Math-Net.Ru

Общероссийский математический портал

Е. М. Суханова, Матричная корреляция, *Теория вероятн. и ее примен.*, 2009, том 54, выпуск 2, 383–391

DOI: <https://doi.org/10.4213/tvp2716>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.208.226.222

26 апреля 2023 г., 17:12:09

© 2009 г.

СУХАНОВА Е. М.*

МАТРИЧНАЯ КОРРЕЛЯЦИЯ¹⁾

Предложена новая матричнозначная корреляционная мера для пары многомерных случайных величин. Матричная корреляция в основных чертах повторяет свойства классического коэффициента корреляции с тем отличием, что роль чисел выполняют квадратные матрицы. Матричная корреляция также позволяет объединить многие известные понятия многомерного корреляционного и регрессионного анализа.

Ключевые слова и фразы: многомерная выборка, корреляция, асимптотическое распределение.

1. Введение. Пусть даны два случайных вектора $x \in \mathbf{R}^p$ и $y \in \mathbf{R}^q$ с конечными невырожденными (положительно определенными) дисперсионными матрицами. Рассмотрим вариацию линейной комбинации $Kx + y \in \mathbf{R}^q$, где $K \in \mathbf{R}_p^q$ — произвольная матрица. (Здесь и далее символом $\mathbf{R}_p^q$ обозначаем множество вещественных матриц с q строками и p столбцами; индексы, равные 1, опускаются.) Легко проверить, что

$$0 \preccurlyeq \text{Var}[Kx + y] = \text{Var } y - \text{Cov}(y, x)[\text{Var } x]^{-1} \text{Cov}(x, y) + (K + \text{Cov}(y, x)[\text{Var } x]^{-1}) \text{Var } x (K + \text{Cov}(y, x)[\text{Var } x]^{-1})^T.$$

Для двух симметричных матриц W, V запись $W \preccurlyeq V$ означает, что разность $V - W$ положительно полуопределена. Выбрав $K = -\text{Cov}(y, x)[\text{Var } x]^{-1}$, получаем матричный аналог неравенства Коши–Буняковского–Шварца (см. [1, гл. 2, § 16, лемма 2, с. 159]):

$$\text{Cov}(y, x)[\text{Var } x]^{-1} \text{Cov}(x, y) \preccurlyeq \text{Var } y. \quad (1)$$

Равенство в (1) достигается в том и только том случае, когда $\text{Var}[Kx + y] = 0$, т.е. с вероятностью 1 векторы x и y аффинно связаны.

Неравенство (1) можно переписать в виде $\rho \rho^T \preccurlyeq I_q$, где

$$\rho = \rho(y, x) = [\text{Var } y]^{-1/2} \text{Cov}(y, x)[\text{Var } x]^{-1/2} \quad (2)$$

и $I_q \in \mathbf{R}_q^q$ — единичная матрица. Отметим, что запись (2) корректна, поскольку положительно определенный квадратный корень из симметричных положительно определенных матриц $\text{Var } x$, $\text{Var } y$ существует и притом единственен.

Для $p = q = 1$ формула (2) дает коэффициент корреляции Пирсона. Для произвольных p, q определение ρ , по существу, не меняется: ρ есть ковариационная матрица, нормированная таким образом, что $|\rho| \preccurlyeq I$, где $|\rho| = (\rho \rho^T)^{1/2}$ — матричный модуль; при этом равенство $|\rho| = I$ выполняется только для аффинно связанных векторов. Исходя из этого, будем называть величину ρ из (2) *матричной корреляцией*.

Элементарные свойства ρ как меры (линейной) связи будут описаны в следующем пункте. Определение выборочной матричной корреляции $\mathbf{r}$ и ее геометрический смысл приведены в п. 3. В гауссовском случае найдены точное распределение матрицы $\mathbf{r}$ (при условии, что $\rho = 0$) и асимптотическое распределение выборочной матричной корреляции для больших n , см. п. 4. Связь матрицы ρ с известными понятиями многомерной взаимозависимости отражена в п. 5. Наконец, в п. 6 показана роль матричной корреляции в задаче многомерной линейной регрессии. Доказательства результатов приведены в п. 7. Краткие выводы содержатся в заключительном п. 8.

* Московский государственный университет им. М.В. Ломоносова, механико-математический факультет, кафедра теории вероятностей, Ленинские горы, 119992 Москва, Россия; e-mail: suhanova.katya@mail.ru

¹⁾ Работа выполнена при поддержке РФФИ (проект № 06-01-00454).

2. Простейшие свойства ρ как меры связи. В [2] и [3] были сформулированы несколько аксиом, которым должна удовлетворять любая мера зависимости. Расширяя запас скаляров до кольца квадратных матриц, «переведем» на многомерный язык желаемые свойства.

(А) *Нормировка.* Данное свойство является прямым следствием неравенства (1) Коши–Буняковского–Шварца: $0 \leq |\rho| \leq I$. Это означает, что все собственные значения матрицы $|\rho|$ не превосходят 1.

(В) *Независимость.* Матрицы $\rho(y, x)$ и $\text{Cov}(y, x)$ одновременно обращаются или не обращаются в нуль (см. определение (2)). Будем говорить, что случайные векторы x и y некоррелированы, если $\rho = 0$. Таким образом, если x и y независимы, то они некоррелированы. В гауссовском случае понятия независимости и некоррелированности совпадают.

