$.

Лемма 4. Если $\mu$ - мера, $\nu$ - полумера и $\nu(x)/\mu(x) < \epsilon$, то расстояние Хеллингера в вершине $x$ между $\mu$ и $\mu+\nu$, меньше $\epsilon$. Доказательство очевидно.

Рассмотрим вычислимую меру $\mu$ и случайную относительно нее последовательность $Y$. По номеру $\mu$ в обычной нумерации и деффекту случайности $Y$ мы оценим сверху сумму вдоль $Y$ расстояний Хеллингера между условными вероятностями по $\mu$ и по Г. Из соображений непрерывности достаточно оценить эту сумму для конечного начала $Y$ (назовем его $y$), а также заменив полумеру Г на сумму конечного множества слагаемых, из которых состоит Г (сумму первых $k$ слагаемых назовем $\gamma_k$).

Оценим сверху сумму расстояний Хеллингера вдоль $y$ между $\gamma_k$ и $\gamma_{k-1}$ для $k$, больших номера $\mu$. Расставим на $y$ точки (если они есть), где „обрываются“ полумеры с номерами $\leq k$. Таких точек не больше $k$. По леммам 3 и 4 в каждой из этих точек расстояние между $\gamma_k$ и $\gamma_{k-1}$ очень мало ($<\epsilon_k\cdot\mu\cdot k^2/c$). Так что сумма расстояний по отмеченным точкам тоже очень мала ($<\epsilon_k\cdot\mu\cdot k^3/c$). Теперь докажем, что интересующая нас сумма расстояний на каждом отрезке между двумя соседними отмеченными точками не превышает $z = \sqrt{(\epsilon_k)}$ (тогда по всем отрезкам она не превышает $k \cdot \sqrt{(\epsilon_k)}$). Рассмотрим один такой отрезок. Пусть $J$ - множество тех номеров $j \leq k$, для которых $j$-я полумера почти совпадает с мерой на рассматриваемом отрезке. Можно считать, что $k \in J$, иначе все расстояния на отрезке нулевые. Пусть $\delta$ - сумма полумер из $\Gamma$ с номерами из $J$, а $\delta' = \delta \setminus \{k\}$. Понятно, что интересующая нас сумма по отрезку не превышает суммы расстояний между $\delta$ и $\delta'$ вдоль начала $y$, на котором определены „почти меры“ с номерами из $J$. Поскольку $\delta$ и $\delta'$ - „почти меры“, то по лемме 1 усреднение по $\delta'$ последней суммы расстояний не превышает $3\epsilon_k$. Так как номер $\mu$ (обозначим его $i$) принадлежит $J$, то среднее по $\mu$ последней суммы расстояний не превышает $3\epsilon_k/\epsilon_i$ (тоже очень малое число). Мера $\mu$ множества последовательностей, для которых последняя сумма больше $z$, не превышает $(3\epsilon_k/\epsilon_i)/z$. Кроме того, это множество перечислимо при известном $J$. Для задания $J$ достаточно $k$ битов; дефект случайности, возникающий из оценки на меру $\mu$ рассматриваемого множества, гораздо больше. Это противоречит случайности $Y$.

Теперь индукцией по $j$ будем оценивать сумму расстояний Хеллингера вдоль $y$ между $\gamma_k$ и $\gamma_{k-j}$ для $k - j \leq$ номера $\mu$. Последовательно применяем лемму 2. При этом расстояние между $\gamma_k$ и $\gamma_{k-1}$ умножается на $2^j$, расстояние между $\gamma_{k-1}$ и $\gamma_{k-2}$ умножается на $2^{j-1}$, и так далее. Множитель $2^{k-j}$ (который получится в конце) гораздо меньше $\epsilon_{k-j}$.

Сумма расстояний Хеллингера между $\gamma_i$ и $\mu$ оценивается сверху так же, как и раньше, разбиением $y$ на отрезки, только теперь $i$ (номер $\mu$) фиксировано, и применение леммы 4 для точек „обрывания“ становится не нужной.

Расстояние между $\mu$ и $\gamma_k$ оценивается через лемму 2.

Конец доказательства.

# Письмо 4

Часть 1.

