

ВЕРОЯТНОСТЬ СОВПАДЕНИЯ ЗНАКОВ ЦЕНТРИРОВАННЫХ ОТНОСИТЕЛЬНО ВЫБОРОЧНОГО СРЕДНЕГО СЛУЧАЙНЫХ ВЕЛИЧИН¹

Колданов П.А.

Нижегородский филиал Национального исследовательского университета
«Высшая школа экономики», г. Нижний Новгород

Поступила в редакцию 20.10.2018, после переработки 04.12.2018.

Одной из мер связи между случайными величинами является вероятность совпадения знаков их центрированных аналогов. В [4] показано, что в классе эллиптических распределений при известном параметре сдвига такая вероятность не зависит от образующей функции. В настоящей работе доказано, что вероятность совпадения знаков случайных величин, центрированных относительно их выборочного среднего, также не зависит от образующей функции при любом объеме наблюдений. Более того, вероятность совпадения знаков случайных величин, центрированных относительно их выборочного среднего, равна вероятности совпадения знаков случайных величин, центрированных относительно их параметра сдвига.

Ключевые слова: матричные эллиптические распределения, вероятность совпадения знаков, инвариантность к образующей функции, инвариантность к параметру сдвига.

Вестник ТвГУ. Серия: Прикладная математика. 2018. № 4. С. 23–30.
<https://doi.org/10.26456/vtpmk515>

Введение

Одно из интенсивно развивающихся направлений математической статистики связано с изучением многомерных распределений. При этом один из вопросов связан с изучением мер связи между компонентами случайного вектора. Наиболее изученным многомерным распределением является многомерное нормальное распределение. Естественным обобщением многомерного нормального распределения являются распределения с плотностями, постоянными на многомерных эллипсоидах, или эллиптические распределения $ECD(\mu, \Sigma, g)$ с параметром сдвига - вектором μ , параметром масштаба - матрицей Σ и образующей функцией g [1].

Хорошо известно, что для нормального распределения полной характеристикой меры зависимости компонент случайного вектора является матрица Σ , а

¹Работа выполнена при финансовой поддержке РФФИ (проект № 18-07-00524).

наиболее популярной мерой зависимости между двумя случайными величинами X_i, X_j является коэффициент корреляции Пирсона

$$\rho_{i,j} = \frac{\text{cov}(X_i, X_j)}{\sqrt{DX_i DX_j}} = \frac{\sigma_{i,j}}{\sqrt{\sigma_{i,i} \sigma_{j,j}}}.$$

Известны также и другие коэффициенты корреляции:

- коэффициент корреляции Спирмена $\rho_{i,j}^{Sp}$ [2]

$$\rho_{i,j}^{Sp} = P((X_i(t) - X_i(t+1))(X_j(t) - X_j(t+2)) > 0), \forall t = 1, \dots, n-2,$$

вектора $\begin{pmatrix} X_i(1) \\ X_j(1) \end{pmatrix}, \dots, \begin{pmatrix} X_i(n) \\ X_j(n) \end{pmatrix}$ – независимые вектора, распределенные как вектор $\begin{pmatrix} X_i \\ X_j \end{pmatrix}, i, j = 1, \dots, p; i \neq j;$

- коэффициент корреляции Кендалла $\tau_{i,j}$ [2]

$$\tau_{i,j} = P((X_i(t) - X_i(t+1))(X_j(t) - X_j(t+1)) > 0), \forall t = 1, \dots, n-1.$$

В [4] для случайного вектора $X = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{pmatrix}$, имеющего эллиптическое распределение $ECD(\mu, \Sigma, g)$, использована мера зависимости двух случайных величин, определяемая вероятностью

$$p^{i,j} = P((X_i - \mu_i)(X_j - \mu_j) > 0), \forall i, j = 1, \dots, p. \quad (1)$$

Доказано, что в классе эллиптических распределений $ECD(\mu, \Sigma, g)$ между мерой (1) и элементами матрицы Σ существует функциональная связь

$$p^{i,j} = \frac{1}{2} + \frac{1}{\pi} \arcsin \frac{\sigma_{i,j}}{\sqrt{\sigma_{i,i} \sigma_{j,j}}}.$$

Этот результат доказан для известного параметра сдвига – вектора μ .