(С) *Функциональная зависимость.* Пусть случайные векторы $x \in \mathbf{R}^p$ и $y \in \mathbf{R}^q$ аффинно связаны: $y = Kx + b$ для некоторой неслучайной матрицы $K \in \mathbf{R}_p^q$ полного ранга и некоторого вектора $b \in \mathbf{R}^q$. Из определения (2) простой выкладкой получаем, что $\rho\rho^T = I_q$, $|\rho| = I_q$.

(D) *Свойство возрастания.* Вольно говоря, данное свойство означает увеличение $|\rho|$ при возрастании линейной зависимости. Опишем его формально. Пусть $\mu^*(x) = K^*x + b^*$, $K \in \mathbf{R}_p^q$, $b \in \mathbf{R}^q$ — оптимальная в среднеквадратическом смысле линейная оценка y по x , т.е. $\mathbf{E}y = \mathbf{E}\mu^*(x)$ и

$$\Delta^* := \text{Var}[y - \mu^*(x)] = \inf_{K, b} \text{Var}[y - Kx - b].$$

С учетом (1), минимум дисперсионной матрицы достигается при

$$K^* = \text{Cov}(y, x) [\text{Var } x]^{-1} = [\text{Var } y]^{1/2} \rho(y, x) [\text{Var } x]^{-1/2}. \quad (3)$$

Оптимальный параметр K^* можно назвать *коэффициентом регрессии*. Формула (3) связывает коэффициент регрессии с матричной корреляцией. Значение минимального квадратичного отклонения равно

$$\Delta^* = [\text{Var } y]^{1/2} (I - |\rho(y, x)|^2) [\text{Var } y]^{1/2}.$$

Таким образом, чем больше (в матричном смысле) $|\rho|^2$, тем меньше среднеквадратическая ошибка оценивания Δ^* , и наоборот.

(Е) *Эквивариантность.* Рассмотрим поведение ρ при аффинных преобразованиях векторов x и y . Прежде всего, назовем ортогональную матрицу $\text{sign } C := |C|^{-1}C$ *матричным знаком*. Отметим, что $C = |C| \text{sign } C$ — это полярное разложение матрицы C .

Для $\tilde{x} = A_1x + b_1$ и $\tilde{y} = A_2y + b_2$ с произвольными невырожденными матрицами A_i и векторами b_i , $i = 1, 2$, выполняется соотношение:

$$\rho(\tilde{y}, \tilde{x}) = \text{sign}\{A_2[\text{Var } y]^{1/2}\} \rho(y, x) \text{sign}^T\{A_1[\text{Var } x]^{1/2}\}.$$

В частности, если A_1, A_2 ортогональны, то $\rho(\tilde{y}, \tilde{x}) = A_2\rho(y, x)A_1^T$. Итак, при аффинных преобразованиях матричная корреляция изменяется на ортогональные матрицы, а ее сингулярные значения не меняются.

(F) **Симметричность.** Строго говоря, порядок векторов x и y в определении (2) матричной корреляции важен, ибо $\rho(y, x) = \rho^T(x, y)$. Но такая «симметрия с точностью до транспонирования» не является существенным ограничением, поскольку понятие корреляции тесно связано с несимметричной задачей линейной регрессии (см. (D)).

3. Выборочная матричная корреляция. Перейдем теперь к выборочной версии матричной корреляции (2).

Пусть $(x_i^T, y_i^T)^T$, $i = 1, \dots, n$, суть n независимых реализаций пары $(x^T, y^T)^T$, $x \in \mathbf{R}^p$, $y \in \mathbf{R}^q$. Пусть

$$\text{Var} \begin{pmatrix} x \\ y \end{pmatrix} = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

где $\Sigma_{11} = \text{Var } x$, $\Sigma_{22} = \text{Var } y$ и $\Sigma_{12} = \Sigma_{21}^T = \text{Cov}(x, y)$. В этих обозначениях корреляция (2) записывается в виде

$$\rho = \rho(y, x) = \Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1/2}. \quad (4)$$

Как и в одномерном случае, для оценки ρ воспользуемся выборочной матрицей ковариаций

$$\mathbf{s} = (n-1)^{-1} \sum_{i=1}^n \begin{pmatrix} x_i - \bar{x} \\ y_i - \bar{y} \end{pmatrix} \begin{pmatrix} x_i - \bar{x} \\ y_i - \bar{y} \end{pmatrix}^T = \begin{pmatrix} \mathbf{s}_{11} & \mathbf{s}_{12} \\ \mathbf{s}_{21} & \mathbf{s}_{22} \end{pmatrix}.$$

В матрице $\mathbf{s}$ блоки $\mathbf{s}_{ij}$ имеют те же размерности, что и Σ_{ij} , $i = 1, 2$.

