

Вячеслав Евгеньевич МОСЯГИН —
доцент кафедры математического анализа
и теории функций,
кандидат физико-математических наук
vmosyagin@mail.ru

Мария Евгеньевна СЕНЬКОВА —
студентка V курса
marysen@mail.ru

Институт математики и компьютерных наук,
Тюменский государственный университет

УДК 519. 213. 2

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ УПОРЯДОЧЕННОЙ ВЫБОРКИ ИЗ ДВУМЕРНОГО НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

NUMERICAL CHARACTERISTICS OF THE ORDERED SAMPLES FROM BIVARIATE NORMAL DISTRIBUTION

АННОТАЦИЯ. Найдены значения некоторых функционалов, зависящих от распределения порядковых статистик из двумерного нормального распределения.

SUMMARY. Some values of functionals, depending on the order statistics distributions from the bivariate normal distribution, were obtained.

КЛЮЧЕВЫЕ СЛОВА. Двумерное нормальное распределение, двумерный вариационный ряд.

KEY WORDS. Bivariate normal distribution, bivariate ordered samples.

1. Введение. Пусть $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ — выборка из двумерного нормального распределения $N(a_x, a_y, \sigma_x^2, \sigma_y^2, \rho)$ с плотностью:

$$f(X, Y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ \frac{1}{-2(1-\rho^2)} \left(\frac{(x-a_x)^2}{\sigma_x^2} - \frac{2\rho(x-a_x)(y-a_y)}{\sigma_x\sigma_y} + \frac{(y-a_y)^2}{\sigma_y^2} \right) \right\}.$$

Здесь $a_x = EX_1$, $a_y = EY_1$, $\sigma_x^2 = DX_1$, $\sigma_y^2 = DY_1$ — соответственно математические ожидания и дисперсии координат выборки, а $\rho = \rho(X_1, Y_1)$ — коэффициент корреляции.

Распределение $N(0, 0, 1, 1, \rho)$ назовем стандартным двумерным нормальным распределением. Линейные преобразования:

$$U_i = \frac{X_i - a_x}{\sigma_x}, \quad V_i = \frac{Y_i - a_y}{\sigma_y}, \quad i = \overline{1, n}, \quad (1)$$

приводят выборку к стандартной с плотностью:

$$\varphi(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{1}{-2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right\}. \quad (2)$$

Занумеруем в порядке возрастания первые координаты выборки: $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Получим так называемый вариационный ряд, соответствующие членам этого ряда вторые координаты выборки обозначим $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_n$.

Определение 1. Перенумерованную выборку

$$(X_{(1)}, \tilde{Y}_1), (X_{(2)}, \tilde{Y}_2), \dots, (X_{(n)}, \tilde{Y}_n) \quad (3)$$

назовем двумерным вариационным рядом.

Если элементы выборки нанести на плоскость, то точки ряда (3) располагаются в порядке слева направо. Заметим, что в отличие от первых координат вторые координаты имеют произвольный порядок.

Упорядоченные многомерные выборки впервые были рассмотрены в работе [1], посвященной линейному оцениванию параметров многомерного нормального распределения по цензурированным выборкам. Здесь же приводятся без доказательства некоторые частные результаты теорем 3-4.

Настоящая работа посвящена нахождению совместных распределений и числовых характеристик координат двумерного вариационного ряда.

Далее функция $\varphi(u, v)$ будет обозначать плотность (2),

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \quad \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-t^2/2} dt, \quad u \in (-\infty, \infty).$$

Число $C_n^k = \frac{n!}{k!(n-k)!}$, $C_n^k = 0$, если $n < k$. Напомним, что $\Phi(+\infty) = 1$, $\Phi(-\infty) = 0$.

В работе принята сквозная нумерация утверждений и замечаний. Нумерация формул в каждом параграфе самостоятельная. При ссылке на формулу из другого параграфа возникает двойная нумерация, где первая цифра — номер параграфа. Значок ■ означает окончание доказательства.