7-го апреля Маркус прислал мне письмо, в котором опровергается одно из утверждений в книге Ли и Витаньи. А именно, Маркус доказывает, что среднее по мере $\mu$ величины $(\mu(x_t|x_{<t}) - M(x_t|x_{<t}))^2$ не убывает, как утверждают Ли и Витаньи, быстрее $1/t$, а наоборот, оценивается снизу

априорной полумерой $t$ (как натурального числа). При этом доказательство Маркуса годится для любой (не только вычислимой) меры $\mu$. Но надо учесть, что в этом доказательстве существенно используется определение Ли и Витаньи априорной полумеры на дереве $M$ как взвешенной суммы всех перечислимых снизу полумер, а не более общее определение $M$ как перечислимой снизу полумеры, мажорирующей с точностью до мультипликативной константы все перечислимые снизу полумеры. Кроме того, Ли и Витаньи, непонятно почему, в отличие от других авторов (начиная с Соломонова) используют квадрат разности, а не квадрат разности корней, что было бы гораздо естественнее. Можно ли распространить рассуждение Маркуса на расстояние Хеллингера?

Часть 3.

Вы помните, что когда я начал думать над тематикой Хуттера, меня в первую очередь заинтересовал вопрос о том, какое качество сходимости можно обеспечить, предсказывая меру, принадлежащую заранее известному множеству мер. Эта проблема по существу принадлежит классической теории вероятностей. Именно ее в 50-е годы начал решать Соломонов, именно она в 60-е годы привела его к понятию априорной полумеры. Соломонов более или менее сразу предположил, что хорошим, а возможно, и оптимальным предсказателем является среднее арифметическое мер из заранее известного множества. Однако первую оценку он получил только в 70-е годы, и мы знаем, что эта оценка была далека от оптимальной. В 80-е годы, работая у Семенова, этим занимался Вовк. Он все излагал на языке колмогоровской сложности. В математическом отношении это было несущественно, но затрудняло понимание (по крайней мере, для меня). Вовк также рассматривал только предсказатель Соломонова.

Поразительно, но этот предсказатель не всегда оптимален, как показывает теорема 2 (замечание 1) из следующей части. Как показывает теорема 1 из следующей части, предсказатель, о котором мы с Семеновым сделали доклад на семинаре Ширяева (я писал Вам о нем), является оптимальным, когда предсказываются две меры. Хотелось бы наконец ответить на первоначальный вопрос Соломонова и найти оптимальный предсказатель для $n$ мер ($n > 2$). Этот вопрос имеет и философское значение (мерам соответствуют научные теории, претендующие на объяснение какого-то явления, предсказателю соответствует поведение исследователя), и практическое значение.

Как Вы помните, в верхней оценке возникает коэффициент $1/2$ в показателе экспоненты, и он не зависит от количества "предсказываемых"мер (причем в качестве предсказателя используется среднее арифметическое, по Соломонову). Мы с Семеновым доказали, что для предсказателя Соломонова этот коэффициент нельзя увеличить, если количество "предсказываемых"мер неограничено (теорема 2 в следующей части письма). Было бы интересно получить точную оценку коэффициента для фиксированного количества мер. Напомню, что для получения точной оценки 1 в случае двух мер мы использовали не предсказатель Соломонова, а некоторый новый предсказатель, который всегда совпадает с одной из двух мер (в смысле условных вероятностей).

Как я уже писал Вам, мы с Семеновым сделали доклад на семинаре Ширяева. По мнению участников, результаты новые. Еще мы собираемся спросить у Вовка. Можно написать Соломонову (в прошлом году он сделал доклад на конференции по этой теме). По моему мнению было бы разумно разделить совместную с Хуттером работу на две публикации. Одна "алгоритмическая все, что относится к перечислимым снизу полумерам. Соответствующий текст Хуттера и Мучника можно надеяться успеть представить на ALT. Другая публикация (Хуттера, Мучника и Семенова) - "классическая над которой еще надо поработать. Конечно, обе публикации могут ссылаться друг на друга. Впрочем, по Вашим письмам у меня сложилось впечатление, что, возможно, Маркуса больше интересует "алгоритмическая"часть проблематики и не очень интересует "классическая"часть. Если это действительно так, то я совсем не хочу уводить Маркуса от его основных интересов. В этом случае мы с Семеновым постараемся сами завершить "классическую"часть.

Часть 4.

Когда мы доказывали верхнюю оценку для коэффициента в показателе экспоненты, я исходил из того, что пример можно получить в одном из двух крайних случаев: равномерная мера и ее небольшой сдвиг или мера, сосредоточенная на одной последовательности, и ее небольшой сдвиг. Почему-то мне показалось, что скорее оценку даст равномерная мера (она дала оценку 2). Оказалось же, что дискретная $\delta$-мера дает оценку 1 (как мы знаем, точную для двух мер).

Теорема 1. Для каждого $\epsilon > 0$ есть две меры, для которых любой предсказатель дает расходимость одного из двух интегралов при коэффициенте в показателе экспоненты больше $1 - \epsilon$.

