ИССЛЕДОВАНИЕ РАСПРЕДЕЛЕНИЙ КОЭФФИЦИЕНТА ПАРНОЙ КОРРЕЛЯЦИИ ПИРСОНА

Танасейчук А.В., Коденко А.И., Лемешко Б.Ю.
НГТУ, Новосибирск
E-mail: AWTan@yandex.ru

В данной работе исследовалась зависимость распределения оценок коэффициента парной корреляции и z-преобразования оценок от истинного значения коэффициента корреляции и наблюдаемого многомерного закона. Показано, что распределение оценок коэффициента корреляции r с ростом модуля истинного значения коэффициента отклоняется от нормального закона и становится асимметричным, тогда как z-преобразование Фишера позволяет приблизить распределение к нормальному закону, сгладив асимметричность и исключив зависимость дисперсии от истинного значения корреляции. Показано, что при $r=0$ распределение оценок коэффициента корреляции устойчиво к отклонениям наблюдаемого многомерного закона от нормального, тогда как при $|r|>>0$ появляется зависимость от вида наблюдаемого закона.

Оценка коэффициента парной корреляции между элементами i и j многомерного случайного вектора вычисляется по формуле [1]

$$\hat{r}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}},$$

где $\hat{\sigma}_{ij}$ - элементы $\hat{\Sigma}$ оценки ковариационной матрицы Σ [2]. Предельным распределением оценок $\hat{r}$ (при $r=0$) является нормальное распределение с параметрами $E[\hat{r}]=r$, $D[\hat{r}]=(1-r^2)^2/n$ [3].

При $r \neq 0$ для оценок коэффициента корреляции используют z-преобразование Фишера

$$z(\hat{r}) = \frac{1}{2}\ln\left(\frac{1+\hat{r}}{1-\hat{r}}\right) = arcth\ \hat{r}\ .$$

Преобразование Фишера приводит к статистике $z(\hat{r})$, приближенно описываемой нормальным законам, а также исключает зависимость дисперсии получаемых величин от истинного значения r. Предельным распределением $z(\hat{r})$ считают [1] нормальный закон с параметрами

$$E[z(\hat{r})] \approx \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right),\ D[z(\hat{r})] \approx 1/(n-3). \qquad (1)$$

Исследования проводились с использованием методики статистического моделирования. Моделирование многомерного нормального закона с заданным вектором математического ожидания M и ковариационной матрицей Σ проводилось по методике, описанной в [2], в основании которой лежит линейное преобразование вектора независимых нормально распределенных случайных величин. Процедура моделирования многомерных величин, распределенных по законам, отличным от нормального, с заданными M и Σ

реализована в соответствии с тем же алгоритмом. При этом вектор формируется в соответствии с некоторым одномерным законом распределения. В данном случае удобно использовать двустороннее экспоненциальное распределение ($BiExp(\theta_0;\theta_1;\lambda)$) [2], которое определяет широкий класс симметричных распределений, отличных от нормального. При проведении исследований в качестве распределения с тяжелыми хвостами использовалось многомерное распределение, полученное по методике [2] на основе одномерного распределения Лапласа, в качестве распределения с легкими хвостами – распределение на основе двустороннего экспоненциального распределения с параметром формы $\lambda = 10$.

Результаты. Проведенные статистические эксперименты показали, что зависимость распределения $\hat{r}$ от истинного значения r не ограничивается вариацией параметров нормального распределения. С ростом величины истинного значения r распределение оценок $\hat{r}$ значительно отклоняется от нормального, демонстрируя ярко выраженную асимметричность (рис.1).