Будем называть *выборочной матричной корреляцией*, построенной по $\mathbf{y} = (y_1, \dots, y_n) \in \mathbf{R}_n^q$ и $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}_n^p$, величину

$$\mathbf{r} = \mathbf{r}(\mathbf{y}, \mathbf{x}) = \mathbf{s}_{22}^{-1/2} \mathbf{s}_{21} \mathbf{s}_{11}^{-1/2}. \quad (5)$$

К определению (5) можно было прийти и другим путем. Пусть сейчас $p = q$. В недавней работе [4] было предложено рассматривать $(p \times n)$ -матрицы $\mathbf{x}$ и $\mathbf{y}$ как «векторы» размерности n , координатами которых являются p -столбцы. Обобщенное скалярное произведение так называемых $(p \times n)$ -векторов $\mathbf{x}$ и $\mathbf{y}$ задается формулой

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i^T. \quad (6)$$

Получаем метрическое пространство $(p \times n)$ -матриц с расширенным (до кольца квадратных матриц $\mathbf{R}_p^p$) запасом скаляров. На него переносятся многие понятия векторных пространств: ортогональности, проекции, линейных подпространств и проч. Такой подход, в частности, позволил изложить линейный многомерный статистический анализ, практически дословно повторяя одномерную теорию.

Вспомним, что выборочная корреляция Пирсона ($p = q = 1$) есть скалярное произведение нормированных остатков $\mathbf{x} - \bar{x}\mathbf{1}$, $\mathbf{y} - \bar{y}\mathbf{1}$, где $\mathbf{1} = (1, \dots, 1) \in \mathbf{R}_n$:

$$r = r(\mathbf{y}, \mathbf{x}) = \left\langle \frac{\mathbf{x} - \bar{x}\mathbf{1}}{|\mathbf{y} - \bar{y}\mathbf{1}|}, \frac{\mathbf{x} - \bar{x}\mathbf{1}}{|\mathbf{x} - \bar{x}\mathbf{1}|} \right\rangle = \frac{\langle \mathbf{y} - \bar{y}\mathbf{1}, \mathbf{x} - \bar{x}\mathbf{1} \rangle}{|\mathbf{y} - \bar{y}\mathbf{1}| |\mathbf{x} - \bar{x}\mathbf{1}|}.$$

Здесь $|z| = \langle z, z \rangle^{1/2} = (zz^T)^{1/2}$ — длина вектор-строки $z \in \mathbf{R}_n^1$.

Точно также для произвольных p, q и матричного скалярного произведения (6)

$$\begin{aligned} \mathbf{r} = \mathbf{r}(\mathbf{y}, \mathbf{x}) &= \langle |\mathbf{x} - \bar{x}\mathbf{1}|^{-1}(\mathbf{y} - \bar{y}\mathbf{1}), |\mathbf{x} - \bar{x}\mathbf{1}|^{-1}(\mathbf{x} - \bar{x}\mathbf{1}) \rangle \\ &= |\mathbf{y} - \bar{y}\mathbf{1}|^{-1} \langle \mathbf{y} - \bar{y}\mathbf{1}, \mathbf{x} - \bar{x}\mathbf{1} \rangle |\mathbf{x} - \bar{x}\mathbf{1}|^{-1}, \end{aligned} \quad (7)$$

где длина $(p \times n)$ -вектора определяется аналогичным образом:

$$|\mathbf{z}| = \langle \mathbf{z}, \mathbf{z} \rangle^{1/2} = (\mathbf{z}\mathbf{z}^T)^{1/2}, \quad \mathbf{z} \in \mathbf{R}_n^p.$$

Полученное определение (7) совпадает с (5).

Поскольку выборочная матрица ковариаций $\mathbf{s}$ является сильно состоятельной оценкой Σ , и $\mathbf{r}$ есть непрерывный функционал $\mathbf{s}$, то, с учетом (4), имеем сходимость матриц: $\mathbf{r}(\mathbf{y}, \mathbf{x}) \rightarrow \rho(y, x)$ п.н. при $n \rightarrow \infty$.

4. Распределение в гауссовском случае. При дополнительных предположениях нормальности и независимости ($\rho = 0$) можно вывести точное распределение матричной корреляции. Прежде чем сформулировать результат, напомним необходимые понятия из многомерного анализа [5]. Линейный оператор векторизации $\text{vec}(A)$ записывает последовательно столбцы матрицы A один под другим в виде вектора. Можно считать, что A и $\text{vec } A$ синонимы. Случайная $(q \times p)$ -матрица $\mathbf{z}$ имеет *матричное нормальное* распределение с параметрами $M \in \mathbf{R}_p^q$ и $\Omega \in \mathbf{R}_{qp}^{qp}$ ($\mathbf{z} \sim N_p^q(M, \Omega)$), если $\text{vec}(\mathbf{z}^T) \sim N^{pq}(\text{vec}(M^T), \Omega)$. Случайная $(p \times p)$ -матрица $\mathbf{w}$ имеет *распределение Уишарта* с m степенями свободы ($\mathbf{w} \sim W_p(m)$), если $\mathbf{w} \stackrel{d}{=} \sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^T$, где $\mathbf{z}_i$ — независимые одинаково распределенные случайные величины с общим распределением $N^p(0, I)$.

Теорема 1. Пусть $\{(x_i^T, y_i^T)^T\}_{i=1}^n$ суть n независимых реализаций пары $x_i \in \mathbf{R}^p$, $y_i \in \mathbf{R}^q$, имеющей совместное нормальное распределение с $\rho(y, x) = 0$, и $\mathbf{r}$ — выборочная версия ρ . Пусть $n \geq p + q + 1$. Тогда

$$|\mathbf{r}|^2(I_q - |\mathbf{r}|^2)^{-1} \stackrel{d}{=} |\mathbf{t}|^2,$$

где $\mathbf{t} \stackrel{d}{=} \mathbf{z}\mathbf{w}^{-1/2}$, $\mathbf{z} \perp \mathbf{w}$, $\mathbf{z} \sim N_p^q(0, I_{qp})$ — матричное нормальное распределение и $\mathbf{w} \sim W_p(n - 1 - q)$ — распределение Уишарта.