2. Вспомогательные результаты. Изучим некоторые свойства плотности $\varphi(u, v)$, которые будут полезны при нахождении моментов вторых координат двумерного вариационного ряда.

Лемма 1. Справедливы равенства:

$$\int_{-\infty}^x \varphi(u, v) dv = \varphi(u) \Phi\left(\frac{x - \rho u}{\sqrt{1 - \rho^2}}\right), \quad (1)$$

$$\int_{-\infty}^{+\infty} v^k \varphi(u, v) dv = ((\rho u)^k + C_k^2 (\rho u)^{k-2} (1 - \rho^2)) \varphi(u), \quad k = 1, 2, 3. \quad (2)$$

Доказательство. Формула (1) вытекает из следующего очевидного представления:

$$\varphi(u, v) = \frac{1}{\sqrt{1 - \rho^2}} \varphi(u) \varphi\left(\frac{x - \rho u}{\sqrt{1 - \rho^2}}\right). \quad (3)$$

Подставляя (3) в интеграл (2), получим равенство:

$$\int_{-\infty}^{+\infty} v^k \varphi(u, v) dv = \frac{\varphi(u)}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left(t\sqrt{1-\rho^2} + \rho u \right)^k e^{-t^2/2} dt.$$

Раскрывая под интегралом бином Ньютона, с помощью равенств

$$\int_{-\infty}^{+\infty} t^{2k-1} e^{-t^2/2} dt = 0, \quad k = 1, 2, \dots, \quad \int_{-\infty}^{+\infty} t^2 e^{-t^2/2} dt = \sqrt{2\pi}$$

приходим к формуле (2) ■

Замечание 1. Ввиду симметричности плотности $\varphi(u, v)$ относительно переменных, равенства (1-3) останутся в силе, если в них переменные u и v поменять местами. Отметим также, что выражение для интеграла (2) можно найти при любом целом $k \geq 0$, если воспользоваться известной формулой:

$$\int_{-\infty}^{+\infty} t^{2k} e^{-t^2/2} dt = \sqrt{2\pi} (2k-1)!!.$$

Самостоятельный интерес представляет следующая

Лемма 2. Справедливы равенства:

$$\int_{-\infty}^{+\infty} \varphi(t) \Phi(at + b) dt = \Phi\left(\frac{b}{\sqrt{a^2 + 1}}\right), \quad a, b \in (-\infty, +\infty) \quad (4)$$

$$\int_{-\infty}^{+\infty} \varphi(u, v) \Phi(v) dv = \varphi(u) \Phi\left(\frac{\rho u}{\sqrt{2-\rho^2}}\right). \quad (5)$$

Доказательство. Обозначим

$$F(b) = \int_{-\infty}^{+\infty} \varphi(t) \Phi(at + b) dt, \quad b \in (-\infty, +\infty). \quad (6)$$

Так как подынтегральная функция совместно с ее производной по параметру b мажорируются интегрируемыми функциями:

$$|\varphi(t) \Phi(at + b)| \leq \varphi(t), \quad |\varphi(t) \Phi'_b(at + b)| = \varphi(t) \varphi(at + b) \leq \varphi(t) / \sqrt{2\pi},$$

не зависящими от параметра, то функцию (6) можно дифференцировать и осуществлять в ней предельный переход под знаком интеграла:

$$\begin{aligned} \frac{dF(b)}{db} &= \int_{-\infty}^{+\infty} \varphi(t) \varphi(at + b) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{(t^2 + (at+b)^2)}{2}} dt = \frac{1}{2\pi} e^{\frac{-b^2}{2(a^2+1)}} \int_{-\infty}^{+\infty} e^{-\frac{((a^2+1)t+ab)^2}{2(a^2+1)}} dt = \\ &= \left| z = t\sqrt{a^2+1} + \frac{ab}{\sqrt{a^2+1}} \right| = \frac{1}{2\pi\sqrt{a^2+1}} e^{\frac{-b^2}{2(a^2+1)}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \frac{e^{\frac{-b^2}{2(a^2+1)}}}{\sqrt{2\pi}\sqrt{a^2+1}}. \end{aligned}$$