В настоящей работе изучается мера зависимости, не зависящая от значения вектора μ , задаваемая вероятностью

$$P((X_i(t) - \bar{X}_i)(X_j(t) - \bar{X}_j) > 0), \forall t = 1, \dots, n, \quad (2)$$

где $\bar{X}_i = \frac{1}{n} \sum_{t=1}^n X_i(t)$. Доказано, что в классе матричных эллиптических распределений [3], мера (1) и мера (2) совпадают, что позволяет переносить результаты работы [4] на класс матричных эллиптических распределений с неизвестным параметром сдвига μ .

Статья организована следующим образом: в секции 1 вводятся основные определения и обозначения. В секции 2 приведены используемые теоремы о матричных эллиптических распределениях. В секции 3 доказана основная теорема настоящей работы. В конце статьи приведено заключение.

1. Основные определения и постановка задачи

Случайный вектор $X = (X_1, X_2, \dots, X_p)$ имеет эллиптическое распределение $ECD_p(\mu, \Sigma, g)$, если его плотность имеет вид:

$$f(x; \mu, \Lambda) = |\Sigma|^{-\frac{1}{2}} g\{(x - \mu)' \Sigma^{-1} (x - \mu)\}, \tag{3}$$

где Σ – положительно определенная матрица, функция $g(x) \geq 0$, и

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y' y) dy_1 \dots dy_p = 1.$$

Этот класс включает, в частности, многомерное нормальное распределение и многомерное распределение Стьюдента. Известно, также, что если математическое ожидание существует, то $E(X_i) = \mu_i, i = 1, 2, \dots, p$.

В [4] показано, что если вектор $X = (X_1, X_2, \dots, X_p)$ имеет эллиптическое распределение $ECD_p(\mu, \Sigma, g)$, то при известном векторе μ вероятности

$$P((X_i - \mu_i)(X_j - \mu_j) > 0) = \frac{1}{2} + \frac{1}{\pi} \arcsin \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

и не зависят от g .

Основная цель настоящей работы заключается в распространении этих результатов на случай неизвестного вектора μ . Рассмотрим

$$P((X_i(t) - \bar{X}_i)(X_j(t) - \bar{X}_j) > 0), \forall i, j = 1, \dots, p; \forall t = 1, \dots, n. \tag{4}$$

Изучение (4) проводится с использованием аппарата матричных эллиптических распределений.

2. Необходимые сведения о матричных эллиптических распределениях

Пусть

$$X = \begin{pmatrix} X_1(1) & X_1(2) & \dots & X_1(n) \\ X_2(1) & X_2(2) & \dots & X_2(n) \\ \dots & \dots & \dots & \dots \\ X_p(1) & X_p(2) & \dots & X_p(n) \end{pmatrix}$$

случайная матрица размера $p \times n$ [3].

Определение 1. Матрица X имеет матричное эллиптическое распределение $ME_{p,n}(M, \Sigma \otimes \Phi, \psi)$, если ее характеристическая функция имеет вид:

$$\phi_X(t) = \exp(tr(iT'M))\psi(tr(T'\Sigma T\Phi)), \tag{5}$$

где T, M – матрицы размера $p \times n$; Σ – матрица размера $p \times p$; Φ – матрица размера $n \times n$, матрицы Σ, Φ положительно определены, и $\psi : [0, +\infty) \rightarrow R$.

Согласно [3] (теорема 1.17 (ii))

$$tr(T'\Sigma T\Phi) = (vec(T'))'(\Sigma \otimes \Phi')vec(T'),$$

где $\Sigma \otimes \Phi'$ – произведение Кронекера.