Распределение случайной матрицы $\mathbf{t}^T$ из теоремы 1 можно называть *матричным* $(p \times q)$ -распределением Стьюдента с $n - 1 - q$ степенями свободы. Мы получили обобщение известного свойства корреляции Пирсона: если (x_i, y_i) — выборка из двумерного нормального распределения с $\rho = 0$, то

$$\frac{r}{\sqrt{1 - r^2}} \stackrel{d}{=} \frac{t_{n-2}}{\sqrt{n-2}},$$

где t_{n-2} — распределение Стьюдента с $n - 2$ степенями свободы.

Теперь займемся асимптотическим поведением матричной корреляции, вообще говоря, зависимых случайных признаков. Из теории многомерного анализа [5] известно, что для выборки из совместного нормального распределения $N_l(\mu, \Sigma)$, $l = p + q$, асимптотическое распределение выборочной дисперсии $\mathbf{s}$ будет нормальным

$$n^{1/2}(\mathbf{s} - \Sigma) \xrightarrow{d} N_l^l(0, (I + K_{l,l})(\Sigma \otimes \Sigma)), \quad n \rightarrow \infty.$$

Коммутационная матрица $K_{l,m} \in \mathbf{R}_{lm}^{lm}$ — это блочная матрица, в которой (i, j) -й блок размера $m \times l$ имеет единицу на месте (j, i) и нуль на остальных местах; $i = 1, \dots, l$, $j = 1, \dots, m$. Умножая (слева) произвольную матрицу из $\mathbf{R}_{lm}^{lm}$ на $K_{l,m}$, мы получим матрицу с переставленными строками [6].

Кронекерово произведение для двух матриц $A \in \mathbf{R}_q^p$ и $B \in \mathbf{R}_s^r$ определяется формулой $A \otimes B = (a_{ij}B) \in \mathbf{R}_{qs}^{pr}$. Кроме того, для удобства записи будем использовать $\oplus$ для *кронекеровой суммы* двух квадратных матриц $A \in \mathbf{R}_p^p$ и $B \in \mathbf{R}_q^q$: $A \oplus B = A \otimes I_q + I_p \otimes B$.

С помощью дельта-метода (см. [5, с. 79]) немедленно получаем асимптотическую нормальность выборочной матричной корреляции $\mathbf{r}$. Действительно, $\mathbf{r}$ есть функция выборочной матрицы ковариаций $\mathbf{s}$, которая, в свою очередь, асимптотически нормальна. Формулу асимптотической дисперсионной матрицы $\mathbf{r}$ дает следующая теорема.

Теорема 2. Асимптотическое распределение выборочной матричной корреляции $\mathbf{r}$, когда выборка $\{(x_i^T, y_i^T)^T\}_{i=1}^n$, $x_i \in \mathbf{R}^p$, $y_i \in \mathbf{R}^q$, имеет совместное нормальное распределение с $\text{Var } x_i = I_p$, $\text{Var } y_i = I_q$, является нормальным:

$$n^{1/2}(\mathbf{r} - \rho) \xrightarrow{d} N_p^q(0, \Omega), \quad n \rightarrow \infty,$$

где

$$4\Omega = 2(I_q - \rho\rho^T) \otimes (I_p - \rho^T\rho) + ((I_q - \rho\rho^T) \oplus (I_p - \rho^T\rho))(I_{qp} - K_{q,p}\rho^T \otimes \rho).$$

Заметим, что в случае $p = q = 1$ теорема 2 дает хорошо известный результат о распределении пирсоновского коэффициента корреляции: для выборки (x_i, y_i) из двумерного нормального распределения

$$n^{1/2}(r - \rho) \xrightarrow{d} N(0, (1 - \rho^2)^2), \quad n \rightarrow \infty.$$

5. Связь ρ с известными понятиями многомерной взаимозависимости.

Когда $q = 1$, модуль матричной корреляции $|\rho|$ есть вещественное число. Его называют *коэффициентом множественной корреляции*. Это максимально возможная корреляция Пирсона между $y \in \mathbf{R}$ и линейной комбинацией $t^T x \in \mathbf{R}$, $t, x \in \mathbf{R}^p$:

$$|\rho(y, x)| = \max_{t \in \mathbf{R}^p} \rho(y, t^T x).$$

Соответственно, $|r|$ называется выборочным коэффициентом множественной корреляции.

В случае большей размерности, $p \geq q > 1$, модуль матричной корреляции $|\rho|$ теряет свойство инвариантности при аффинных преобразованиях (поскольку классы ортогонально подобных $(q \times q)$ -матриц перестают быть вырожденными; см. п. 2 (Е)). Но аффинно инвариантным будет множество сингулярных значений $\rho_1, \dots, \rho_q$ матричной корреляции:

$$\rho = \rho(y, x) = G(D_\rho, 0)H^T, \quad (8)$$

где $D_\rho = \text{diag}\{\rho_1, \dots, \rho_q\}$, $1 \geq \rho_1 \geq \dots \geq \rho_q \geq 0$ и $H \in \mathbf{R}_p^p$, $G \in \mathbf{R}_q^q$ — ортогональные матрицы.

