

ИССЛЕДОВАНИЕ РАСПРЕДЕЛЕНИЙ СТАТИСТИК, ИСПОЛЬЗУЕМЫХ ПРИ ПРОВЕРКЕ ГИПОТЕЗ О КОВАРИАЦИОННОЙ МАТРИЦЕ, ПРИ НАРУШЕНИИ ОСНОВНОГО ПРЕДПОЛОЖЕНИЯ КОРРЕЛЯЦИОННОГО АНАЛИЗА¹

С.С. Помадин, Б.Ю. Лемешко

Новосибирский государственный технический университет

Тел. (383-2) 46-37-54. e-mail: headrd@fpm.ami.nstu.ru

Аннотация. Проведены исследования распределений статистик, используемых при проверке гипотез о ковариационной матрице (математическое ожидание известно, либо оно оценивается по выборке), в случае наблюдения многомерного закона отличного от нормального. Полученные результаты позволяют определить рамки применимости классического аппарата корреляционного анализа в случае многомерных законов, описываемых моделями, построенными на основе экспоненциального семейства распределений.

В различных приложениях статистического анализа многомерных данных одну из ключевых позиций занимают задачи корреляционного анализа. В процессе решения этих задач вычисляются оценки коэффициентов и матриц парной, частной и множественной корреляции; проверяются различные статистические гипотезы относительно параметров многомерного распределения и коэффициентов корреляции; выявляется наличие и характер взаимосвязи величин, взаимозависимости величин при устранении влияния некоторой совокупности других или зависимости одной случайной величины от группы величин. На основании результатов корреляционного анализа может делаться вывод о наличии и характере функциональной зависимости или о предпочтительности для описания исследуемого объекта регрессионной модели того или иного вида.

В основе классического аппарата корреляционного анализа лежит предположение о принадлежности наблюдаемого случайного вектора многомерному нормальному закону. Базирываясь на этом, получены предельные распределения статистик, используемых в корреляционном анализе [1-2]. На практике предпосылки классического корреляционного анализа выполняются далеко не всегда. Естественно, возникает вопрос о справедливости выводов, получаемых на основании классического аппарата, при нарушении предположения о нормальности.

Ответить на эти вопросы, опираясь только на аналитические методы, чрезвычайно сложно из-за нетривиальности задач. Поэтому в основу проводимого исследования положена развиваемая методика компьютерного анализа статистических закономерностей. С ее помощью проводились исследования распределений статистик, связанных с проверкой гипотез классического корреляционного анализа. Данная работа продолжает начатые ранее исследования, основные результаты которых были описаны в [3-4].

Пусть $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n$ – m -мерная выборка случайной величины объема n ; $\bar{M}$ – вектор математического ожидания; Σ – ковариационная матрица с элементами σ_{ij} ; $\hat{\bar{M}}$ и $\hat{\Sigma}$ – оценки максимального правдоподобия (ОМП) для вектора математического ожидания и ковариационной матрицы:

$$\hat{\bar{M}} = \frac{1}{n} \sum_{i=1}^n \bar{X}_i, \quad \hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (\bar{X}_i - \hat{\bar{M}})(\bar{X}_i - \hat{\bar{M}})^T.$$

Нами проводились исследования распределений статистик, связанных с проверкой следующих гипотез классического корреляционного анализа.

1. При проверке гипотез о ковариационной матрице вида $H_0: \Sigma = \Sigma_0$ в случае неизвестного вектора математических ожиданий $\bar{M}_0$ вычисляемая статистика имеет вид

¹ Работа выполнена при финансовой поддержке МО РФ (проекты № T02-3.3-3356 и № A03-2.8-280)

$$L_1 = mn(\ln n - 1) - n \ln |B \Sigma_0^{-1}| + \text{tr}(B \Sigma_0^{-1}), \quad (1)$$

где

$$B = \sum_{i=1}^n \left(\bar{X}_i - \hat{\bar{M}} \right) \left(\bar{X}_i - \hat{\bar{M}} \right)^T.$$

При справедливости гипотезы H_0 данная статистика подчиняется χ^2 -распределению с числом степеней свободы $m(m+1)/2$, то есть условное распределение $G(L_1 | H_0) = \chi^2_{\frac{m(m+1)}{2}}$ [1].

2. Если проверяется гипотеза вида $H_0 : \Sigma = \Sigma_0, \bar{M} = \bar{M}_0$, то используется статистика

$$L_2 = mn(\ln n - 1) - n \ln |B \Sigma_0^{-1}| + \text{tr}(B \Sigma_0^{-1}) + n \left(\hat{\bar{M}} - \bar{M}_0 \right)^T \Sigma_0^{-1} \left(\hat{\bar{M}} - \bar{M}_0 \right), \quad (2)$$

которая при справедливости проверяемой гипотезы H_0 в качестве предельного распределения $G(L_2 | H_0)$ имеет $\chi^2_{\frac{m(m+1)}{2} + m}$ -распределение, с числом степеней свободы $m(m+1)/2 + m$ [1].

Подчеркнем, что рассмотренные выше статистики имеют в качестве предельных указанные распределения лишь в случае принадлежности наблюдаемой величины многомерному нормальному закону. Как изменятся предельные распределения статистик, если наблюдаемый многомерный закон отличается от нормального, заранее сказать нельзя.

Одной из проблем компьютерного моделирования распределений статистик корреляционного анализа является задача генерирования последовательностей псевдослучайных векторов по законам, «заданным образом» отличающимся от многомерного нормального. Удачный подход к решению данной проблемы был предложен в [5].

