О Б О З Р Е Н И Е
П Р И К Л А Д Н О Й И П Р О М Ы Ш Л Е Н Н О Й
Т о м 23 М А Т Е М А Т И К И В ы п у с к 3
2016

А. А. Л и п а т ь е в (Москва, МГУ). **О распределении следа произведения двух случайных матриц.**

В рамках многомерного анализа вариации (MANOVA) исследуется следующая многомерная линейная модель:

$$X = Q\mathbb{B} + \mathcal{E},$$

где X — случайная матрица наблюдений размеров $N\times p$, Q — неслучайная матрица плана размеров $N \times k$, $\mathbb{B}$ — неслучайная матрица регрессионных коэффициентов размеров $k\times p$ и $\mathcal{E}$ — матрица ошибок размеров $N\times p$ с распределением $N_{N\times p}(O, I_N\otimes \Sigma)$. Рассмотрим следующую линейную гипотезу:

$$H_0 : C\mathbb{B} = O,$$

где C — известная матрица размеров $q \times k$ ранга q. При наложении требований инвариантности процедуры проверки гипотезы при ряде аффинных преобразований статистики критериев оказываются функциями собственных значений матрицы $S_h S_e^{-1}$, где

$$S_h = \widehat{\mathbb{B}}^T C^T \big(C(Q^TQ)^{-1}C^T\big)^{-1} C\widehat{\mathbb{B}} \text{ и } S_e = (X - Q\widehat{\mathbb{B}})^T(X - Q\widehat{\mathbb{B}}),$$

при $\widehat{\mathbb{B}} = (Q^TQ)^{-1}Q^TX$ (см. [1]). Одними из наиболее известных статистик, для которых выполнены требования инвариантности, являются статистика Лоули-Хотеллинга (Lawley-Hotelling) T_{LH} и статистика Бартлетта–Нанда–Пиллай (Bartlett–Nanda–Pillai) T_{BNP} :

$$T_{LH} = \mathrm{tr}\ S_h S_e^{-1},$$
$$T_{BNP} = \mathrm{tr}\ S_h (S_h + S_e)^{-1}.$$

Известно, что в случае большой размерности данных:

$$A : q \text{ фиксировано, } p \to \infty,\ n \to \infty,\ \frac{p}{n} \to c \in (0; 1)$$

статистики Лоули-Хотеллинга и Бартлетта–Нанда–Пиллай сходятся по распределению к нормальному закону (см. [2]). При этом вычислимые оценки точности аппроксимации распределений указанных статистик нормальным распределением в условиях предположения A не известны.

Доклад посвящен новому результату, дающему оценку распределения следа произведения двух матриц, полученных нормировкой из двух независимых матриц с распределением Уишарта. Данный результат позволит в дальнейшем построить вычислимые оценки точности аппроксимации распределений статистик Лоули-Хотеллинга и Бартлетта–Нанда–Пиллай нормальным распределением в случае A большой размерности данных.

Теорема. *Пусть квадратные матрицы* B *и* W *размерности* q *независимы и имеют распределение Уишарта* $W_q(p, I_q)$ *и распределение Уишарта* $W_q(m, I_q)$ *с*

$m = n - p - q$ соответственно, а матрицы U и V являются нормированными вариантами матриц B и W:

$$U = \frac{1}{\sqrt{p}}\left(B - pI_q\right), \quad V = \frac{1}{\sqrt{m}}\left(W - mI_q\right).$$

Введем статистику E_1 следующим образом: $E_1 = |\operatorname{tr} UV|$.
Для функции распределения $F_{E_1}(z)$ случайной величины E_1 выполнено следующее неравенство:

$$\sup_z |F_{E_1}(z) - F_2(z)| \leqslant 4q(q+1)C\frac{(\sqrt{m/p}+1)}{\sqrt{m}},$$

где $F_2(z)$ — функция распределения модуля разности двух независимых случайных величин, имеющих распределение $\chi^2_{q(q+1)/2}$, а C — абсолютная постоянная, значение которой может быть вычислено в явном виде.

СПИСОК ЛИТЕРАТУРЫ

1. *Anderson T. W.* An Introduction to Multivariate Analysis. 3rd ed. New York: Wiley, 2003, 721 p.
2. *Fujikoshi Y.*, *Ulyanov V. V.*, *Shimizu R.* Multivariate statistics. High dimensional and large-sample approximations. Hoboken, N. J.: John Wiley and Sons, 2010, 533 p.