Доказательство. В качестве этих мер можно взять меру $\mu_1$, полностью сосредоточенную на последовательности 000... и меру $\mu_2$, у которой условная вероятность нуля всегда равна $1-\epsilon$. Для любого предсказателя $\nu$ сходимость интеграла, соответствующего $\mu_1$, эквивалентна тому, что $\sum_i \nu(0^i 1|0^i)$ сходится. Следовательно, $\nu(0^i 1|0^i)$ стремится к нулю. Это позволяет оценить снизу сумму в показателе экспоненты для интеграла, соответствующего $\mu_2$. При коэффициенте $> 1-\epsilon$ получается расходимость интеграла (даже если интегрировать только по конусам над $0^i$). Замечание. Мы знаем, что меры, у которых условные вероятности не отделены от 0 (как у $\mu_1$), могут иметь необычные свойства (это отмечено и в статье Вовка 1988 года). Искомый пример можно построить, рассмотрев в качестве $\mu_1$ меру, у которой вероятность нуля всегда равна $1-\gamma$, где $0 < \gamma << \epsilon$. Для доказательства оценим снизу взвешенную сумму двух интегралов по конусам над $0^i$. Пусть интеграл, соответствующий $\mu_1$, берется с весом $\alpha > 0$, а интеграл, соответствующий $\mu_2$, берется с весом $\beta > 0$, $\alpha + \beta = 1$. Пользуясь выпуклостью экспоненты, имеем $\alpha * x + \beta * y > exp(\alpha \ln(x) + \beta \ln(y))$. При $\beta$, близком к 1, нужная оценка получается прямым подсчетом.

Теорема 2. Для предсказателя Соломонова (среднее арифметическое предсказываемых мер) коэффициент 1/2 в показателе экспоненты нельзя увеличить, по-крайней мере если количество предсказываемых мер неограничено.

Доказательство. Рассмотрим следующие меры $\mu_1$, ..., $\mu_n$ (нас будет интересовать только их поведение на последовательности $000\ldots$). Пусть $\epsilon > 0$ достаточно мало. Условная вероятность $\mu_i(0^{k+1}|0^k)$ равна 1 при $k \neq i(mod n)$ и $\mu_i(0^{k+1}|0^k)$ равна $1-\epsilon$ при $k = i(mod n)$. Простой подсчет показывает, что для предсказателя Соломонова $\nu$ сумма расстояний Хеллингера между $\mu_i$ и $\nu$ вдоль последовательности $0^L$ примерно равна $2\epsilon * (1 - 1/\sqrt(n)) * (L/n)$. Мера $\mu_i$ последовательности $0^L$ примерно равна $e^( - \epsilon * (L/n))$, что дает нижнюю оценку интеграла по конусу над $0^L$. Замечание 1. Если в качестве предсказателя использовать $\delta$-меру, сосредоточенную в 000..., то сумма расстояний Хеллингера вдоль последовательности $0^L$, станет вдвое меньше - примерно $\epsilon * (L/n)$. Замечание 2. Как и в предыдущей теореме, можно вместо условных вероятностей, равных 1, рассматривать условные вероятности $1-\gamma$, где $0 < \gamma << \epsilon$. Причем, в отличие от предыдущей теоремы, это рассмотрение никак не усложняет доказательство.

Часть 7.

Ромащенко давно интересовался, верен ли аналог одной теоремы о шенноновской энтропии для колмогоровской сложности. Мотивировка следующая. Пусть случайная величина состоит из нескольких корелированных частей. Пусть каждая часть измеряется в точках, достаточно удаленных друг от друга. Цель состоит в том, чтобы измеритель каждой части послал свои результаты в центр, собирающий информацию о совокупной случайной величине. Так как части величины корелированы, то в принципе измеритель может послать только какой-то "фрагмент"своих результатов, лишь бы в центре по всем полученным фрагментам смогли восстановить исходную информацию. Точная формулировка. Пусть даны слова $x_1, ..., x_n$. Для каких чисел $a_1, ..., a_n$ существуют такие слова $y_1, ..., y_n$, что $K(y_i|x_i) = 0$, $K(x_1, ..., x_n|y_1, ..., y_n) = 0$, $K(y_i) < a_i$? Очевидные необходимые условия состоят в том, что при любом разбиении множества индексов 1, ..., n на две части U и V имеют место неравенства $K(x_U|x_V) < \sum_{i \in U} a_i$. Эти условия оказываются достаточными. Когда на колмогоровском семинаре рассказывалось сложное доказательство для n=3, Шень предположил, что индуктивное доказательство для всех n можно получить, применив мою теорему о том, что для слов $b_1$, ..., $b_k$, $d$ есть одна оптимальная программа, переводящая каждое $b_i$ в $d$ и простая относительно $d$. Предположение Шеня подтвердилось, рассуждение совсем простое.

