М. Ф. Бордок, А. А. Илюхин (Москва, ТВП). **О сложности определения двоичного вектора при наличии случайной информации.**

Пусть задана двоичная последовательность фиксированной длины: $\Theta = (\theta_1, \theta_2, \ldots, \theta_n)$, $\theta_i \in \{0,1\}$, $i = 1, 2, \ldots, n$. Предполагается, что существует тест, который позволяет определить истинную последовательность. В этом случае среднее число опробований до нахождения истинной последовательности при тотальном переборе составит $S = (N+1)/2$, где $N = 2^n$ — число последовательностей.

Пусть на основании наблюдений некоторого случайного вектора $\overline{X} = (X_1, X_2, \ldots, X_n)$ можно получить информацию об истинной последовательности, тогда среднее число опробований может быть существенно снижено. А именно, имеет место соотношение: $S(\overline{X}) = \sum_{k=1}^{N} kP(\Theta^{(k)}|\overline{X})$, где $P(\Theta^{(1)}|\overline{X}) \geqslant P(\Theta^{(2)}|\overline{X}) \geqslant \cdots \geqslant P(\Theta^{(N)}|\overline{X})$. Для определения вероятности, когда $S(\overline{X})$ станет меньше некоторого заданного числа, необходимо знать функцию распределения $S(\overline{X})$. Нахождение распределения $S(\overline{X})$ представляет собой сложную задачу даже для случая, когда распределение случайного вектора $\overline{X}$ имеет простой вид. Поэтому целесообразно получить оценку снизу величины $S(\overline{X})$ некой случайной величиной, распределение которой будет иметь известный вид.

В работе, представленной данным докладом, для получения оценки снизу величины $S(\overline{X})$ используется квадрат длины вектора $\overline{P}(\overline{X}) = (P(\Theta^{(1)}|\overline{X}), P(\Theta^{(2)}|\overline{X}), \ldots, P(\Theta^{(N)}|\overline{X}))$ в N-мерном пространстве, т. е. $\|\overline{P}(\overline{X})\| = \sum_{i=1}^{N} P^2(\Theta^{(i)}|\overline{X})$. Если $S(\overline{P}(\overline{X})) \geqslant (N+1)/3$, то из условия $\|\overline{P}(\overline{X})\| \leqslant 12(c-(N+1)/2)^2/[N(N^2-1)]$ следует, что $S(\overline{P}(\overline{X})) \geqslant c \geqslant (N+1)/3$, где c — некоторый параметр.

Если $(N+1)/3 - k/3 \geqslant S(\overline{P}(\overline{X})) \geqslant (N+1)/3 - (k+1)/3$, то из условия

$$\|\overline{P}(\overline{X})\| \leqslant \frac{12(c-(N-k)/2)^2}{(N-k)(N-k-1)(N-k-2)} + \frac{k+1}{N(N-k-1)}$$

следует, что $S(\overline{P}(\overline{X})) \geqslant c$.

Использование для оценки $\|\overline{P}(\overline{X})\|$ целесообразно в силу того, что при определенных условиях распределение случайной величины $\|\overline{P}(\overline{X})\|$ асимптотически при $n \to \infty$ будет стремиться к логарифмически нормальному распределению. Действительно, легко видеть, что $\|\overline{P}(\overline{X})\| = \prod_{i=1}^{n}(1/2+2(1/2-P(\Theta^{(i)}|\overline{X}))^2)$. Следовательно, если случайные величины $P(\Theta^{(i)}|\overline{X})$ при $i = 1, 2, \ldots, n$ независимы и одинаково распределены, то при $n \to \infty$ распределение случайной величины $\|\overline{P}(\overline{X})\|$ стремится к логарифмически нормальному распределению. Плотность данного распределения имеет вид $p(x) = (1/(\sigma x\sqrt{2\pi}))e^{-(\ln(x)-a)^2/(2\sigma^2)}$, где a и σ —- параметры. Моменты данного распределения $\mathbf{E}\,X^k = e^{ka+k^2\sigma^2/2}$. Следовательно, параметры распределения a и σ можно определить, исходя из первых двух моментов $\mathbf{E}\,X$ и $\mathbf{E}\,X^2$. В нашем случае имеем

$$\mathbf{E}\,\|\overline{P}(\overline{X})\|^j = \left(\mathbf{E}\left(\frac{1}{2} + 2\left(\frac{1}{2} - P(\Theta^{(1)}|\overline{X})\right)^2\right)^j\right)^n, \quad j = 1, 2.$$

Таким образом, при достаточно большом n для нахождения распределения случайной величины $\|\overline{P}(\overline{X})\|$ достаточно знать значения первых четырех моментов распределения случайной величины $P(\Theta|\overline{X})$. Это позволит получить оценку $P\{S(\overline{P}) \leqslant x\} \leqslant 1 - F_{LN(a,\sigma)}(y(x))$, где $F_{LN(a,\sigma)}$ — функция распределения логарифмически-нормального закона с параметрами a и σ, $y(x) = (4x-2)/(9x^2-9x+2) - 1/N$, $1 \leqslant x \leqslant (N+1)/2$.

Для случая, когда $X_j = \Theta_j + \eta_j$, $j = 1, 2, \ldots, n$, где η_j — независимые одинаково распределенные случайные величины, не зависимые с Θ_j, условные вероятности имеют вид

$$P\{\Theta_j = 0 | X_j = y_J\} = \frac{P\{\eta_j = y_j\}}{P\{\eta_j = y_j\} + P\{\eta_j = y_{j-1}\}}.$$

Для проверки точности приведенной оценки рассмотрим случай, когда $P\{\eta_j = k\} = 2^{-t}\binom{t}{k}$, следовательно, $P\{\Theta_j|X\} = 1 - \eta_j/t$. Тогда

$$\mathbf{E}\,\|\overline{P}\| = \left(\frac{1}{2}\right)^n \left(1 + \frac{1}{t}\right), \quad \mathbf{E}\,\|\overline{P}\|^2 = \left(\frac{1}{2}\right)^n \left(\frac{1}{2} + \frac{1}{t} + \frac{3}{2t^2} - \frac{1}{t^3}\right).$$

Исходя из данных соотношений при $t = 2$, получим

$$a = 2n \ln\left(\frac{n+1}{2n}\right) - \frac{n}{2} \ln\left(\frac{n^3 + 2n^2 + 3n - 2}{4n^2}\right),$$

$$\sigma^2 = n \ln\left(\frac{n+1}{2n}\right) - 2n \ln\left(\frac{n^3 + 2n^2 + 3n - 2}{4n^2}\right).$$

При $t = 2$ распределение среднего числа перебора до нахождения истинной последовательности будет иметь вид

$$\widetilde{F}(x) = \sum_{j=n-1-|\log_2(x-1/2)|} \frac{1}{2^n}\binom{n}{k}.$$

Для $n = 500$ разность между оценкой $\widetilde{F}(x)$ функции распределения и функцией распределения $F(x)$ составила: при $x \leqslant 2,4 \cdot 10^{67}$ и $x \geqslant 5,2 \cdot 10^{92}$ — меньше 10^{-7}; на промежутке от $2,44 \cdot 10^{57}$ до $5,19 \cdot 10^{92}$ максимальное значение разности равно $0,2227$ в точке $x = 2,26 \cdot 10^{74}$.