Определение 2. Пусть $A = (a_{ij})$ – матрица размера $p \times q$, B – матрица размера $r \times s$. Тогда произведение Кронекера матриц A и B есть матрица размера $pr \times qs$, равная

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1q}B \\ a_{21}B & a_{22}B & \dots & a_{2q}B \\ \dots & \dots & \dots & \dots \\ a_{p1}B & a_{p2}B & \dots & a_{pq}B \end{pmatrix}.$$

Здесь операция $\text{vec}(X')$ дается следующим определением.

Определение 3. Пусть X – матрица размера $p \times n$. Обозначим столбцы X через x_1, x_2, \dots, x_n . Тогда $\text{vec}(X) = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$.

Следующая теорема (теорема 2.1 из [3]) доказывает связь между эллиптическими распределениями (3) и матричными эллиптическими распределениями.

Теорема 1. Случайная матрица X размера $p \times n$ имеет матричное эллиптическое распределение $ME_{p,n}(M, \Sigma \otimes \Phi, \psi)$ тогда и только тогда, когда $x = \text{vec}(X')$ имеет эллиптическое распределение (3) $ECD_{pn}(\text{vec}(M'), \Sigma \otimes \Phi, \psi)$.

Следующая теорема (теорема 2.2 из [3]) доказывает, что линейная функция от случайной матрицы с матричным эллиптическим распределением также имеет матричное эллиптическое распределение.

Теорема 2. Пусть случайная матрица X имеет матричное эллиптическое распределение $ME_{p,n}(M, \Sigma \otimes \Phi, \psi)$, A, B, C – матрицы с элементами из R^1 размеров $q \times p$; $n \times t$; $q \times t$ соответственно. Тогда случайная матрица $AXB + C$ имеет матричное эллиптическое распределение $ME_{q,m}(AMB + C, (A\Sigma A)' \otimes (B'\Phi B), \psi)$.

Следующая теорема (теорема 2.8 из [3]) доказывает, что маргинальное распределение подматрицы случайной матрицы с матричным эллиптическим распределением также имеет матричное эллиптическое распределение.

Теорема 3. Пусть матрица X имеет матричное эллиптическое распределение $ME_{p,n}(M, \Sigma \otimes \Phi, \psi)$. Разобьем матрицы X, M, Σ на подматрицы

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

где X_1 – матрица размера $q \times n$, M_1 – матрица размера $q \times n$, Σ_{11} – матрица размера $q \times q$. Тогда матрица X_1 имеет матричное эллиптическое распределение $ME_{q,n}(M_1, \Sigma_{11} \otimes \Phi, \psi)$.

3. Основной результат

Основной результат настоящей работы заключается в доказательстве следующей теоремы.

Теорема 4. Пусть матрица

$$X = \begin{pmatrix} X_1(1) & X_1(2) & \dots & X_1(n) \\ X_2(1) & X_2(2) & \dots & X_2(n) \\ \dots & \dots & \dots & \dots \\ X_p(1) & X_p(2) & \dots & X_p(n) \end{pmatrix}$$

имеет матричное эллиптическое распределение $ME_{p,n}(M, \Sigma \otimes \Phi, \psi)$ с произвольной симметричной положительно определенной матрицей Σ и единичной матрицей Φ . Предположим, также, что $n > 1$. Тогда для $\forall i, j = 1, \dots, p; i \neq j$ для $\forall t = 1, \dots, n$ матрица $\begin{pmatrix} X_i(t) - \overline{X}_i \\ X_j(t) - \overline{X}_j \end{pmatrix}$ имеет распределение

$$ME_{2,1} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{n-1}{n} \sigma_{ii} & \frac{n-1}{n} \sigma_{ij} \\ \frac{n-1}{n} \sigma_{ij} & \frac{n-1}{n} \sigma_{jj} \end{pmatrix}, \psi \right).$$

Доказательство. Не уменьшая общности, рассмотрим случай $i = 1, j = 2, t = 1$. Согласно теореме 3 матрица

$$\begin{pmatrix} X_1(1) & X_1(2) & \dots & X_1(n) \\ X_2(1) & X_2(2) & \dots & X_2(n) \end{pmatrix}$$