Числа ρ_i^2 являются решениями уравнения $\det(\rho \rho^T - \lambda I) = 0$. Так как

$$\det(\rho \rho^T - \lambda I) = \det(\Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2} - \lambda I) = \det(\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \lambda I),$$

то сингулярные значения ρ_i , $i = 1, \dots, q$, суть *канонические корреляции* двух векторов $x \in \mathbf{R}^p$ и $y \in \mathbf{R}^q$ [5, с. 175]. Впервые это понятие употребил Хотеллинг (1936) [7].

С помощью матриц H и G из (8) определим случайные векторы $u \in \mathbf{R}^p$ и $v \in \mathbf{R}^q$ следующим образом:

$$u = H^T \Sigma_{11}^{-1/2} x = (u^1, \dots, u^p)^T, \quad v = G^T \Sigma_{22}^{-1/2} y = (v^1, \dots, v^q)^T. \quad (9)$$

Тогда

$$\text{Var} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} I_p & (D_\rho, 0) \\ \begin{pmatrix} D_\rho \\ 0 \end{pmatrix} & I_q \end{pmatrix}.$$

Компоненты $u^1, \dots, u^p$ и $v^1, \dots, v^q$ являются *каноническими величинами* (см. [5, с. 175], [7]).

Определения выборочных канонических корреляций и выборочных канонических величин получаются из (8), (9) заменой теоретических величин ρ , Σ_{11} , Σ_{22} на их оценки r , s_{11} , s_{22} соответственно.

Во многих работах по каноническому корреляционному анализу матрицу ρ зачастую используют как вспомогательную, не придавая ей особого смысла. Интересно отметить, что в [8] такая матрица названа *когерентной* и указано ее геометрическое истолкование: (i, j) -й элемент есть косинус угла между i -й и j -й компонентами двух нормированных векторов (в гильбертовом пространстве случайных величин с конечными вторыми моментами). Но и здесь ρ применяется только для перехода к каноническим координатам. Об одном интересном приложении непосредственно матричной корреляции мы расскажем в следующем пункте.

6. Матричная корреляция и оценивание коэффициента многомерной регрессии. Рассмотрим многомерную линейную регрессионную модель:

$$\mathbf{y} = K\mathbf{x} + b\mathbf{1} + \mathbf{e},$$

где $K \in \mathbf{R}_p^q$, $b \in \mathbf{R}^q$ — неизвестные параметры и $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}_n^p$, $\mathbf{y} = (y_1, \dots, y_n) \in \mathbf{R}_n^q$ — экспериментальные данные. Столбцы матрицы ошибок $\mathbf{e} = (e_1, \dots, e_n) \in \mathbf{R}_n^q$ суть независимые одинаково распределенные случайные векторы.

Подберем матрицу K так, чтобы фактор $\mathbf{x}$ и остатки $\mathbf{y} - K\mathbf{x}$ были некоррелированы:

$$\mathbf{r}(\mathbf{x}, \mathbf{y} - K\mathbf{x}) = 0. \quad (10)$$

Воспользуемся определением выборочной матричной корреляции (7), записанной с помощью скалярного произведения (6). Тогда условие некоррелированности (10) имеет вид:

$$\langle \mathbf{x} - \bar{x}\mathbf{1}, \mathbf{y} - K\mathbf{x} - \bar{y}\mathbf{1} + K\bar{x}\mathbf{1} \rangle = \langle \mathbf{x} - \bar{x}\mathbf{1}, \mathbf{y} - \bar{y}\mathbf{1} \rangle - \langle \mathbf{x} - \bar{x}\mathbf{1}, \mathbf{x} - \bar{x}\mathbf{1} \rangle K^T = 0.$$

Решением этого уравнения будет

$$\hat{K} = \langle \mathbf{y} - \bar{y}\mathbf{1}, \mathbf{x} - \bar{x}\mathbf{1} \rangle \langle \mathbf{x} - \bar{x}\mathbf{1}, \mathbf{x} - \bar{x}\mathbf{1} \rangle^{-1}.$$

Это — известная оценка наименьших квадратов. Итак, использование выборочной матричной корреляции $\mathbf{r}$ приводит для K к оценке наименьших квадратов.

7. Доказательства. Доказательство теоремы 1. Итак, пусть $(x_i^T, y_i^T)^T \sim N_{p+q}(\mu, \Sigma)$ — независимые одинаково распределенные случайные векторы, $\Sigma_{21} = \rho = 0$. Поскольку ρ определена, то Σ_{11} , Σ_{22} — невырожденные и, значит, $\Sigma \succ 0$. Из многомерного анализа известно, что

$$(n-1)\mathbf{s} \stackrel{d}{=} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} (\mathbf{u}^T, \mathbf{v}^T) = \begin{pmatrix} \mathbf{u}\mathbf{u}^T & \mathbf{u}\mathbf{v}^T \\ \mathbf{v}\mathbf{u}^T & \mathbf{v}\mathbf{v}^T \end{pmatrix},$$

где $(\mathbf{u}^T, \mathbf{v}^T)^T \sim N_{m+q}^T(0, \Sigma \otimes I_m)$, $m = n-1$. Очевидным образом, $\mathbf{u} \perp \mathbf{v}$. Тогда для матричной корреляции $\mathbf{r} = \mathbf{r}(\mathbf{y}, \mathbf{x})$ справедливо

$$\mathbf{r} \stackrel{d}{=} (\mathbf{v}\mathbf{v}^T)^{-1/2} \mathbf{v}\mathbf{u}^T (\mathbf{u}\mathbf{u}^T)^{-1/2}.$$