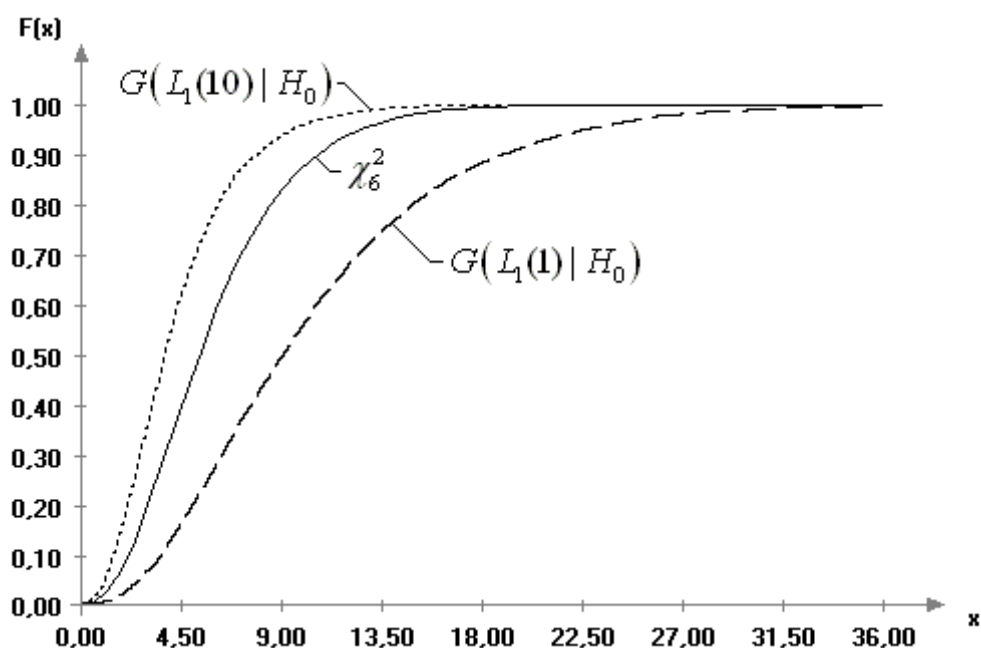


Рис. 1. Распределения статистики L_1 при значения параметра формы экспоненциально-го семейства $\lambda = 1, \lambda = 10$ и классическое предельное распределение статистики L_1 :

χ^2_6 -распределение ($m=3, n=150$)

Распределения статистик L_1 и L_2 , используемые при проверке гипотез о ковариационной матрице, очень чувствительны к виду наблюдаемого закона распределения. Эту зависимость хорошо иллюстрирует приводимый рисунок 1, на котором изображены графики эмпи-

рических функций распределения статистики L_1 , полученные в случае принадлежности наблюдаемой многомерной величины законам, отличающимся от многомерного нормального. На рисунке представлено также предельное распределение статистики L_1 в случае наблюдения многомерного нормального закона: χ^2_6 – распределение.

Из представленной картины очевидно, что распределения статистики L_1 , построенные по «ненормальным» многомерным законам, которые были смоделированы с использованием процедуры, изложенной в [5], по отличному от нормального закону на основе одномерного экспоненциального семейства при значениях параметра формы равных 1 и 10, существенно отличаются от предельного распределения, полученного для многомерного нормального закона.

Аналогичную зависимость от вида наблюдаемого закона демонстрирует и распределение статистики L_2 при проверке совместной гипотезы о ковариационной матрице и математическом ожидании.

Таким образом, результаты проведенных исследований, с одной стороны, подтверждают, что распределения статистик, используемых при проверке гипотез о ковариационной матрице, значительно отличаются от классических предельных при отклонениях наблюдаемого закона от многомерного нормального. В этом существенное отличие от критериев проверки гипотез о векторе математических ожиданий, где распределения статистик весьма устойчивы к отклонениям наблюдаемого закона от многомерного нормального [3,6]. Поэтому при использовании классических процедур для проверки гипотез о ковариационной матрице, также как и при проверке гипотез о дисперсии в одномерном случае [7], целесообразно удостовериться в том, что наблюдаемый многомерный закон является нормальным.

С другой стороны, результаты показывают возможность построения моделей распределений статистик (1) и (2) в случае принадлежности наблюдаемых величин любым многомерным законам. При этом единственной проблемой остаются трудности моделирования многомерных случайных величин в соответствии с заданным законом..

Литература

1. Андерсон Т. Введение в многомерный статистический анализ. – М.: Физматгиз, 1963. – 500 с.
2. Кендалл М., Стьюарт А. Статистические выводы и связи. – Москва: Издательство «Наука», 1973. – 900 с.
3. Лемешко Б.Ю., Помадин С.С. Исследование распределений статистик корреляционного анализа при отклонении многомерного закона от нормального // Тр. V международной конференции «Актуальные проблемы электронного приборостроения» АПЭП-2000. Новосибирск, 2000. – Т. 7. – С. 184-187.
4. Лемешко Б.Ю., Помадин С.С. Распределения статистик корреляционного анализа при отклонении многомерного закона от нормального // Тр. VI международной конференции «Актуальные проблемы электронного приборостроения» АПЭП-2002. Новосибирск, 2002. – Т. 6. – С. 32-35.
5. 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.
6. Лемешко Б.Ю., Помадин С.С. Корреляционный анализ наблюдений многомерных случайных величин при нарушении предположений о нормальности // Сибирский журнал индустриальной математики. 2002. - Т.5. - № 3. - С.115-130.
7. Лемешко Б.Ю., Ванюкевич О.Н. Проверка гипотез о дисперсии при нарушении предположений о нормальности // Сб. научных трудов НГТУ. – Новосибирск, Изд-во НГТУ, 2002. – № 3(29). – С. 27-32.