А. Л. Костевич, И. С. Никитина (Минск, БГУ). **Уточнение процедуры Бонферрони для семейства χ^2-критериев, построенных по l-граммам различной длины.**

Одной из наиболее известных процедур множественной проверки гипотез является процедура Бонферрони, которая по тестируемой выборке отвергает гипотезу $\mathcal{H}_0$, если ее отверг хотя бы один из M использованных критериев. Оптимизация процедуры требует учета зависимостей между статистиками критериев при выборе их уровней значимости. Рассмотрим задачу выбора уровней значимости M критериев $\chi^2(l)$, построенных по непересекающимся l-граммам длин $l \in \{l_1, \ldots, l_M\}$, при их совместном применении в процедуре Бонферрони для проверки гипотезы $\mathcal{H}_0$ о независимости и равновероятности испытаний Бернулли.

Для исследования выбрана схема тестирования [1]: выборка разбивается на K непересекающихся подвыборок объема n; для каждых k $(1 \le k \le K)$ и m $(1 \le m \le M)$ по k-ой подвыборке вычисляются статистики $S^{(k)}_{\chi^2(l_m)}$ критериев $\chi^2(l_m)$ для проверки гипотезы $\mathcal{H}_0$; затем для критериев $\chi^2(l_m)$ вычисляются агрегированные статистики:

$$\tilde{S}_{\chi^2(l_m)} = \frac{1}{\sqrt{2K(2^{l_m}-1)}} \sum\nolimits_{k=1}^{K} S^{(k)}_{\chi^2(l_m)} - K\left(2^{l_m}-1\right), \quad 1 \le m \le M.$$

Теорема 1. *При верной $\mathcal{H}_0$ в асимптотике $n, K \to \infty$ предельным распределением вектора $(\tilde{S}_{\chi^2(l_1)}, \ldots, \tilde{S}_{\chi^2(l_M)})'$ является M-мерное нормальное распределение с нулевым вектором математического ожидания и ковариационной матрицей $\Sigma = (\sigma_{ij})$:*

$$\sigma_{ii} = 1, \quad \sigma_{ij} = \frac{1}{\sqrt{(2^{l_i}-1)(2^{l_j}-1)}} \left(\frac{l_i l_j}{HOK^2(l_i, l_j)} \sum_{k=1}^{|A(l_i,l_j)|-1} 2^{a_{k+1}-a_k} - 1 \right), \qquad (1)$$

где a_k — k-й по величине член множества $A(l_i, l_j) = \{0,\, l_i,\, 2l_i,\, \ldots,\, (HOK(l_i,l_j)/l_i)\, l_i,$ $l_j,\, 2l_j,\, \ldots,\, (HOK(l_i,l_j)/l_j - 1)\, l_j\}$, $a_k < a_{k+1}$, $|A(l_i,l_j)| = HOK(l_i,l_j)(1/l_i + 1/l_j)$.

Следствие. *Если l_i делит l_j, то $\sigma_{ij} = \sqrt{(2^{l_i}-1)/(2^{l_j}-1)}$.*

Для описанной схемы в условиях теоремы 1 для критерия $\chi^2(l_m)$ следует отклонить гипотезу $\mathcal{H}_0$, если его агрегированная статистика больше заданного порога: $\tilde{S}_{\chi^2(l_m)} > \Delta$, где $\Delta = \Phi^{-1}(1-\alpha_c)$, α_c — уровень значимости критерия, $\Phi(\cdot)$ — ф.р. стандартного нормального распределения. В случае совместного использования критериев χ^2 на базе агрегированных статистик для уточнения оценки сверху для вероятности ошибки первого рода процедуры Бонферрони

$$\varepsilon = \mathsf{P}\left\{\text{отвергнуть } \mathcal{H}_0 \text{ любым критерием} \mid \mathcal{H}_0 \text{ истинна}\right\} \qquad (2)$$

может быть применен подход [2]:

Теорема 2. *В условиях теоремы 1 для вероятности* (2) *выполняется:*

$$\varepsilon \le M\alpha_c - (M-1)(2\alpha_c - 1) - \max\nolimits_{1 \le i \le M} \sum\nolimits_{j \ne i} F_{ij}(\Delta, \Delta),$$

где $F_{ij}(\cdot)$ — маргинальная ф.р. подвектора $(\tilde{S}_{\chi^2(l_i)}, \tilde{S}_{\chi^2(l_j)})'$.

Утверждение теоремы 2 с учетом результатов [2] позволяет находить уровни значимости α_c отдельных критериев по заданной границе для ε и наоборот.

СПИСОК ЛИТЕРАТУРЫ

1. NIST SP 800-22. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, 2000.
2. *Костевич А.Л., Милованова И.С.* Улучшенная процедура Бонферрони множественной проверки гипотез для семейства критериев с нормальным распределением статистик. — Обозр. прикл. и пром. матем., 2004, т. 11, вып. 2, с. 242-243.