Отсюда, учитывая, что

$$F(-\infty) = \int_{-\infty}^{+\infty} \varphi(t) \Phi(-\infty) dt = 0,$$

находим

$$F(b) = \frac{1}{\sqrt{2\pi}\sqrt{a^2+1}} \int_{-\infty}^b e^{\frac{-x^2}{2(a^2+1)}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{b}{\sqrt{a^2+1}}} e^{\frac{-z^2}{2}} dz = \Phi\left(\frac{b}{\sqrt{a^2+1}}\right).$$

Утверждение (5) является прямым следствием (4):

$$\int_{-\infty}^{+\infty} \varphi(u, v) \Phi(v) dv = \frac{\varphi(u)}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} e^{\frac{-(v-u\rho)}{2(1-\rho^2)}} dv = \varphi(u) \int_{-\infty}^{+\infty} \varphi(t) \Phi(t\sqrt{1-\rho^2} + \rho u) dt.$$

Для завершения доказательства остается в формулу (4) подставить $a = \sqrt{1-\rho^2}$, $b = \rho u$ ■

Замечание 2. Формулу (4) можно истолковать следующим образом. Функция

$$F(x) = \int_{-\infty}^{+\infty} \varphi(t) \Phi(at + x) dt$$

при любом $a \in (-\infty, +\infty)$ является функцией распределения, причем

$$F(0) = \int_{-\infty}^{+\infty} \varphi(t) \Phi(at) dt = \Phi(0) = \frac{1}{2}.$$

Последнее соотношение можно найти в [2; 46], где указан способ рекуррентного вычисления интегралов вида:

$$I_n(a) = \int_{-\infty}^{+\infty} \varphi(t) \Phi^n(at) dt, \quad n = 1, 2, \dots,$$

используемых для вычисления моментов порядковых статистик из нормального распределения.

3. Совместные распределения координат двумерного вариационного ряда. Рассмотрим выборку:

$$(U_1, V_1), (U_2, V_2), \dots, (U_n, V_n) \in N(0, 0, 1, 1, \rho) \tag{1}$$

Для соответствующего ей двумерного вариационного ряда

$$(U_{(1)}, \tilde{V}_1), (U_{(2)}, \tilde{V}_2), \dots, (U_{(n)}, \tilde{V}_n), \tag{2}$$

найдем совместное распределение координат.

Теорема 1. Совместные плотности распределения компонент ряда (2) представимы в виде:

$$f_{U_{(i)}, \tilde{V}_i}(u, v) = C_n(i) \varphi(x, y) \Phi^{i-1}(u) (1 - \Phi(u))^{n-i}, \tag{3}$$

$$f_{\tilde{V}_i, \tilde{V}_j}(v_1, v_2) =$$

$$= C_n(i, j) \int_{-\infty}^{+\infty} \int_{u_1}^{+\infty} \varphi(u_1, v_1) \varphi(u_2, v_2) \Phi^{i-1}(u_1) (\Phi(u_2) - \Phi(u_1))^{j-i-1} (1 - \Phi(u_2))^{n-j} du_1 du_2, \tag{4}$$

где $1 \leq i < j \leq n$, $C_n(i) = \frac{n!}{(i-1)!(n-i)!}$, $C_n(i, j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}$.

