А.Л. Костевич, И.С. Милованова (Минск, ННИЦ ППМИ БГУ). **Улучшенная процедура Бонферрони множественной проверки гипотез для семейства критериев с нормальным распределением статистик.**

Рассмотрим задачу выбора уровней значимости критериев при проверке нескольких гипотез по одной выборке. Задача такого вида возникает, например, при генерации ключей [1], при тестировании выходных последовательностей криптографических алгоритмов [2].

Пусть выбрано m критериев $\{C_i\}$ со статистиками $\{S_i\}$, предназначенных для проверки гипотез $\{\mathcal{H}_{0,i}\}$ против альтернатив $\{\mathcal{H}_{1,i}\}$ соответственно. Требуется построить такую процедуру проверки гипотезы $\mathcal{H}_0 = \bigcap_{i=1}^m \mathcal{H}_{0,i}$ против объединенной альтернативы $\mathcal{H}_1 = \bigcup_{i=1}^m \mathcal{H}_{1,i}$, для которой "обобщенная" вероятность ошибки первого рода не превосходит заранее заданного значения α:

$$\varepsilon = \mathbf{P}\left\{\text{отвергнуть любую } \mathcal{H}_{0,i} \mid \mathcal{H}_0\right\} \le \alpha. \tag{1}$$

Пусть регистрируется выборка X и по ней вычислены P–значения критериев $\{C_i\}$: $p_1(S_1), \ldots, p_m(S_m)$. Одной из наиболее известных процедур множественной проверки гипотез является процедура Бонферрони [3]:

$$\text{принимается } \begin{cases} \mathcal{H}_0, & \text{если } p_{\min} \ge \alpha_c, \\ \mathcal{H}_1, & \text{иначе,} \end{cases} \qquad p_{\min} = \min_{1\le i\le m} p_i(S_i), \tag{2}$$

где α_c — уровень значимости критериев, скорректированный для множественной проверки гипотез. Известно [3], что для процедуры (2) выполняется: $\alpha_c \le \varepsilon \le m\alpha_c$. Поэтому по заданному α согласно (1) выбирают $\alpha_c = \alpha/m$.

Известно [3], что если статистики критериев являются зависимыми, то выбор $\alpha_c = \alpha/m$ может привести к завышенной оценке сверху в (1): $\varepsilon/\alpha \ll 1$, что приводит к снижению мощности процедуры Бонферрони. Поэтому рассмотрим задачу уточнения границ для значения ε в широко распространенном случае, когда статистики $\{S_i\}$ являются центрированными и имеют совместное многомерное нормальное распределение вероятностей:

$$\mathcal{L}\left\{(S_1, \ldots, S_m)'\right\} = \mathcal{N}_m\left(\mathbf{0}, \Sigma\right), \;\; \mathbf{D}\left\{S_i\right\} = 1, \;\; p_i(S_i) = 2\Phi\left(-|S_i|\right), \;\; i = \overline{1,m}, \tag{3}$$

где $\Phi(\cdot)$ — функция распределения (ф.р.) стандартного нормального закона.

Теорема. *Для вероятности* (1) *процедуры* (2) *для модели* (3) *выполняется:*

$$1 + (m-2)(1-\alpha_c) - \textstyle\sum_{j=1}^m B_j(\Delta) \le \varepsilon \le 1 + (m-2)(1-\alpha_c) - \max\limits_{1\le j\le m} B_j(\Delta),$$

$$B_j(\Delta) = \textstyle\sum_{i\ne j}^m \left(F_{ij}(-\Delta,-\Delta) - 2F_{ij}(-\Delta,\Delta) + F_{ij}(\Delta,\Delta)\right), \quad \Delta = \Phi^{-1}\left(1-\alpha_c/2\right),$$

где $F_{ij}(\cdot)$ — *маргинальная ф.р. подвектора* $(S_i, S_j)'$, $\mathbf{Corr}\left\{S_i, S_j\right\} = \rho_{ij}$.

Верхняя граница для ε легко вычислима и она может использоваться для расчета α_c, учитывающего попарную зависимости статистик $\{S_i\}$.

В докладе приводятся вычисленные границы для вероятности (1) для критерия серий из FIPS 140-2 [1] и эквивалентного ему набора критериев поиска шаблонов [2]; также обсуждается методика тестирования конкурса NESSIE.

СПИСОК ЛИТЕРАТУРЫ

1. FIPS 140–2. Security Requirements for Cryptographic Modules, 2001.
2. NIST SP 800-22. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, 2000.
3. *Simes R.J.* An Improved Bonferroni Procedure for Multiple Tests of Significance. — Biometrika, 1986, v. 73, p. 751–754.