

Исследование распределений статистик, используемых при проверке гипотез о ковариационных матрицах, при наблюдаемых законах, отличных от нормального¹

Лемешко Б.Ю., Помадин С.С.
НГТУ, г. Новосибирск, E-mail: ser@fpm.ami.nstu.ru

В основе существующего аппарата классического корреляционного анализа лежит предположение о принадлежности наблюдаемого случайного вектора многомерному нормальному закону. Базируясь на этом, получены предельные распределения статистик, используемых в соответствующих задачах проверки гипотез. На практике предположение о нормальности наблюдаемых многомерных величин не всегда оказывается справедливым. Что происходит с распределениями статистик критериев, если наблюдаемый закон в той или иной степени отличается от нормального?

Проводимые исследования опирались на методику статистического моделирования. При этом задача моделирования псевдослучайных векторов, «заданным образом» отличающихся от многомерного нормального, решалась с использованием подхода, уже изложенного ранее в [1].

Рассматривались статистики [2]

$$-2 \ln \lambda_1 = mn(\ln n - 1) - n \ln |\mathbf{B}\Sigma_0^{-1}| + \text{tr}(\mathbf{B}\Sigma_0^{-1}), \quad (1)$$

$$-2 \ln \lambda_2 = mn(\ln n - 1) - n \ln |\mathbf{B}\Sigma_0^{-1}| + \text{tr}(\mathbf{B}\Sigma_0^{-1}) + n(\hat{\mathbf{M}} - \mathbf{M}_0)^T \Sigma_0^{-1}(\hat{\mathbf{M}} - \mathbf{M}_0), \quad (2)$$

где m – размерность случайного вектора $\mathbf{X}$, n – объем выборки, $\hat{\mathbf{M}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$ –

оценка вектора математических ожиданий, $\mathbf{B} = \sum_{i=1}^n (\mathbf{X}_i - \hat{\mathbf{M}})(\mathbf{X}_i - \hat{\mathbf{M}})^T$, кото-

рые используются при проверке гипотез о ковариационных матрицах вида $H_0: \Sigma = \Sigma_0$ и $H_0: \Sigma = \Sigma_0, \mathbf{M} = \mathbf{M}_0$. В случае многомерного нормального закона и справедливости гипотезы $H_0: \Sigma = \Sigma_0$ предельным распределением статистики (1) является χ^2 -распределение с $m(m+1)/2$ степенями свободы, а при справедливости $H_0: \Sigma = \Sigma_0, \mathbf{M} = \mathbf{M}_0$ распределение статистики (2) представляет собой χ^2 -распределение с числом степеней свободы $m(m+1)/2 + m$.

Для подтверждения корректности используемой методики были проведены исследования распределений данных статистик в случае многомерного нормального закона. Эти исследования показали, что получаемые эмпирические распределения статистик хорошо согласуются с теоретическими предельными,

¹ Работа выполнена при поддержке Минобразования РФ (проекты № Т02-3.3-3356 и № А03-2.8-280)

полученными в классическом корреляционном анализе, как при известном, так и при неизвестном векторе математических ожиданий.

Далее исследования распределений статистик (1) и (2) были проведены для широкого спектра симметричных многомерных законов от более «островершинных» по сравнению с нормальным, до более «плосковершинных». На рис. 1 в виде $-2\ln\lambda_1^a$ обозначена статистика (1), вычисляемая при наблюдении многомерного закона распределения, смоделированного на базе одномерного экспоненциального семейства с параметром формы a [1]. Исследования проводились при изменении параметра формы в диапазоне от 1 до 10.

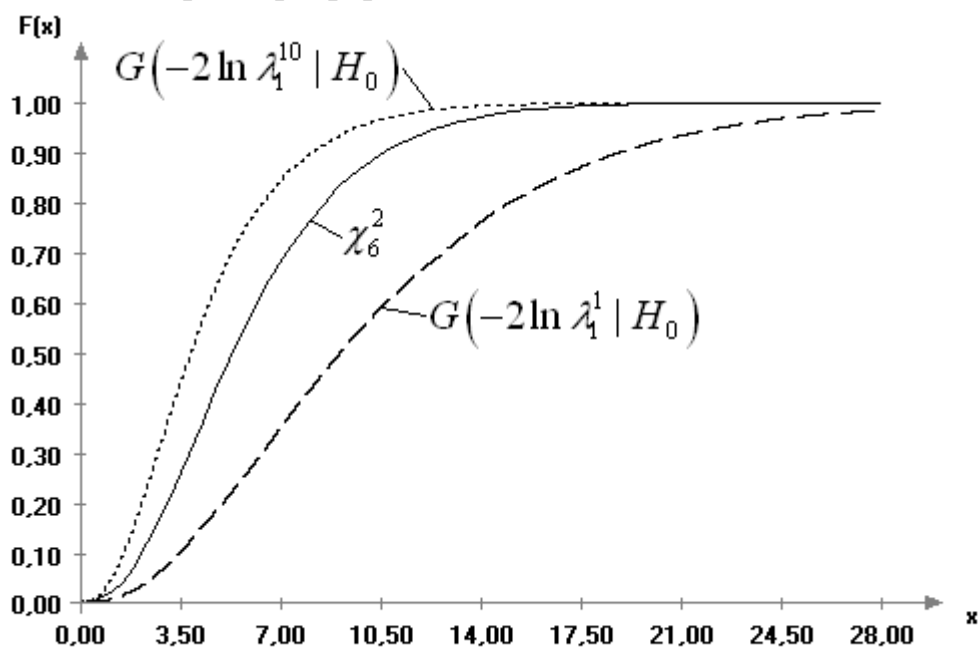


Рис. 1. Эмпирические функции распределения статистики (1) при проверке гипотезы $H_0 : \Sigma = \Sigma_0$ при неизвестном векторе $\mathbf{M}$, $m = 3$; $n = 100$

Результаты исследования статистик, используемых при проверке гипотез о ковариационной матрице, при наблюдении законов, отличающихся от нормального, показали, что распределения данных статистик очень чувствительны к виду наблюдаемого закона: даже при незначительном отклонении от многомерного нормального закона распределения статистик значительно отличаются от классических. Похожая картина наблюдается и в одномерном случае [3].

1. Лемешко Б.Ю., Помадин С.С. Один подход к моделированию псевдослучайных векторов с «заданными» числовыми характеристиками по законам, отличным от нормального // Материалы МНТК «Информатика и проблемы телекоммуникаций». Новосибирск, 2002. – С. 121-122.
2. Андерсон Т. Введение в многомерный статистический анализ. – М.: Физматгиз, 1963. – 500 с.
3. Лемешко Б.Ю., Помадин С.С. Исследование распределений статистик, используемых при проверке гипотез о математическом ожидании и дисперсии, в случае принадлежности наблюдаемых величин экспоненциальному семей-

ству распределений // Материалы региональной конференции «Вероятностные идеи в науке и философии». Новосибирск. 2003. – С.102-105.