

И. А. Кравченко (Москва, ИТМ и ВТ РАН). **Об определении многократного применения общей двоичной последовательности по искаженным случайным подпоследовательностям.**

Относительно двух двоичных последовательностей $\bar{Y}_N^{(1)} = \{Y_1^{(1)}, \dots, Y_N^{(1)}\}$, $\bar{Y}_N^{(2)} = \{Y_1^{(2)}, \dots, Y_N^{(2)}\}$ рассматриваются две альтернативные гипотезы $\mathbf{H}_0$ и $\mathbf{H}_1$.

Гипотеза $\mathbf{H}_0$ состоит в том, что $\bar{Y}_N^{(1)}, \bar{Y}_N^{(2)}$ представляют собой реализации случайных последовательностей $\bar{U}_N^{(1)}, \bar{U}_N^{(2)}$ соответственно, которые образуются следующим образом: $U_t^{(i)} = X_{\tau_i(t)} \oplus Z_t^{(i)}$, $t = 1, \dots, N$, $i = 1, 2$, где

1) $\bar{X} = \{X_j\}_{j=1}^\infty$ — последовательность независимых в совокупности случайных величин, имеющих распределение $\mathbf{P}\{X_j = 0\} = \mathbf{P}\{X_j = 1\} = 1/2$, $j = 1, 2, \dots$;

2) $\bar{Z}^{(i)} = \{Z_j^{(i)}\}_{j=1}^N$ — последовательность независимых в совокупности случайных величин, имеющих распределение $\mathbf{P}\{Z_j^{(i)} = 0\} = (1/2)(1 + \delta)$, $\mathbf{P}\{Z_j^{(i)} = 1\} = (1/2)(1 - \delta)$, $0 \leq \delta \leq 1$, $j = 1, \dots, N$, $i = 1, 2$;

3) $\tau_i(t) = \sum_{r=1}^t h_r^{(i)}$, $t = 1, 2, \dots$, $\bar{h}^{(i)} = \{h_r^{(i)}\}_{r=1}^\infty$ — независимые в совокупности одинаково распределенные случайные величины; распределение $h_1^{(i)}$ невырождено и сосредоточено на решетке натуральных чисел с максимальным шагом 1; математическое ожидание a и дисперсия σ^2 случайной величины $h_1^{(i)}$ конечны ($a > 1$, $\sigma^2 > 0$), $i = 1, 2$;

4) последовательности $\bar{X}$, $\bar{Z}^{(1)}$, $\bar{Z}^{(2)}$, $\bar{h}^{(1)}$, $\bar{h}^{(2)}$ независимы в совокупности.

Таким образом, $\bar{U}_N^{(1)}, \bar{U}_N^{(2)}$ — «искаженные» последовательностями $\bar{Z}_N^{(1)}, \bar{Z}_N^{(2)}$ случайные подпоследовательности из $\bar{X}$, полученные в соответствии с последовательностями $\{\tau_1(t)\}$, $\{\tau_2(t)\}$, которые задают номера используемых знаков последовательности $\bar{X}$.

Гипотеза $\mathbf{H}_1$ состоит в том, что $\bar{Y}_N^{(1)}, \bar{Y}_N^{(2)}$ представляют собой реализации и случайных последовательностей $\bar{B}_N^{(1)}, \bar{B}_N^{(2)}$, где $\bar{B}_N^{(i)} = \{B_t^{(i)}\}_{t=1}^N$ — последовательность независимых в совокупности случайных величин, имеющих распределение $\mathbf{P}\{B_t^{(i)} = 0\} = \mathbf{P}\{B_t^{(i)} = 1\} = 1/2$; $t = 1, \dots, N$; $i = 1, 2$.