Матрица $P = \mathbf{v}^T (\mathbf{v}\mathbf{v}^T)^{-1} \mathbf{v} \in \mathbf{R}_m^m$ задает оператор проектирования на линейное подпространство строк $\mathbf{v}$, $\text{rank } P = q$. Тогда для $H^T = (\mathbf{v}\mathbf{v}^T)^{-1/2} \mathbf{v} \in \mathbf{R}_m^q$ выполняются равенства

$$P = HH^T, \quad H^T H = I_q.$$

По сути H задает базис оператора проектирования P . Оператор $Q \equiv I_m - P$, $\text{rank } Q = m - q$, является проектором на подпространство, ортогональное к пространству строк $\mathbf{v}$.

Изучаемая статистика имеет вид:

$$|\mathbf{r}|^2 (I - |\mathbf{r}|^2)^{-1} = \mathbf{r} (I - \mathbf{r}^T \mathbf{r})^{-1} \mathbf{r}^T \stackrel{d}{=} H^T \mathbf{u}^T (\mathbf{u} Q \mathbf{u}^T)^{-1} \mathbf{u} H. \quad (11)$$

С учетом свойств распределения Уишарта (см. [5, §§ 6, 7]),

(i) условное распределение $\mathbf{u} Q \mathbf{u}^T \mid \mathbf{v} \sim W_p(\text{rank } Q, \Sigma_{11})$ не зависит от $\mathbf{v}$; следовательно, $\mathbf{u} Q \mathbf{u}^T \perp \mathbf{v}$ и безусловное распределение совпадает с условным: $\mathbf{u} Q \mathbf{u}^T \sim \Sigma_{11}^{1/2} W_p(n-m) \Sigma_{11}^{1/2}$;

(ii) $\mathbf{u} H \mid \mathbf{v} \sim N_q^p(0, \Sigma_{11} \otimes H^T H) = N_q^p(0, \Sigma_{11} \otimes I_q)$; следовательно, $\mathbf{u} H \perp \mathbf{v}$ и безусловное распределение совпадает с условным $\mathbf{u} H \sim \Sigma_{11}^{1/2} N_q^p(0, I_p \otimes I_q)$.

(iii) независимость $\mathbf{u} H \perp \mathbf{u} Q \mathbf{u}^T$ следует из того, что $QH = 0$ и $\mathbf{u} H \perp \mathbf{v}$, $\mathbf{u} Q \mathbf{u}^T \perp \mathbf{v}$.

Пункты (i)–(iii) вместе с представлением (11) влекут за собой утверждение теоремы для $\mathbf{w} = \Sigma_{11}^{-1/2} \mathbf{u} Q \mathbf{u}^T \Sigma_{11}^{-1/2}$ и $\mathbf{z} = H^T \mathbf{u}^T \Sigma_{11}^{-1/2}$. Условие $n \geq p + q + 1$ необходимо для того, чтобы с вероятностью 1 матрица $\mathbf{w}$ была невырожденной.

Доказательство теоремы 2. По условию теоремы $\Sigma = \begin{pmatrix} I_p & \rho^T \\ \rho & I_q \end{pmatrix}$. Как мы уже ранее отмечали,

$$n^{1/2}(\mathbf{s} - \Sigma) \xrightarrow{d} \mathbf{z} \sim N_{p+q}^{p+q}(0, (I + K_{p+q})(\Sigma \otimes \Sigma)). \quad (12)$$

Следовательно, справедливы разложения:

$$\begin{aligned} \mathbf{s}_{11} &= I + n^{-1/2} \mathbf{z}_{11} + o_p(n^{-1/2}), \\ \mathbf{s}_{21} &= \rho + n^{-1/2} \mathbf{z}_{21} + o_p(n^{-1/2}), \\ \mathbf{s}_{22} &= I + n^{-1/2} \mathbf{z}_{22} + o_p(n^{-1/2}). \end{aligned}$$

Символ $o_p(n^{-1/2})$ означает, что $n^{-1/2} o_p(n^{-1/2})$ сходится по вероятности к нулю. Прямое вычисление дает:

$$\begin{aligned} r(\mathbf{y}, \mathbf{x}) &= \mathbf{s}_{22}^{-1/2} \mathbf{s}_{21} \mathbf{s}_{11}^{-1/2} = \left[I - \frac{1}{2} n^{-1/2} \mathbf{z}_{22} + o_p(n^{-1/2}) \right] \\ &\quad \times [\rho + n^{-1/2} \mathbf{z}_{21} + o_p(n^{-1/2})] \left[I - \frac{1}{2} n^{-1/2} \mathbf{z}_{11} + o_p(n^{-1/2}) \right] \\ &= \rho - \frac{1}{2} n^{-1/2} \mathbf{z}_{22} \rho - \frac{1}{2} n^{-1/2} \rho \mathbf{z}_{11} + n^{-1/2} \mathbf{z}_{21} + o_p(n^{-1/2}). \end{aligned} \quad (13)$$