Часть 8.

В этом семестре в Москве Беклемишев, и он читает курс о комбинаторных фактах, недоказуемых в арифметике Пеано. Необычность его подхода в том, что он основан на модальной логике доказуемости. Я был удивлен очень большим интересом, проявленным к этой теме. На первых лекциях в аудитории буквально негде было сесть. При полном изложении (Беклемишев дал мне тексты) доказательства довольно-таки сложны.

Три основные давно известные результата в этой области можно сформулировать так. Для канонической системы обозначений индукция до ординала $\epsilon_0$ позволяет финитно (кроме этой индукции) доказать непротиворечивость PA. Есть примитивно-рекурсивная система обозначений, в которой индукция до $\omega$ позволяет финитно доказать непротиворечивость PA (то же для перечислимых непротиворечивых расширений PA). Для каждого рекурсивного ординала есть такая примитивно-рекурсивная система обозначений, что индукция до этого ординала доказуема в PA (а значит, по теореме Геделя не доказывает непротиворечивость PA). Основная открытая проблема (неформальная) - определить, что такое естественная система обозначений рекурсивного ординала.

Второй из перечисленных фактов сравнительно прост. Мне удалось понять, как можно не слишком сложно доказать первый факт и третий факт. Использованные методы можно назвать классическими (не те, что у Беклемишева). Раньше я не знал не слишком сложных доказательств.

Часть 9.

На семинаре Адяна был доклад о раскрасках графов. Возможно, Вам понравится следующая задача. Ее решение очень простое, но не сразу придумаешь. Почему для произвольного графа количество его раскрасок в n цветов выражается многочленом от n. (От раскраски требуется, чтобы вершины, соединенные ребром, были разноцветными.) Сто лет назад этот многочлен пытались использовать для решения проблемы четырех красок.

# Письмо 5

## Часть 1

Здравствуйте, Алексей!

Письмо Хуттера благополучно дошло до меня, и через кафедральный компьютер, и повторно. Вторая часть настоящего письма непосредственно адресована Хуттеру. Поэтому Вы, пожалуйста, переведите ему устно эту часть (компьютерный перевод может не сохранить оттенков). Математические материалы находятся в третьей и четвертой частях.

[...]

Попросите, пожалуйста, Маркуса вставить в наши тексты упоминание о грантах (желательно, на первой странице): The work was partially supported by Russian Foundation for Basic Research (grants N 04-01-00427, N 02-01-22001) and Council on Grants for Scientific Schools.

[...]

Недавно в Москве один голландец сделал доклад о логике интерпретируемости. Арифметическое предложение B интерпретирует арифметическое предложение C, если в PA+B доказуема непротиворечивость каждого конечно-аксиоматизируемого фрагмента PA+C. Логика с двуместной модальностью „интерпретируемость“ обладает более сложными свойствами, чем логика доказуемости. Например, она разрешима, но не обладает интерполяционным свойством Крейга. Логикой интерпретируемости сейчас много занимаются на западе. По словам голландца, улучшенное изложение результатов из этой области опубликовано в “On the proofs of arithmetical completeness for interpretability logic” by Domenico Zambella, Notre Dame Journal of Formal Logic 1992. Журнала с этим необычным названием нет в университетской библиотеке и нет в компьютере за соответствующий год. Не могли бы Вы сделать мне копию этой статьи, если она доступна в Швейцарии?

## Часть 2

Дорогой Маркус!

Я очень рад, что Вы готовы поработать со мною и Семеновым над „классической“ частью проблематики. Удастся ли нам получить существенно новые результаты, это не так важно. Важно, чтобы эти занятия были интересны всем участникам.

Я понимаю Вас, что оптимизация оценок не всегда является увлекательным занятием (хотя и это иногда нужно, если относиться к математической статье как к произведению искусства). Однако вопрос о величине коэффициента $\alpha$, для которого $E_\mu[e^{\alpha\sum h_t}] < \infty$, — это принципальный вопрос. При переводе на язык колмогоровской сложности на $\alpha$ делится дефект случайности. В колмогоровской теории можно пренебрегать аддитивными, но не мультипликативными константами.