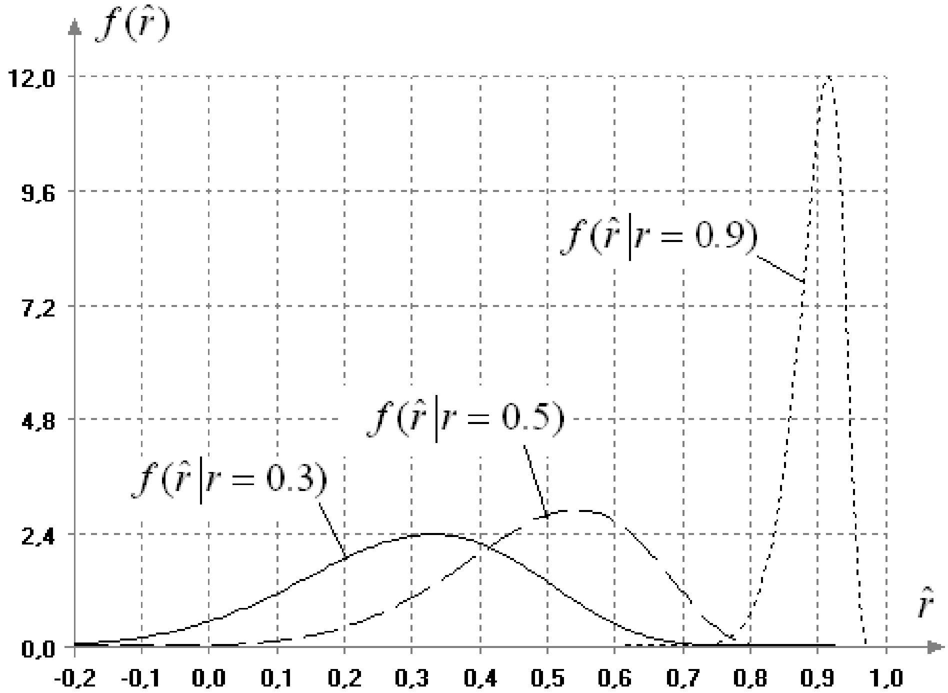

Рис. 1. Плотности распределения для различных значений коэффициента парной корреляции r в случае многомерного нормального закона для выборки объема n=30

Применение z-преобразования Фишера действительно позволяет сгладить асимметричность и исключить зависимость дисперсии распределения оценок от истинного значения r, что можно видеть на рис. 2. При этом следует отметить, что в случае r=0 согласие $z(\hat{r})$ с нормальным законом с параметрами (1) достигается медленнее, чем при $|r|$>0.

Исследования распределений оценок коэффициента корреляции в зависимости от наблюдаемых многомерных законов проводились для трех различных r: при r=0, r=0.3, r=0.9.

Проведенные исследования показали, что в случае $r=0$ распределение оценок $\hat{r}$ устойчиво к отклонениям наблюдаемого многомерного закона от нормального.

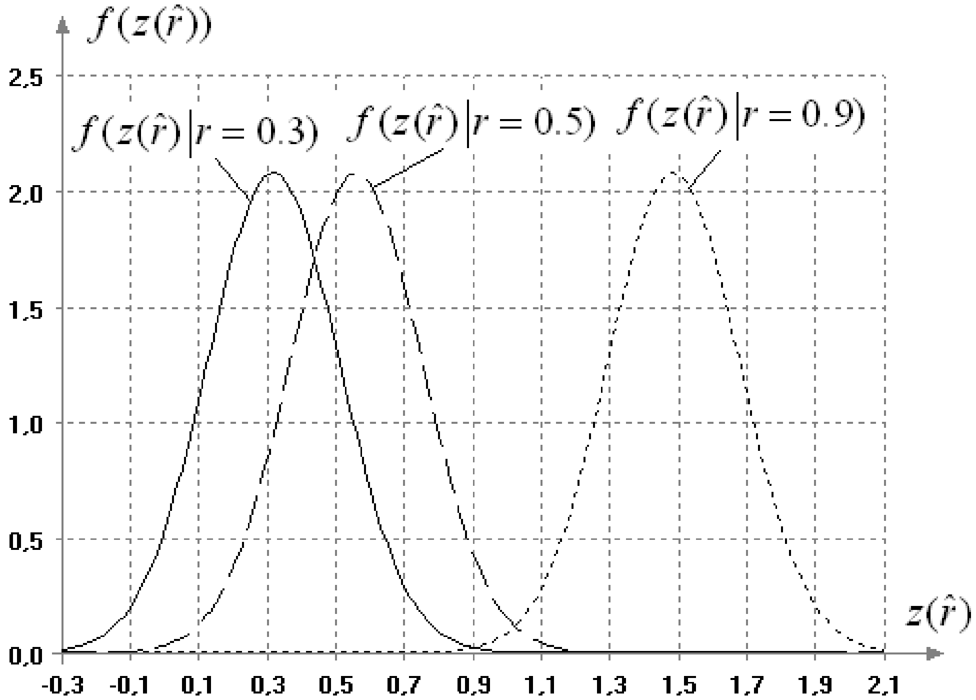

Рис. 2. Плотности распределения z-преобразования для различных значений коэффициента парной корреляции r в случае многомерного нормального закона для выборки объема $n=30$

При $|r|>0$ наблюдается зависимость распределения оценок от наблюдаемого многомерного закона. При этом стоит отметить, что в случае малых объемов выборок ($n=10,\ldots,30$) при многомерных законах с более легкими хвостами (по сравнению с нормальным случаем) отклонения от соответствующих теоретических распределений менее выражены, чем в случае многомерного нормального закона, а также в случае многомерных распределений с тяжелыми хвостами.

Работа выполнена при поддержке РФФИ (проект № 09-01-00056а), Аналитической ведомственной целевой программы "Развитие научного потенциала высшей школы" (проект № 2.1.2/11855) и Федеральной целевой программы Минобрнауки РФ "Научные и научно-педагогические кадры инновационной России".

Литература

1 Кобзарь А.И. Прикладная математическая статистика. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
2 Лемешко Б.Ю., Постовалов С.Н., Чимитова Е.В., Помадин С.С., Французов А.В. Компьютерные методы исследований статистических закономерностей // Тезисы докладов всероссийской НТК «Информационные системы и технологии ИСТ-2001». – Нижний Новгород, 2001. – с. 87-89.
3 Кендалл М., Стьюарт А. Статистические выводы и связи. – М.: Наука, 1973. – 900 с.