В. Н. З у б о в (Москва, МГУ). **Исследование скорости слабой сходимости степенных статистик согласия к хи-квадрат распределению.**

В контексте критериев согласия, построенных по сгруппированным данным, мы рассматриваем вектор $\mathbf{Y} = (Y_1, \ldots, Y_k)^T$, имеющий полиномиальное распределение $M_k(n, \pi)$, т. е. (обозначим $\mathbf{P} = \mathbf{P}\{Y_1 = n_1, \ldots, Y_k = n_k\}$)

$$\mathbf{P} = \begin{cases} n! \prod_{j=1}^{k} (\pi_j^{n_j}/n_j!), & n_j = 0, 1, \ldots, n \ (j = 1, \ldots, k) \text{ и } \sum_{j=1}^{k} n_j = n, \\ 0, & \text{иначе,} \end{cases}$$

где $\pi = (\pi_1, \ldots, \pi_k)^T$, $\pi_j > 0$, $\sum_{j=1}^{k} \pi_j = 1$. Мы предполагаем, что число полиномиальных ячеек k фиксированно. Далее будем считать выполненной основную гипотезу H_0: $\pi = \mathbf{p}$.

Основным объектом изучения является построенное по вышеуказанному распределению семейство *степенных статистик согласия* вида

$$t_\lambda(\mathbf{Y}) = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{k} Y_j \left[\left(\frac{Y_j}{np_j} \right)^{\lambda} - 1 \right], \qquad \lambda \in \mathbf{R}.$$

Статистика хи-квадрат и многие другие известные статистики получаются как частные случаи, при этом, как показывают многие исследования (см. [1]), статистика хи-квадрат в этом классе не всегда является оптимальным выбором для построения критериев согласия.

Предполагая выполненной основную гипотезу, рассмотрим преобразование $X_j = (Y_j - np_j)/\sqrt{n}$, $j = 1, \ldots, k$, $r = k - 1$, $\mathbf{X} = (X_1, \ldots, X_r)^T$.

Обозначим $T_\lambda(\mathbf{x})$ статистику t_λ, выраженную в координатах $\mathbf{X}$.

Функцию распределения статистик семейства можно записать в виде вероятности попадания случайного вектора $\mathbf{X}$, имеющего решетчатое распределение, в некоторое множество, обозначаемое B^λ.

В терминах преобразованных переменных X_j множество B^λ ограничено поверхностью $T_\lambda(\mathbf{x}) = c$, где

$$T_\lambda(\mathbf{x}) = \frac{2n}{\lambda(\lambda+1)} \left(-1 + \sum_{i=1}^{r} p_i \left(1 + \frac{x_i}{\sqrt{n} p_i} \right)^{\lambda+1} + p_{r+1} \left(1 - \frac{x_1 + \cdots + x_r}{\sqrt{n} p_{r+1}} \right)^{\lambda+1} \right).$$

Назовем сечение множества B^λ *максимальным* по переменной x_1, если результат ортогонального проектирования этого сечения на плоскость $x_1 = \mathrm{const}$, рассматриваемый как множество в $(r-1)$-мерном пространстве, совпадает с проекцией на ту же плоскость всего множества. Определение очевидным образом обобщается на случай остальных координат.

Автор показал, что справедливо следующее утверждение.

Теорема. *Максимальное сечение (например, для переменной x_1) задается следующей ограничивающей поверхностью*:

$$\frac{c\lambda(\lambda+1)}{2n} = -1 + \sum_{i=2}^{r} p_i \left(1 + \frac{x_i}{\sqrt{n} p_i} \right)^{\lambda+1} + (p_1 + p_{r+1}) \left(1 - \frac{x_2 + \cdots + x_r}{\sqrt{n}(p_1 + p_{r+1})} \right)^{\lambda+1}.$$

Следствие. *Максимальное сечение множества B^λ ограничено поверхностью $T_\lambda(x') = c$, где $x' = (x_2, \ldots, x_r)$, $p'_2 = p_2, \ldots, p'_r = p_r$, $p'_{r+1} = p_1 + p_{r+1}$. Это означает, что проекция максимального сечения множества B^λ на $(r-1)$-мерное пространство переменных есть то же множество B^λ на единицу меньшей размерности, относительно другого набора независимых переменных и вероятностей.*

Последний факт позволяет построить индуктивный шаг, с помощью которого в работе [2] показывается общая для всех степенных статистик согласия оценка аппроксимации их распределения предельным хи-квадрат распределением с r степенями свободы: $\mathbf{P}\left\{T_\lambda(\mathbf{x}) < c\right\} = K_r(c) + O\left(n^{-1+1/(r+1)}\right)$, $r \geqslant 3$.

СПИСОК ЛИТЕРАТУРЫ

1. *Cressie N. A. C.*, *Read T. R. C.* Goodness-of-fit statistics for discrete multivariate data. New York: Springer, 1988.
2. *Ulyanov V. V.*, *Zubov V. N.* Refinement on the convergence of one family of goodness-of-fit statistics to chi-squared distribution. — Hiroshima Mathematical Journal, 2009, v. 39, № 1, p. 133–161.