В третьей части письма содержатся математические подробности для статьи Хуттера и Мучника, которую мы посылаем в Падую [*на конференцию ALT-04 — прим. при публикации*], а в четвертой части заготовки для следующей статьи Хуттера, Мучника и Семенова по „классической“ тематике.

Лемму 5 [*см. mlconvx-2.zip; кажется, эта лемма превратилась в теорему 2 в predictor-edited.pdf*] из текста Хуттера и Мучника, посылаемого в Падую, я предлагаю убрать и отложить ее до публикации „классической“ статьи.

Мне по-прежнему кажется, что в теореме 12 [*теорема 8 конференционной версии mlconvx.pdf и теорема 11 журнальной HutterMuchnik2007.pdf*] имеет место эффективная (хотя и грубая) оценка на скорость сходимости. Математические объяснения приведены в третьей части письма (раздел 1). Доказательство асимптотической сходимости Вы написали очень хорошо.

По поводу леммы 14 я думаю, что дополнительные ссылки не нужны.

Ваш вопрос о том, почему мы используем именно расстояние Хеллингера, по-моему, является очень важным. Некоторые понятные мне мотивировки приведены в третьей части письма (раздел 2). Наверное, имеет смысл спросить о том же у специалистов по математической статистике (каковым я все-таки не являюсь).

В 3 разделе третьей части письма доказывается теорема 11 [*теорема 7 конференционной версии mlconvx.pdf и теорема 10 журнальной HutterMuchnik2007.pdf*] (которая в теперешнем тексте приведена без доказательства).

В книге Ли и Витаньи универсальная полумера M определяется как взвешенная сумма всех перечислимых снизу полумер. Это определение более узкое, чем наше определение 2. Для их определения удается оценить снизу расстояние Хеллингера между $M$-предсказаниями и $\mu$-предсказаниями на шаге $t$, где $\mu$ - любая мера (Вы как-то прислали мне аналогичную оценку для других расстояний). Математическое обсуждение находится в 4 разделе третьей части письма. Я думаю, что этот результат естественно включить в нашу теперешнюю статью, так как он относится к универсальной полумере.

Так как до 22 мая (предельного срока конференции в Падуе) времени мало, то Вы пошлите то, что успеете написать. В дальнейшем я внимательно прочту весь текст (к тому же могут появиться замечания рецензентов).

По поводу Вашего вопроса „что будет, если вместо усреднения Соломонова рассмотреть условные вероятности относительно той гипотезы, которая в текущий момент более вероятна", я могу сказать следующее. Асимптотическая сходимость при таком способе предсказания имеет место, но этот способ предсказания не оптимален. Математические объяснения находятся в разделах 1 и 2 четвертой части письма.

## Часть 3

### Раздел 3.1. Эффективность оценки в теореме 12

Пусть $\nu_i$ обрывается на высоте $L_i$. За счет быстрого убывания коэффициентов $\varepsilon_j$ вклад $\nu_j$ в первые $L_i$ слагаемых окончательной суммы пренебрежимо мал для $j > L_i$. То есть на отрезках между точками обрывания, лежащих ниже $L_i$, достаточно рассматривать $\gamma_k$ только для $k <= L_i$. Разность $1 - \nu_i(\epsilon)$ дает вклад во все эти $\gamma_k$. Но поскольку $1 - \nu_i(\epsilon)$ убывает очень быстро с ростом $i$ и с ростом $L_i$, то и суммарный вклад оказывается мал.

### Раздел 3.2. Чем расстояние Хеллингера лучше других

Пусть $P$ и $Q$ — распределения вероятностей на множестве $A$, а S — распределение вероятностей на множестве $B$. Мне кажется, что при разумном определении расстояния $\rho$ между распределениями вероятностей должно быть выполнено $\rho(P,Q) = \rho(P \times S, Q \times S)$, где $\times$ обозначает произведение независимых сомножителей. Действительно, если две гипотезы различаются в информации о части объекта исследования (соответствующей $A$), в остальном же совпадают, то расстояние между этими гипотезами должно определяться по $A$. Независимость распределений на $A$ и на $B$ обеспечивает, что информация о $B$ ничего не сообщает про $A$. Расстояние Хеллингера, расстояние Кульбака, расстояние $\chi^2$ и расстояние в $L_1$ (сумма модулей) указанным свойством обладают, а расстояния в $L_p$ (кроме $p = 1$) и сумма квадратов указанным свойством не обладают. Преимущество расстояния Хеллингера и суммы модулей по сравнению с расстоянием Кульбака и $\chi^2$, с одной стороны, в том, что первые симметричны, а вторые нет, а с другой стороны в том, что первые непрерывны, а вторые разрывны (и даже неограничены около 0). Преимущество расстояния Хеллингера по сравнению с суммой модулей в следующем. Роль небольшого изменения нулевого значения вероятности на $\varepsilon > 0$ гораздо больше роли изменения ненулевого значения вероятности на то же малое $\varepsilon$. Расстояние Хеллингера улавливает эту разницу, а сумма модулей не улавливает.

