Асимптотические разложения для распределений обобщенных статистик типа Хотеллинга, построенных по выборкам случайного размера

Монахов М. М.

Рассматривается статистика типа Хотеллинга

$$T_{N_n} = g(n)\,\mathrm{tr}\, S_h S_{N_n}^{-1},$$

построенная по выборкам случайного размера N_n, где $g(n) \to \infty$ при $n \to \infty$. Обозначим через $G_k(x)$ функцию распределения χ^2-распределения с k степенями свободы. Предположим, что выполнены следующие условия.

Условие 1. *Пусть для статистики фиксированного объёма*

$$T_n = n\,\mathrm{tr}\, S_h S_n^{-1}$$

существует константа $C_1 > 0$ такая, что

$$\sup_x \left| \mathbb{P}\big(n\,\mathrm{tr}\, S_h S_n^{-1} \le x\big) - G_k(x) - \frac{k}{4n}\sum_{j=0}^{2} a_j G_{k+2j}(x) \right| \le C_1 n^{-2}.$$

Условие 2. *Существуют $m \in \mathbb{N}$, $\beta > m/2$, $C_2 > 0$, функция распределения $H(y)$ с $H(0+) = 0$, функции ограниченной вариации $h_i(y)$, $i = 1, \ldots, m$, и последовательность $0 < g(n) \uparrow \infty$, $n \to \infty$, такие, что*

$$\sup_y \left| \mathbb{P}\left(\frac{N_n}{g(n)} < y\right) - H(y) - \sum_{i=1}^{m} \frac{1}{n^{i/2}} h_i(y) \right| \le \frac{C_2}{n^\beta}.$$

Теорема 1 (аналог теоремы переноса)**.** *Пусть выполнены условия 1, 2. Тогда существует константа $C_3 > 0$ такая, что*

$$\sup_x \left| \mathbb{P}\big(g(n)\,\mathrm{tr}\, S_h S_{N_n}^{-1} \le x\big) - F_n(x) \right| \le C_1 \mathbb{E} N_n^{-2} + \frac{C_3 + C_2 M_n}{n^\beta},$$

где

$$\begin{aligned} F_n(x) = &\int_{1/g(n)}^{\infty} G_k(xy)\,dH(y) + \sum_{i=1}^{m} \frac{1}{n^{i/2}} \int_{1/g(n)}^{\infty} G_k(xy)\,dh_i(y) \\ &+ \frac{k}{4g(n)} \int_{1/g(n)}^{\infty} \sum_{j=0}^{2} \frac{a_j}{y} G_{k+2j}(xy)\,dH(y) \\ &+ \frac{k}{4g(n)} \sum_{i=1}^{m} \frac{1}{n^{i/2}} \int_{1/g(n)}^{\infty} \frac{1}{y} \sum_{j=0}^{2} a_j G_{k+2j}(xy)\,dh_i(y), \end{aligned}$$

a

$$M_n = \sup_x \int_{1/g(n)}^{\infty} \left| \frac{\partial}{\partial y} \left(G_k(yx) + \frac{k}{4g(n)y} \sum_{j=0}^{2} a_j G_{k+2j}(yx) \right) \right| dy.$$

В частном случае, когда N_n имеет отрицательное биномиальное распределение, смещённое на 1, из теоремы 1 получаются явные разложения Чебышева–Эджворта второго порядка на базе предельного F-распределения.

Asymptotic Expansions for Distributions of Generalized Hotelling-Type Statistics Based on Samples with Random Sizes

M. M. Monakhov

We consider a Hotelling-type statistic

$$T_{N_n} = g(n)\,\mathrm{tr}\, S_h S_{N_n}^{-1},$$

constructed from samples with random sizes N_n, where $g(n) \to \infty$ as $n \to \infty$. Let $G_k(x)$ denote the distribution function of the chi-square distribution with k degrees of freedom. Assume that the following conditions hold.

Condition 1. *For the fixed-size statistic*

$$T_n = n\,\mathrm{tr}\, S_h S_n^{-1}$$

there exists a constant $C_1 > 0$ such that

$$\sup_x \left| \mathbb{P}\big(n\,\mathrm{tr}\, S_h S_n^{-1} \le x\big) - G_k(x) - \frac{k}{4n}\sum_{j=0}^{2} a_j G_{k+2j}(x) \right| \le C_1 n^{-2}.$$

Condition 2. *There exist $m \in \mathbb{N}$, $\beta > m/2$, $C_2 > 0$, a distribution function $H(y)$ with $H(0+) = 0$, functions of bounded variation $h_i(y)$, $i = 1, \dots, m$, and a sequence $0 < g(n) \uparrow \infty$, $n \to \infty$, such that*

$$\sup_y \left| \mathbb{P}\left(\frac{N_n}{g(n)} < y\right) - H(y) - \sum_{i=1}^{m} \frac{1}{n^{i/2}} h_i(y) \right| \le \frac{C_2}{n^\beta}.$$

Theorem 1 (an analogue of the transfer theorem)**.** *Assume that Conditions 1 and 2 hold. Then there exists a constant $C_3 > 0$ such that*

$$\sup_x \big|\mathbb{P}\big(g(n)\,\mathrm{tr}\, S_h S_{N_n}^{-1} \le x\big) - F_n(x)\big| \le C_1 \mathbb{E} N_n^{-2} + \frac{C_3 + C_2 M_n}{n^\beta},$$

where

$$\begin{aligned} F_n(x) = &\int_{1/g(n)}^{\infty} G_k(xy)\, dH(y) + \sum_{i=1}^{m} \frac{1}{n^{i/2}} \int_{1/g(n)}^{\infty} G_k(xy)\, dh_i(y) \\ &+ \frac{k}{4g(n)} \int_{1/g(n)}^{\infty} \sum_{j=0}^{2} \frac{a_j}{y} G_{k+2j}(xy)\, dH(y) \\ &+ \frac{k}{4g(n)} \sum_{i=1}^{m} \frac{1}{n^{i/2}} \int_{1/g(n)}^{\infty} \frac{1}{y} \sum_{j=0}^{2} a_j G_{k+2j}(xy)\, dh_i(y), \end{aligned}$$

and

$$M_n = \sup_x \int_{1/g(n)}^{\infty} \left| \frac{\partial}{\partial y} \left(G_k(yx) + \frac{k}{4g(n)y} \sum_{j=0}^{2} a_j G_{k+2j}(yx) \right) \right| dy.$$

In the particular case where N_n has a negative binomial distribution shifted by 1, Theorem 1 yields explicit second-order Chebyshev–Edgeworth expansions based on the limiting F-distribution.