Решается задача построения состоятельного критерия различения гипотез $\mathbf{H}_0$ и $\mathbf{H}_1$, при этом параметр δ , введенный при определении последовательностей $\bar{Z}^{(1)}, \bar{Z}^{(2)}$, которые можно рассматривать как «искажающие» последовательность $\bar{X}$, будет рассматриваться как функция от N . Если $\delta = 0$, то гипотезы $\mathbf{H}_0$ и $\mathbf{H}_1$ совпадают и, следовательно, неразличимы, так как в этом случае $\bar{Z}^{(1)}, \bar{Z}^{(2)}$ — две независимые последовательности случайных величин, имеющие равномерное распределение.

Рассмотрим следующую статистику:

$$S_N = \sum_{t=1}^N \sum_{m=1}^N (-1)^{Y_t^{(1)} \oplus Y_m^{(2)}} P_{t,m},$$

где $P_{t,m}$ — вероятность события $A_{t,m} = \{\tau_1(t) = \tau_2(m)\}$, $t \geq 1$, $m \geq 1$.

Статистика S_N представляет собой коэффициент при квадратичном члене разложения статистики отношения правдоподобия рассматриваемых гипотез по степеням параметра δ (коэффициент при линейном члене равен 0).

Обозначим $\tilde{U}$ стандартную нормальную величину, независимую от случайной величины $\tilde{W}$, определяемой следующим равенством:

$$\tilde{W} = \int_0^1 (2x)^{-1/2} \exp \left\{ -\frac{w^2(x)}{2x} \right\} dx,$$

где $w(x)$ — стандартный винеровский процесс на отрезке $[0, 1]$, траектории которого можно считать непрерывными почти наверное.

Имеют место следующие соотношения: $\mathbf{E} \widetilde{W} = 1$, $\mathbf{D} \widetilde{W} = \pi/3 - 1$.

Введем обозначение: $D_N = (2\pi\sigma^2)^{-1/2} a^{-1} N^{1/2}$.

Пусть $\mathbf{E}_i$ ($\mathbf{D}_i$) обозначает знак математического ожидания (дисперсии), которое берется от случайных величин в условиях гипотезы $\mathbf{H}_i$, $i = 0, 1$.

Далее приводятся утверждение и теоремы относительно статистики S_N .

Утверждение. *Имеют место асимптотические соотношения при $N \rightarrow \infty$:*

$$\mathbf{E}_1 S_N = 0, \quad \mathbf{D}_1 S_N = D_N(1 + o(1)), \quad \mathbf{E}_0 S_N = \delta^2 D_N(1 + o(1)).$$

Теорема 1. *В условиях гипотезы $\mathbf{H}_1$ случайная величина $S_N D_N^{-1/2}$ при $N \rightarrow \infty$ слабо сходится к случайной величине, имеющей стандартное нормальное распределение.*

Теорема 2. *В условиях гипотезы $\mathbf{H}_0$ при $N \rightarrow \infty$:*

1) *если $\delta^2 D_N^{1/2} \rightarrow \infty$, то случайная величина $S_N(\delta^2 D_N)^{-1}$ слабо сходится к случайной величине $\widetilde{W}$;*

2) *если $\delta^2 D_N^{1/2} \rightarrow \lambda$ ($0 \leq \lambda < \infty$), то случайная величина $S_N D_N^{-1/2}$ слабо сходится к случайной величине $\widetilde{U} + \lambda \widetilde{W}$.*

Если при $N \rightarrow \infty$ величина $\delta^8 N$ стремится к бесконечности, то существует состоятельный критерий различения гипотез $\mathbf{H}_0$ и $\mathbf{H}_1$. Критерий может быть определен, например, заданием величины $C_N = \delta D_N^{3/4}$ так, что при $S_N < C_N$ принимается гипотеза $\mathbf{H}_1$, а при $S_N \geq C_N$ принимается гипотеза $\mathbf{H}_0$.

При подсчете статистики S_N можно пользоваться (теоремы 1 и 2 будут оставаться справедливыми) следующим приближением для вероятности $P_{t,m}$:

$$P_{t,m} \cong \frac{1}{\sqrt{2\pi\sigma^2(t+m)}} \exp \left\{ -\frac{1}{2} \frac{a^2}{\sigma^2} \frac{(t-m)^2}{(t+m)} \right\}.$$