Конечно, в приведенных выше соображениях есть что-то субъективное.

### Раздел 3.3

Существует такая двоичная последовательность $\alpha_n$, что она неслучайна относительно равномерной меры $\mu$ и в то же время для любой универсальной полумеры $M$ выполнено $M(\alpha_n|\alpha_{<n})/\mu(\alpha_n|\alpha_{<n}) \to 1$.

Рассмотрим быстро возрастающую последовательность чисел $L_i$ и определим вспомогательную меру $\nu$ так: $\nu(1|x_{<n}) = 1/2 + 1/i$ для $L_i <= n < L_{i+1}$. Как мы знаем, для каждого $M$ для почти всех относительно $\nu$ последовательностей будет $M(\alpha_n|\alpha_{<n})/\nu(\alpha_n|\alpha_{<n}) \to 1$. Так как множе-

ство всех (перечислимых снизу) $M$ счетно, то для почти всех относительно $\nu$ последовательностей будет $\forall M M(\alpha_n|\alpha_{<n})/\nu(\alpha_n|\alpha_{<n}) \to 1$. Поскольку $\nu(x_n|x_{<n})/\mu(x_n|x_{<n}) \to 1$, то для почти всех относительно $\nu$ последовательностей будет $\forall M M(\alpha_n|\alpha_{<n})/\nu(\alpha_n|\alpha_{<n}) \to 1$. Кроме того, почти все относительно $\nu$ последовательности случайны относительно $\nu$. Нам годится любая случайная относительно $\nu$ последовательность $\alpha$, для которой $\forall M M(\alpha_n|\alpha_{<n})/\nu(\alpha_n|\alpha_{<n}) \to 1$. Действительно, из быстрого роста $L_i$ следует, что для случайной относительно $\nu$ последовательности $\alpha$, начиная с некоторого $i$, доля единиц среди цифр $\alpha$ на отрезке $[L_i, L_{i+1})$ больше $1/2 + 1/2i$. Из этого (и из быстрого роста $L_i$) в свою очередь следует, что последовательность $\alpha$ неслучайна относительно $\mu$.

## Раздел 3.4. Нижняя оценка расстояния Хеллингера между $M$-предсказаниями и предсказаниями по произвольной мере $\mu$

Пусть $K$ - префиксная сложность, а $M$ - универсальная полумера (в широком смысле). Тогда функция $\phi(x) := M(x) * \sum_{n>=l(x)} 2^{-K(n)}$, очевидно, является перечислимой снизу полумерой. Для дальнейшего рассуждения нам годится любая универсальная полумера $W$ вида $a*M+b*\phi$ (каждая полумера, универсальная в узком смысле, имеет такой вид). Пусть $Z$ - алфавит. Понятно, что $\sum_{z\in Z} W(xz|x) < 1-c2^{-K(l(x))}$ для некоторого $c > 0$ и для любого $x$. Оценка снизу расстояния Хеллингера между $W(xz|x)$ и $\mu(xz|x)$ использует только то, что $\sum_{z\in Z} \mu(xz|x) = 1$. Мы хотим минимизировать выражение $h = 2 - 2 * \sum_i \sqrt{(p_i q_i)}$ при условии, что $\sum p_i = 1$, $\sum q_i = 1 - \varepsilon < 1$. Считаем значения $p_i$ и $\varepsilon$ фиксированными, а значение $q_i$ переменными. Сначала рассматривается случай, когда $p_j = 0$, а $q_j > 0$. В этом случае $j$ удаляется из множества индексов, сумма $p_i$ остается равной 1, а сумма $q_i$ уменьшается. Случай, когда все $q_j$ равны 0, тривиален. Затем рассматривается случай, когда $p_j > 0$, а $q_j = 0$. Простая проверка показывает, что небольшое увеличение $q_j$ и одновременное уменьшение некоторого $q_k$ на ту же величину уменьшает $h$. Следовательно, минимум $h$ достигается, когда все $q_i > 0$. По классическому правилу во внутренней точке экстремума вектор частных производных пропорционален вектору, перпендикулярному гиперплоскости $\sum q_i = 1 - \varepsilon$, то есть вектору $\langle 1, ..., 1\rangle$. Из этого следует, что в точке минимума вектор $q_i$ пропорционален вектору $p_i$. Последнее сразу приводит к нижней оценке на $h$.