С помощью оператора вес, последовательно записывающего столбцы матрицы в виде вектора, перепишем (13) в векторной форме:

$$\text{vec}[n^{1/2}(\mathbf{r} - \rho)^T] \xrightarrow{d} -\frac{1}{2} (I_q \otimes \rho^T) \text{vec } \mathbf{z}_{22} - \frac{1}{2} (\rho \otimes I_p) \text{vec } \mathbf{z}_{11} + \text{vec } \mathbf{z}_{12}. \quad (14)$$

Далее нам требуется дисперсионная матрица величины

$$(\text{vec}^T \mathbf{z}_{22}, \text{vec}^T \mathbf{z}_{11}, \text{vec}^T \mathbf{z}_{12})^T \sim N_{q^2+p^2+qp}(0, \Omega_z).$$

Ее можно вычислить, используя (12). Поскольку $\mathbf{s}_{11}$ и $\mathbf{s}_{22}$ также являются выборочными матрицами ковариаций для p - и q -мерных наблюдений с единичными вариациями, то $\text{Var } \mathbf{z}_{11} = I_{p^2} + K_p$ и $\text{Var } \mathbf{z}_{22} = I_{q^2} + K_q$.

По определению, дисперсионная матрица случайной матрицы задается формулой $\text{Var } \mathbf{z} \equiv \text{Var}[\text{vec}(\mathbf{z}^T)]$. Представим $\text{Var } \mathbf{z}$ в блочном виде:

$$\text{Var } \mathbf{z} = (I + K_{p+q})(\Sigma \otimes \Sigma) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

где A_{22} — дисперсионная матрица для $(\mathbf{z}_{21}, \mathbf{z}_{22}) \in \mathbf{R}_{p+q}^q$. Представим A_{22} в виде $A_{22} = (B_{ij})_{i,j=1}^q$, где блок $B_{ij} \in \mathbf{R}_{p+q}^{p+q}$ есть вариация i -й и j -й строк матрицы $(\mathbf{z}_{21}, \mathbf{z}_{22})$,

$$B_{ij} = \begin{pmatrix} p \times p & p \times q \\ q \times p & q \times q \end{pmatrix}.$$

Следовательно, оставляя в каждом блоке A_{22} только $(q \times p)$ -матрицу, получим $\text{cov}(\text{vec } \mathbf{z}_{22}, \text{vec } \mathbf{z}_{12})$. Точно так же $(p \times p)$ -матрицы из B_{ij} образуют $\text{Var}[\text{vec } \mathbf{z}_{12}]$.

Чтобы найти B_{ij} , представим матрицы K_{p+q} и $\Sigma \otimes \Sigma$ в виде $(p+q) \times (p+q)$ блоков: $K_{p+q} = (U_{ji})$, где $U_{ji} \in \mathbf{R}_{p+q}^{p+q}$ состоит из нулей и одной единицы на месте (j, i) , а

$$\Sigma \otimes \Sigma = \begin{pmatrix} (\delta_{ij} \Sigma) & (\rho_{ji} \Sigma) \\ (\rho_{ij} \Sigma) & (\delta_{ij} \Sigma) \end{pmatrix}.$$

Здесь $\rho = (\rho_{ij})$ и δ_{ij} — символ Кронекера. Блочные матрицы можно перемножать как обычные, обращаясь с блоками как с элементами. Поэтому

$$B_{ij} = \delta_{ij}\Sigma + \sum_{k=1}^p U_{k,p+i}\rho_{jk}\Sigma + U_{p+j,p+i}\Sigma, \quad i, j = 1, \dots, q.$$

Следовательно, $(q \times p)$ -блок в B_{ij} равен $(\delta_{ij}\rho + U_{ji}^{(q)}\rho)$, где $U_{ji}^{(q)} \in \mathbf{R}_q^q$ отличается от U_{ji} только размером. Следовательно, $\text{cov}(\text{vec } \mathbf{z}_{22}, \text{vec } \mathbf{z}_{12}) = (I_{q^2} + K_q)I_q \otimes \rho$. Кроме того, $(p \times p)$ -блок в B_{ij} имеет вид $(\delta_{ij}I_p + \sum_{k=1}^p U_{ki}^{(p,q)}\rho_{jk}\rho)$, где $U_{ki}^{(p,q)} \in \mathbf{R}_q^p$. Тогда $\text{Var}[\text{vec } \mathbf{z}_{12}] = I_{qp} + K_{q,p}\rho^T \otimes \rho$.

Аналогичным образом рассматривая A_{12} , найдем остальные необходимые ковариационные матрицы. Получим:

$$\Omega_z = \begin{pmatrix} I_{q^2} + K_q & (I_{q^2} + K_q)\rho \otimes \rho & (I_{q^2} + K_q)I_q \otimes \rho \\ \cdot & I_{p^2} + K_p & (I_{p^2} + K_p)\rho^T \otimes I_p \\ \cdot & \cdot & I_{qp} + K_{q,p}\rho^T \otimes \rho \end{pmatrix}.$$

Пропущенные элементы дописываются по симметрии: $\Omega_z = \Omega_z^T$. Получив эту матрицу, легко сосчитать дисперсионную матрицу линейной комбинации из (14). Остается только убедиться, что полученная дисперсия совпадает с приведенной в теореме.