имеет матричное эллиптическое распределение $ME_{2,n}(M, \Sigma \otimes \Phi, \psi)$ где

$$M_{2 \times n} = \begin{pmatrix} \mu_1 & \mu_1 & \dots & \mu_1 \\ \mu_2 & \mu_2 & \dots & \mu_2 \end{pmatrix},$$

$$\Sigma_{2 \times 2} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix},$$

$$\Phi = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Тогда

$$\begin{pmatrix} X_1(1) & X_1(2) & \dots & X_1(n) \\ X_2(1) & X_2(2) & \dots & X_2(n) \end{pmatrix} \times \begin{pmatrix} \frac{n-1}{n} \\ -\frac{1}{n} \\ \dots \\ -\frac{1}{n} \end{pmatrix} = \begin{pmatrix} X_1(1) - \overline{X}_1 \\ X_2(1) - \overline{X}_2 \end{pmatrix}.$$

Поэтому по теореме 2 матрица $\begin{pmatrix} X_1(1) - \overline{X}_1 \\ X_2(1) - \overline{X}_2 \end{pmatrix}$ имеет распределение $ME_{2,1}(AMB + C, (A\Sigma A') \otimes (B'\Phi B), \psi)$ где

$$A_{2 \times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$B_{n \times 1} = \begin{pmatrix} \frac{n-1}{n} \\ -\frac{1}{n} \\ \dots \\ -\frac{1}{n} \end{pmatrix},$$

$$C_{2 \times n} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix},$$

$$\Phi_{n \times n} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Так как

$$AMB+C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 & \mu_1 & \dots & \mu_1 \\ \mu_2 & \mu_2 & \dots & \mu_2 \end{pmatrix} \begin{pmatrix} \frac{n-1}{n} \\ -\frac{1}{n} \\ \dots \\ -\frac{1}{n} \end{pmatrix} + \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$A\Sigma A' = \Sigma_{2 \times 2} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

и

$$B'\Phi B = \left(\frac{n-1}{n}, -\frac{1}{n}, \dots, -\frac{1}{n} \right) \begin{pmatrix} \frac{n-1}{n} \\ -\frac{1}{n} \\ \dots \\ -\frac{1}{n} \end{pmatrix} = \frac{n-1}{n},$$

то

$$(A\Sigma A') \otimes (B'\Phi B) = \begin{pmatrix} \frac{n-1}{n}\sigma_{11} & \frac{n-1}{n}\sigma_{12} \\ \frac{n-1}{n}\sigma_{12} & \frac{n-1}{n}\sigma_{22} \end{pmatrix}.$$

Поэтому матрица $\begin{pmatrix} X_1(1) - \overline{X_1} \\ X_2(1) - \overline{X_2} \end{pmatrix}$ имеет распределение

$$ME_{2,1} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{n-1}{n}\sigma_{11} & \frac{n-1}{n}\sigma_{12} \\ \frac{n-1}{n}\sigma_{12} & \frac{n-1}{n}\sigma_{22} \end{pmatrix}, \psi \right).$$

□

Из теоремы 4 и леммы 2 из [4] следует, что

$$P((X_i(t) - \overline{X_i})(X_j(t) - \overline{X_j}) > 0) =$$

$$= \frac{1}{2} + \frac{1}{\pi} \arcsin \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}, \forall t = 1, \dots, n; \forall i, j = 1, \dots, p, i \neq j.$$

Заключение

В настоящей работе показано, что вероятность совпадения знаков центрированных относительно математического ожидания случайных величин с матричным эллиптическим распределением равна вероятности совпадения знаков центрированных относительно выборочного среднего случайных величин. Этот результат показывает, что соответствующие результаты работы [4] справедливы и при неизвестном параметре сдвига эллиптического распределения. Однако при

этом требуется, чтобы совместное распределение наблюдений над случайным вектором описывалось матричным эллиптическим распределением. В этом случае сетевые структуры в знаковой сети центрированных относительно математического ожидания случайных величин совпадают с сетевыми структурами в знаковой сети центрированных относительно выборочного среднего случайных величин. Кроме того, статистические процедуры идентификации сетевых структур в знаковой сети в классе матричных эллиптических распределений $ME_{n,p}(\mu, \Sigma, g)$ будут обладать инвариантной функцией риска относительно параметров μ, g .