Вопрос о возможности аналогичной оценки для полумеры, универсальной в широком смысле, остается открытым.

## Часть 4

### Раздел 4.1 Асимптотическая сходимость для предсказаний по более вероятной гипотезе

Пусть гипотезы задаются мерами $\mu_i$ и на их номерах задано некоторое ненулевое априорное распределение вероятностей. Тогда в каждый момент возникает апостериорное распределение вероятностей на множестве номеров $i$. В качестве предсказания выдаем условные вероятности $\mu_i(\cdot|x)$ для того $i$, которому $x$ (последовательность предыдущих экспериментов) приписывает наибольшую апостериорную вероятность. Пусть $\nu$ — мера, которая задается нашими предсказаниями. Пусть $\mu_j$ — истинная гипотеза. По классической теореме для каждого $i$ для почти всех относительно $\mu_j$ последовательностей х существует предел $\lim \mu_i(x_n)/\mu_j(x_n)$. Рассмотрим те $i$, которые бесконечно много раз встречаются на данной последовательности $x$ в определении $\nu$. Для таких $i$ бесконечно много раз $\mu_i(x_n) >= \mu_j(x_n)$, следовательно, $\lim \mu_i(x_n)/\mu_j(x_n) >= 1$. Отсюда $\lim \mu_i(x_n|x_{<n})/\mu_j(x_n|x_{<n}) = 1$. Остается учесть, что $\nu(x_n|x_{<n}) = \mu_i(x_n|x_{<n})$ для соответствующего $i$.

### Раздел 4.2. Неоптимальность предсказателя из предыдущего раздела

Рассмотрим алфавит $\{0,1\}$ и меры $\mu_1$ и $\mu_2$. Число $\varepsilon > 0$ достаточно мало. Для ненулевых последовательностей $y$ всегда $\mu_1(0|y) = \mu_2(0|y) = 1$. Для нулевых последовательностей $y$ нечетной длины $\mu_1(0|y) = 1$, $\mu_2(0|y) = 1 - \varepsilon$. Для непустых нулевых последовательностей $y$ четной длины $\mu_1(0|y) = 1 - \varepsilon$, $\mu_2(0|y) = 1$. В корне $\mu_1(0) = 1 - 2\varepsilon$, $\mu_2(0) = 1$. На последовательностях из одних нулей мера $\mu_1$ всегда меньше меры $\mu_2$. Если истинная гипотеза — $\mu_1$, то использование в качестве предсказателя $\mu_2$ вдвое менее выгодно, чем всегда с вероятностью 1 предсказывать 0.

# Письмо 6

Я немного посмотрел статьи Хуттера (координаты которых он мне прислал). Как я понимаю, его интересовали разные естественные (с философской или прикладной точки зрения) предсказатели и их свойства. Это интересная и очень обширная область, в которую мне не так просто войти. Сейчас я постараюсь по возможности более отчетливо сформулировать цель нашей с Хуттером и Семеновым следующей статьи, как она мне видится.

Соломонов предложил естественную задачу предсказания и естественный способ ее решения (среднее арифметическое). Есть разные другие способы, более или менее естественные, но есть также вопрос о нахождении оптимального способа. Этот оптимальный способ (или последовательность способов, "сходящихся"к оптимальному) может оказаться странным. Было бы, конечно, хорошо, если бы оптимальный способ нашелся среди естественных. Первоначально я предполагал, что способ Соломонова является оптимальным. То же самое, вероятно, предполагал и Вовк, и даже доказал эту оптимальность для частного случая мер Бернулли (в статье 1989 года).