8. Заключение. Канонические корреляции (и коэффициент множественной корреляции как частный случай) являются известным обобщением понятия коэффициента корреляции Пирсона. Существенное отличие такого многомерного аналога от своего «родителя» заключается в утрате информации о направлении связи. Предложенная матричная корреляция (2), имея своим спектром канонические корреляции, позволяет делать выводы как о силе, так и о направлении зависимости. Как следствие, матричная корреляция зависит от выбора системы координат.

«Унаследованным» недостатком матричной корреляции является необходимость совместного распределения данных по нормальному закону, поскольку в общем случае понятия независимости и некоррелированности не совпадают (см. п. 2 (В)). Статистические свойства матричной корреляции имеют большое сходство с одномерной теорией.

Особенностью многомерного анализа является расширение пространства скаляров до (некоммутативного) кольца квадратных матриц. Такое обобщение чисел достаточно распространено в теории матриц и ее приложениях [9]. Совершенно новая идея определить матричное скалярное произведение (6), предложенная Ю. Н. Тюриным [4], имеет под собой серьезные статистические основания. Порождаемые этим скалярным произведением понятия ортогональности и проекции позволяют рассказать линейный многомерный анализ в терминах векторных пространств. В пользу такого подхода свидетельствует и возможность определить с помощью упомянутого скалярного произведения выборочную матричную корреляцию (7). Дополнительно, для теоретической версии скалярного произведения двух случайных векторов, заданного матрицей вторых моментов, определение (2) повторяет (7).

Автор выражает глубокую благодарность профессорам Ю. Н. Тюрину и Х. Ойа за полезные обсуждения.

СПИСОК ЛИТЕРАТУРЫ

1. Боровков А. А. Математическая статистика. М.: Наука, 1984, 472 с.
2. Rényi A. On measures of dependence. — Acta Math. Acad. Sci. Hungar., 1959, v. 10, p. 441–451.
3. Drouot Mari D., Kotz S. Correlation and Dependence. London: Imperial College Press, 2001, 219 p.

4. Тюрин Ю. Н. Многомерный анализ: геометрическая теория. — Теория вероятн. и ее примен. (в печати).
5. Bilodeau M., Brenner D. Theory of Multivariate Statistics. New York: Springer-Verlag, 1999, 288 p.
6. Магнус Я. Р., Нейдеккер Х. Матричное дифференциальное исчисление с приложениями к статистике и эконометрике. М.: Физматлит, 2002, 496 с.
7. Hotelling H. Relation between two sets of variables. — Biometrika, 1936, v. 28, p. 321–377.
8. Scharf L. L., Mullis C. T. Canonical coordinates and the geometry of inference, rate, and capacity. — IEEE Trans. Signal Process., 2000, v. 48, p. 824–830.
9. Хорн Р., Джонсон Ч. Матричный анализ. М.: Мир, 1989, 655 с.

Поступила в редакцию
11.III.2008

© 2009 г.

СЮЛЮКИН А. В.*

О КВАЗИНЕРАВНОМЕРНЫХ ОЦЕНКАХ ОСТАТОЧНЫХ ЧАСТЕЙ ДЛЯ РАЗЛОЖЕНИЯ БЕРГСТРЁМА–ЧЕБЫШЁВА

Получено несколько квазинеравномерных оценок остаточных частей разложения Бергстрёма–Чебышёва для плотностей нормированных сумм n независимых одинаково распределённых случайных величин, т.е. таких оценок, у которых равномерные части убывают при $n \rightarrow \infty$ быстрее, чем неравномерные части.

Ключевые слова и фразы: центральная предельная теорема, асимптотические разложения, точность аппроксимации, неравномерные оценки.

Пусть P — распределение вероятностей на $\mathbf{R}$ с нулевым средним и единичной дисперсией, $P(x)$ — функция распределения меры P , $P_n(x)$ — функция распределения меры $P^{*n}(\sqrt{n}A)$, где $A \in \mathcal{R}(\mathbf{R})$, $\Phi(x)$ — функция распределения стандартного нормального закона. Хорошо известна неравномерная оценка в центральной предельной теореме

$$|P_n(x) - \Phi(x)| \leq \frac{C\beta_3}{(1 + |x|^3)\sqrt{n}},$$

где β_3 — третий абсолютный момент распределения P и $C \leq 31.94$ (см. [4]).

Известна неравномерная оценка Л. В. Осипова остаточного члена асимптотического разложения Эджворта–Крамера

$$\begin{aligned} & \left| P_n(x) - \Phi(x) - \sum_{l=1}^{k-2} \frac{\Gamma_l(x)}{n^{l/2}} \right| \\ & \leq C(k) \left\{ \frac{1}{n^{(k-2)/2}(1 + |x|)^k} \int_{|y| \geq \sqrt{n}(1 + |x|)} |y|^k dP(y) \right. \\ & \quad + \frac{1}{n^{(k-1)/2}(1 + |x|)^{k+1}} \int_{|y| < \sqrt{n}(1 + |x|)} |y|^{k+1} dP(y) \\ & \quad \left. + \left(\sup_{|t| \geq 1/(12\beta_3)} |f(t)| + \frac{1}{2n} \right)^n \frac{n^{k(k+1)/2}}{(1 + |x|)^{k+1}} \right\}, \end{aligned}$$

* Московский государственный университет им. М. В. Ломоносова, механико-математический факультет, кафедра теории вероятностей, Ленинские горы 1, 119992 Москва, Россия; e-mail: syulyukin@yandex.ru