Доказательство. Начнем со второго утверждения. Событие $\{U_{(i)} \in du_1; \tilde{V}_i \in dv_1; U_{(j)} \in du_2; \tilde{V}_j \in dv_2\}$, $i < j$, $-\infty < u_1 < u_2 < \infty$ происходит всякий раз, когда одна из точек $(U_1, V_1), (U_2, V_2), \dots, (U_n, V_n)$ попадает в прямоугольник $du_1 \times dv_1$, другая — в прямоугольник $du_2 \times dv_2$, $i-1$ точек располагается в области $(-\infty, u_1) \times (-\infty, \infty)$, $j-i-1$ точек в полосе $(u_1, u_2) \times (-\infty, \infty)$, а остальные $n-j$ точек попадают в область $(u_2, \infty) \times (-\infty, \infty)$. Вероятность каждой такой перестановки членов выборки равна:

$$\varphi(u_1, v_1)\varphi(u_2, v_2)\Phi^{i-1}(u_1)(\Phi(u_2) - \Phi(u_1))^{j-i-1}(1 - \Phi(u_2))^{n-j} du_1 du_2 dv_1 dv_2,$$

а поскольку число подобных перестановок равно $C_n(i, j)$, то

$$\begin{aligned} P(U_{(i)} \in du_1; \tilde{V}_i \in dv_1; U_{(j)} \in du_2; \tilde{V}_j \in dv_2) = \\ = C_n(i, j)\varphi(u_1, v_1)\varphi(u_2, v_2)\Phi^{i-1}(u_1)(\Phi(u_2) - \Phi(u_1))^{j-i-1}(1 - \Phi(u_2))^{n-j} du_1 du_2 dv_1 dv_2 \end{aligned}$$

Интегрируя это выражение по области $u_1 < u_2$, приходим к формуле (4).

Формула (3) доказывается аналогично ■

В следующем параграфе нам понадобятся еще три выражения для плотностей первых и вторых координат двумерного вариационного ряда:

$$f_{\tilde{V}_i}(v) = C_n(i) \int_{-\infty}^{+\infty} \varphi(u, v)\Phi^{i-1}(u)(1 - \Phi(u))^{n-i} du, \quad i = \overline{1, n} \quad (5)$$

$$f_{U_i}(u) = C_n(i)\varphi(u)\Phi^{i-1}(u)(1 - \Phi(u))^{n-i}, \quad i = \overline{1, n} \quad (6)$$

$$f_{U_{(i)}, U_{(j)}}(u_1, u_2) = C_n(i, j)\varphi(u_1)\varphi(u_2)\Phi^{i-1}(u_1)(\Phi(u_2) - \Phi(u_1))^{j-i-1}(1 - \Phi(u_2))^{n-j}, \quad u_1 < u_2 \quad (7)$$

Равенство (5) является следствием интегрирования обеих частей равенства (3) по переменной u . Формулы (6-7) можно найти в [2-3].

4. Числовые характеристики двумерного вариационного ряда

Теорема 2. Моменты первых и вторых координат двумерного вариационного ряда (3.1) - (3.2) связаны равенствами:

$$E\tilde{V}_i^k = \rho EU_{(i)}^k + C_k^2 \rho^{k-2} (1 - \rho^2) EU_{(i)}^{k-2}, \quad k = 1, 2, 3, \quad (1)$$

$$D\tilde{V}_i = \rho^2 DU_{(i)} + 1 - \rho^2, \quad i = \overline{1, n} \quad (2)$$

Доказательство. Из формул (3.5), (3.6) и леммы 1 находим:

$$\begin{aligned} E\tilde{V}_i^k &= \int_{-\infty}^{+\infty} v^k f_{\tilde{V}_i}(v) dv = C_n(i) \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} v^k \varphi(u, v) dv \right) \Phi^{i-1}(u)(1 - \Phi(u))^{n-i} du \\ &= C_n(i) \int_{-\infty}^{+\infty} \left((\rho u)^k + C_k^2 (\rho u)^{k-2} (1 - \rho^2) \right) \Phi^{i-1}(u)(1 - \Phi(u))^{n-i} du = \\ &= \rho^k \int_{-\infty}^{+\infty} u^k f_{U_{(i)}}(u) du + C_k^2 \rho^{k-2} (1 - \rho^2) \int_{-\infty}^{+\infty} u^{k-2} f_{U_{(i)}}(u) du = \\ &= \rho EU_{(i)}^k + C_k^2 \rho^{k-2} (1 - \rho^2) EU_{(i)}^{k-2}. \end{aligned}$$