Разумеется, прежде чем говорить об оптимальности, следует точно сформулировать параметр, который мы хотим оптимизировать. В этом вопросе мысль развивается так. По классической теореме для почти всех относительно меры, соответствующей реальности, последовательностей реальные условные вероятности стремятся к условным вероятностям, предсказываемым методом Соломонова. Это дает основание искать верхнюю оценку суммарного отклонения (суммы потерь) через дефект случайности (относительно реальной меры) рассматриваемой последовательности. Отметим, что если мы разобьем последовательность на длинные независимые друг от друга участки, то и отклонения прогнозов от реальности, и дефект случайности будут суммироваться. Поэтому естественно предположить, что искомая верхняя оценка имеет вид: сумма потерь меньше $\beta*$дефект случайности $+O(1)$. Если рассматриваются вычислимые меры и вычислимая процедура прогнозирования, и функция потерь тоже вычислима, то последнее неравенство эквивалентно конечности интеграла по реальной мере от $2^{\sum \text{loss}/\beta}$. Итак, наша цель найти супремум тех $\alpha$, для которых конечен интеграл от $e^{\alpha * \sum \text{loss}}$. Я вполне согласен с Маркусом в том, что чем шире класс рассматриваемых функций потерь, тем было бы лучше. В первом разделе второй части письма я изложу наш с Семеновым подход, который в принципе может быть применен к любой функции потерь. однако пока удалось применить этот подход только к расстоянию Хеллингера. В первом разделе второй части применение этого подхода проиллюстрировано на доказательстве $\sup\alpha \geq 1/2$. Возникающий при этом способ предсказаний отличается от способа Соломонова, однако оценка на $\alpha$ не улучшается. Во втором разделе второй части письма этот подход применяется для доказательства $\sup\alpha = 1$ для случая двух мер. Для способа Соломонова и, например, для static MDL $\alpha$ нельзя сделать больше $1/2$ даже для двух мер.

# Часть 2

## Раздел 1

Пусть нам дано $n$ мер на дереве $T^*$, среди которых одна (неизвестно, какая) соответствует реальности. Рассмотрим следующее подмножество $S$ в $[1,\infty)^n$. Набор $\langle z_1,\ldots,z_n\rangle$ принадлежит $S$, когда существует предсказатель $\nu$, для которого $\forall\mu_1,\ldots,\mu_n \forall i\, E_i \exp(\alpha\sigma_i(x)) < z_i$, где $E_i$ — математическое ожидание по мере $\mu_i$, $\sigma_i(x)$ — сумма расстояний между условными вероятностями для $\mu_i$ и для $\nu$ вдоль последовательности $x$. Понятно, что если $\langle z_1,...,z_n\rangle \in S$, то для любой вершины $v$ можно найти числа $y_{ij} \in [0,\infty)$ со следующими свойствами. Индекс $i$ меняется от 1 до $n$, индекс $j$ меняется в $T$, $\forall j\, \langle y_{1j},\ldots,y_{nj}\rangle \in S$, $\forall i\, \sum_j y_{ij}\mu_i(vj|v)\exp[\alpha\rho(\mu_i(vj|v),\nu(vj|v))] < z_i$, где $\rho$ — функция потерь. (Отметим, что указанные свойства фактически никак не связаны с вершинами дерева $T^*$.) Наоборот, если найдено множество $S$ с указанными свойствами, то мы получаем некоторый предсказатель (для данного значения $\alpha$). Для $\rho$, равного расстоянию Хеллингера, и $\alpha = 1/2$ годится следующее множетсво $S$, которое задается неравенством $\sum_i 1/z_i \leq 1$. Просто проверяется, что в качестве $y_{ij}$ в этом случае можно взять $(\sum_k z_k^{-1}\sqrt{\mu_k(vj|v)})/(z_i^{-1}\sqrt{\mu_i(vj|v)})$, а в качестве $\nu(vj|v)$ можно взять $(\sum_k z_k^{-1}\sqrt{\mu_k(vj|v)})^2/(\sum_d(\sum_k z_k^{-1}\sqrt{\mu_k(vd|v)})^2)$. Как мы видим, описанный способ прогнозирования похож на способ Соломонова, но апостериорные вероятности пропорциональны не $\mu_i$, а $\sqrt{\mu_i}$. Можно ли на таком языке описать способ Соломонова?

## Раздел 2

Пусть $n=2$, $C$ — достаточно большое число, тем большее, чем ближе $\alpha$ к 1. В качестве $S$ возьмем множество пар вида $\langle 1+Cu, 1+C/u\rangle$. Предположим, что $z_i = 1+Cu \leq z_{1-i}$. Тогда положим $\nu(\cdot|v) = \mu_i(\cdot|v)$. При этом $y_{ij}$

определяется как $1 + (Cu\sqrt{\mu_{1-i}(vj|v)/\mu_i(vj|v)})/(\sum_d \sqrt{\mu_i(vd|v)\mu_{1-i}(vd|v)})$.

Я долго и безуспешно старался понять, существует ли универсальная полумера, предсказывающая для всех случайных последовательностей. У меня осталось впечатление, что это очень сложная задача. Найти точную оценку на $\alpha$ для $n > 2$ пока не удалось, но нам с Семеновым кажется, что эта задача в пределах наших с Вами возможностей. Ждем новых идей.

С добрыми пожеланиями, Андрей Мучник