Рассматривается последовательность испытаний, в каждом из которых происходит одно из событий, принадлежащих множеству A. Другими словами, элементы множества A — это возможные (попарно несовместимые) исходы испытания, причем A состоит из всех возможных исходов. Предполагается, что есть несколько экспертов, каждый из которых очередному испытанию сопоставляет свой вероятностный прогноз. То есть каждому элементу из A эксперт сопоставляет вероятность, с которой этот элемент окажется исходом очередного испытания. При этом каждому эксперту известны исходы всех предыдущих испытаний. Таким образом, формально эксперт — это функция, которая каждой конечной последовательности x элементов из A сопоставляет распределение $P(xa|x)$ (при условии, что в предыдущих испытаниях имели место исходы из x, вероятность того, что исходом очередного испытания будет a, с точки зрения эксперта равна $P(xa|x)$).

Наша задача состоит в том, чтобы, зная исходы всех предыдущих испытаний и прогнозы экспертов, сопоставленные всем предыдущим испытаниям и очередному испытанию, сопоставить очередному испытанию свой прогноз. При этом мы хотим, чтобы "точность" нашего прогнозирования была не намного хуже, чем "точность" прогнозирования каждого из экспертов. Для формализации термина "точность" дадим следующие определения.

В первую очередь надо ввести меру того, насколько один прогноз отличается от другого. В качестве этой меры удобно использовать расстояние Хеллингера $\rho(p,q) = \sum\limits_{a\in A} \left(\sqrt{p(a)} - \sqrt{q(a)}\right)^2$. Есть и другие способы измерения расстояний между прогнозами. Если в конце доклада останется время, будут названы некоторые причины, по которым мы будем использовать именно расстояние Хеллингера.

Во-первых, в 1986 г. Вовк доказал следующее (см. [**?**]).

Пусть на пространстве бесконечных последовательностей в конечном алфавите $A = \{a_0, a_1, \dots\}$ (то есть на дереве ветвления $|A|$) даны две вычислимые меры P и Q, положительные на всех конусах. Мера P конуса над вершиной x будет обозначаться $P(x)$. Для $a \in A$ выражение $P(xa|x)$ будет обозначать $P(xa)/P(x)$. Если x фиксированно, то $P(xa|x)$ задачт распределение вероятностей на A. Начало длины n бесконечной последовательности y обозначим через $y_{1:n}$.

Бесконечная последовательность y, случайная по Мартин-Лчфу относительно P, тогда и только тогда случайна по Мартин-Лчфу относительно Q, когда $\sum_n \rho(\lambda a.P(y_{1:n}a|y_{1:n}), \lambda a.Q(y_{1:n}a|y_{1:n})) < \infty$ (в дальнейшем связки $\lambda a.$ будут опускаться). Получить аналогичный критерий в терминах других расстояний не удачтся.

Во-вторых, функция ρ симметрична и ограничена (в отличие, например, от часто используемых в математической статистике расстояния Кульбака и расстояния χ^2). Очевидно, всегда $0 \leqslant \rho \leqslant 2$ и $\rho(p,q) = 0 \Leftrightarrow p = q$. Отметим важную особенность функции ρ. Если $\forall i\ (p(i) = 0 \to q(i) = 0) \wedge \forall i (p(i) \neq 0 \to |p(i) - q(i)| \leqslant \varepsilon \ll p(i))$, то $\rho(p,q) = O(\varepsilon^2)$. Если же $\exists i\ (p(i) = 0\ \wedge\ q(i) = \varepsilon > 0)$, то $\rho(p,q) > \varepsilon$. Это отражает тот факт, что эффект замены нулевой вероятности на ненулевую бывает гораздо больше, чем эффект небольшого изменения положительной вероятности. Возможно, именно поэтому неудобно использовать расстояния

$$L_1(p,q) = \sum_i |p(i) - q(i)|;$$

$$L_2(p,q) = \sqrt{\sum_i (p(i) - q(i))^2};$$

$$L_\infty(p,q) = \max_i |p(i) - q(i)|.$$

С другой стороны, в отличие от L_1, L_2, L_∞, функция ρ не является нормой: $\rho(p,q)$ бывает строго больше $\rho(p,r) + \rho(q,r)$ $\bigl($например, для $p = (1,0)$, $q = (0,1)$, $r = \left(\frac{1}{2}, \frac{1}{2}\right)\bigr)$. Неравенство, близкое к неравенству треугольника, для ρ всч-таки есть: $\rho(p,q) \leqslant 2(\rho(p,r) + \rho(q,r))$.

Пусть произошло N испытаний, в которых имели место исходы $\langle y_1, \ldots, y_N \rangle = Y$. Пусть каждому началу Y двумя способами сопоставлялись вероятностные прогнозы следующего испытания; то есть даны функции $R(Y_{1:n}a|Y_{1:n})$ и $Q(Y_{1:n}a|Y_{1:n})$, которые по $n \in [0,N]$ и $a \in A$ указывают вероятность того, что исходом $n+1$-го испытания будет a при условии, что исходами первых n испытаний были $\langle y_1, \ldots, y_n \rangle$. Расстояние между двумя последовательностями прогнозов естественно определяется как $\sum\limits_{n=0}^{N} \rho(R(Y_{1:n}a|Y_{1:n}), Q(Y_{1:n}a|Y_{1:n}))$.

Предположим, в действительности последовательность Y порождается в соответствии с распределениями вероятностей, заданными функцией R. Для достаточно больших чисел s наша цель — сделать как можно меньше вероятность относительно R события

$\left\{Y \;\middle|\; \sum\limits_{n=0}^{N} \rho(R(Y_{1:n}a|Y_{1:n}), Q(Y_{1:n}a|Y_{1:n})) > s\right\}$. Обозначим эту вероятность $\mathcal{F}_{R,Q}(s)$.

Употребленный выше термин "точность" соответствует скорости убывания по s функций $\mathcal{F}_{R,Q}(s)$ и $\mathcal{F}_{R,P_j}(s)$, где функция P_j задает прогнозы j-го эксперта, а функциия Q задает наши прогнозы, полученные на основании прогнозов всех экспертов.

Основное соображение состоит в следующем. Если для всех j вероятности $\mathcal{F}_{P_j,Q}(s)$ достаточно малы при больших s, то "точность" нашего прогнозирования Q будет достаточно велика в ситуации, когда "точность" прогнозирования *хотя бы одного* из экспертов P_j достаточно велика. Это очевидно следует из неравенства для расстояния Хеллингера $\rho(p,q) \leqslant 2(\rho(p,r)+\rho(q,r))$.