Работа выполнена в лаборатории алгоритмов и технологий анализа сетевых структур.

Список литературы

- [1] Anderson T.W. An Introduction to Multivariate Statistical Analysis. 3rd edition. New York: John Wiley and Sons, 2003. 721 p.
- [2] Kruskal W.H. Ordinal measures of association // Journal of the American Statistical Association. 1958. Vol. 53, № 284. Pp. 814–861.
- [3] Gupta A.K., Varga T., Bodnar T. Elliptically Contoured Models in Statistics and Portfolio Theory. New York: Springer-Verlag, 2013. 321 p. <https://doi.org/10.1007/978-1-4614-8154-6>
- [4] Kalyagin V.A., Koldanov A.P., Koldanov P.A. Robust identification in random variable networks // Journal of Statistical Planning and Inference. 2017. Vol. 181. Pp. 30–40.

Образец цитирования

Колданов П.А. Вероятность совпадения знаков центрированных относительно выборочного среднего случайных величин // Вестник ТвГУ. Серия: Прикладная математика. 2018. № 4. С. 23–30. <https://doi.org/10.26456/vtpmk515>

Сведения об авторах

1. Колданов Петр Александрович

доцент кафедры прикладной математики и информатики факультета информатики, математики и компьютерных наук Нижегородского филиала Национального исследовательского университета «Высшая школа экономики».

Россия, 603155, г. Нижний Новгород, ул. Большая Печерская, д. 25/12, НИУ ВШЭ - Нижний Новгород. E-mail: pkoldanov@hse.ru

PROBABILITY OF SIGN COINCIDENCE CENTERED WITH RESPECT TO SAMPLE MEAN RANDOM VARIABLES

Koldanov Petr Alexandrovich

Associate Professor at Applied Mathematics and Computer Science department,
Nizhny Novgorod branch of the National Research University
"Higher School of Economics"
Russia, 603155, Nizhny Novgorod, Bolshaya Pecherskaya str., 25/12,
NRU – HSE in Nizhny Novgorod.
E-mail: pkoldanov@hse.ru

Received 20.10.2018, revised 04.12.2018.

Probability of sign coincidence of centered random variables is one possible measure of connection. In [4] it was shown that the measure does not depend from generating function in the class of elliptically contoured distributions. This result was obtained for known shift parameter. In the present paper it is proved that for any sample size the probability of sign coincidence centered with respect to the sample mean random variables does not depend on generating function too. Moreover it is proved that the probability of sign coincidence centered with respect to the sample mean random variables is equal to the probability of sign coincidence centered with respect to the shift parameter random variables.

Keywords: matrix elliptically contoured distribution, probability of sign coincidence, invariance with respect to generating function, invariance with respect to shift parameter.

Citation

Koldanov P.A., "Probability of sign coincidence centered with respect to sample mean random variables", *Vestnik TverGU. Seriya: Prikladnaya Matematika* [Herald of Tver State University. Series: Applied Mathematics], 2018, no. 4, 23–30. (in Russian) <https://doi.org/10.26456/vtpmk515>

References

- [1] Anderson T.W., *An Introduction to Multivariate Statistical Analysis*, 3rd edition, John Wiley and Sons, New York, 2003, 721 pp.
- [2] Kruskal W.H., "Ordinal measures of association", *Journal of the American Statistical Association*, **53**:284 (1958), 814–861.
- [3] Gupta A.K., Varga T., Bodnar T., *Elliptically Contoured Models in Statistics and Portfolio Theory*, Springer-Verlag, New York, 2013, 321 pp., <https://doi.org/10.1007/978-1-4614-8154-6>.
- [4] Kalyagin V.A., Koldanov A.P., Koldanov P.A., "Robust identification in random variable networks", *Journal of Statistical Planning and Inference*, **181** (2017), 30–40.