Н. М. Меженная, В. Г. Михайлов (Москва, МИЭМ, МИАН). **Один подход к различению гипотез в задаче о плотном вложении.**

Статистическая версия задачи о плотном вложении (см. [1], [2]) состоит в следующем. Пусть $X_n = \{x_i\}_{i=1}^n$ и $Y = \{y_i\}_{i=1}^\infty$ — последовательности Бернулли с вероятностью успеха $0 < p < 1$. Гипотеза H_1 состоит в том, что X_n и Y связаны соотношениями $x_k = y_{\sum_{j=1}^k Z_j}$, где $Z = \{Z_j\}_{j=1}^\infty$ — последовательность независимых случайных величин с распределением $\mathbf{P}\{Z_1 = 1\} = 1$, $\mathbf{P}\{Z_j = 1\} = \mathbf{P}\{Z_j = 2\} = 1/2$, $j = 2, 3, \dots$. Гипотеза H_0 состоит в том, что X_n и Y независимы.

Отрезок $x_1, \dots, x_k$ из знаков алфавита $A_N = \{0, \dots, N-1\}$ называется *плотно заполненным знаком* $a \in A_N$, если $a \in \{x_i, x_{i+1}\}$, $i = 1, \dots, k-1$. Плотно заполненный (знаком a) отрезок называем *плотной a-серией*, если он не содержится ни в каком плотно заполненном отрезке большей длины. *Длиной* плотной a-серии называется длина минимального отрезка, содержащего все ее знаки a. Число входящих в a-серию знаков a называется *весом* плотной a-серии. Статистические свойства плотных серий заданной длины либо заданного веса в последовательности Бернулли описаны в работах [3] и [4].

Для проверки гипотезы H_1 воспользуемся следующей процедурой. Пусть заданы число $0 < \varepsilon < 1$ и функция двух натуральных аргументов $C(\cdot,\cdot)$ (их выбор зависит от требуемых вероятностей ошибок).

1. Находим в X_n наибольшую завершенную серию из единиц. Обозначим s ее длину, а T — место появления.

2. Пусть $h = \max\{1;\ 1,5T - C(n,T) - 0,5\}$, $H = 1,5T + C(n,T) + 0,5$. Среди тех плотных 1-серий в Y_m, которые начинаются внутри отрезка $y_{[h]}, y_{[h]+1}, \dots, y_{[H]}$, находим серию максимального веса (если их несколько, выбираем любую из них), ее вес обозначим w.

3. Если $w \geqslant s + (s-1)(p/2)(1-\varepsilon)$, то принимаем гипотезу H_1, в противном случае гипотеза H_1 отклоняется.

Пусть $Q = p(2-p)$, $B_{n,p}(x) = \sum_{k=0}^{[np(1-x)]} C_n^k p^k (1-p)^{n-k}$, $0 < x < 1$, и $P(H_i|H_j)$ — вероятность принять гипотезу в том случае, когда верна гипотеза H_j (здесь $i, j \in \{0,1\}$). Можно показать, что

$$\mathbf{P}(H_0|H_1) \leqslant B_{s-1,p/2}(\varepsilon) + 2B_{T-1,1/2}\left(\frac{2C(n,T)}{T-1}\right),$$

$$\mathbf{P}(H_1|H_0) \leqslant 2(C(n,T)+1)p(1-Q)Q^{s+(s-1)p(1-\varepsilon)/2-1}.$$

Асимптотическое свойства критерия описаны в следующей теореме.

Теорема. *Пусть* $C(n,T) = \ln n\sqrt{T}$ *и*

$$\frac{1}{2} + \frac{\ln Q}{\ln(1/p)}\left(1 + \frac{p}{2}\right) < 0. \tag{1}$$

Тогда найдутся такие числа $0 < \varepsilon \leqslant 1 - p$, $0 < \delta \leqslant 1$ *и* $\alpha > 0$, *что при условии* $s \geqslant (1-\delta)\ln n/\ln(1/p)$ *имеет место соотношение* $\mathbf{P}(H_1|H_0) + \mathbf{P}(H_0|H_1) = O(n^{-\alpha})$, $n \to \infty$.

З а м е ч а н и е. Неравенство (1) выполнено для всех p из диапазона $0 < p \leqslant 0,52$. При значениях $p > 0,5$ критерий нужно строить по длинным 0-сериям.

Работа выполнена при поддержке РФФИ, проект № 08-01-00078а.

СПИСОК ЛИТЕРАТУРЫ

1. *Golic J. Dj.* Constrained embedding probability for two binary strings. — SIAM J. Discrete Math., 1996, v. 9, № 3, p. 360–364.
2. *Михайлов В. Г.*, *Меженная Н. М.* Оценки для вероятности плотного вложения одной дискретной последовательности в другую. — Дискретн. матем., 2005, т. 17, в. 3, с. 19–27.
3. *Меженная Н. М.* Предельные теоремы пуассоновского типа для числа плотных серий в случайной последовательности. — Обозрение прикл. и промышл. матем., 2008, т. 15, в. 3, с. 565–566.
4. *Меженная Н. М.* Предельные теоремы для числа плотных серий в случайной последовательности. — Дискретн. матем., 2009, т. 21, в. 1, с. 105–116.