В частности,

$$E\tilde{V}_i = \rho EU_{(i)}, \quad E\tilde{V}_i^2 = \rho^2 EU_{(i)}^2 + 1 - \rho^2. \quad (3)$$

Следовательно,

$$D\tilde{V}_i = EV_i^2 - (EV_i)^2 = \rho^2 DU_{(i)} + 1 - \rho^2 \blacksquare$$

В следующем утверждении представлены формулы для смешанных моментов координат двумерного вариационного ряда.

Теорема 3. Смешанные моменты координат вариационного ряда (3.1)-(3.2) связаны с моментами первых координат формулами:

$$EU_{(i)}^k \tilde{V}_i^m = \rho^m EU_{(i)}^{k+m} + C_m^2 \rho^{m-2} (1 - \rho^2) EU_{(i)}^{k+m-2}, \quad (4)$$

$$E\tilde{V}_i^k \tilde{V}_j^m = \rho^{k+m} EU_{(i)}^k U_{(j)}^m + C_k^2 \rho^{k+m-2} (1 - \rho^2) EU_{(i)}^k U_{(j)}^{m-2} + C_m^2 \rho^{k+m-2} (1 - \rho^2) EU_{(i)}^{k-2} U_{(j)}^m + C_k^2 C_m^2 \rho^{k+m-4} (1 - \rho^2)^2 EU_{(i)}^{k-2} U_{(j)}^{m-2}, \quad 1 \leq i < j \leq n. \quad (5)$$

В частности,

$$\text{cov}(U_{(i)}, \tilde{V}_i) = \rho DU_{(i)}, \quad (6)$$

$$\text{cov}(\tilde{V}_i, \tilde{V}_j) = \rho^2 \text{cov}(U_{(i)}, U_{(j)}), \quad i, j = \overline{1, n}. \quad (7)$$

Доказательство. Равенство (4) вытекает из формул (3.3) - (3.7) и леммы 1:

$$\begin{aligned} EU_{(i)}^k \tilde{V}_i^m &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u^k v^m f_{U_{(i)}, \tilde{V}_i}(u, v) dudv = \\ &= C_n(i) \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} v^m \varphi(u, v) dv \right) u^k \Phi^{i-1}(u) (1 - \Phi(u))^{n-i} du = \\ &= C_n(i) \int_{-\infty}^{\infty} ((\rho u)^m + C_m^2 (\rho u)^{m-2} (1 - \rho^2)) u^k \varphi(u) \Phi^{i-1}(u) (1 - \Phi(u))^{n-i} du = \\ &= \rho^m \int_{-\infty}^{\infty} u^{k+m} f_{U_{(i)}}(u) du + C_m^2 \rho^{m-2} (1 - \rho^2) \int_{-\infty}^{\infty} u^{k+m-2} f_{U_{(i)}}(u) du = \\ &= \rho^m EU_{(i)}^{k+m} + C_m^2 \rho^{m-2} (1 - \rho^2) EU_{(i)}^{k+m-2}. \end{aligned}$$

В частности, $EU_{(i)} \tilde{V}_i = \rho EU_{(i)}^2$. Отсюда и из (3) приходим к (6):

$$\text{cov}(U_{(i)}, \tilde{V}_i) = EU_{(i)} \tilde{V}_i - EU_{(i)} E\tilde{V}_i = \rho EU_{(i)}^2 - \rho (EU_{(i)})^2 = \rho DU_{(i)}.$$

Точно также из формул (3.4) - (3.7) и леммы 1 устанавливаем справедливость равенств (5) и (7) ■

Замечание 3. Вернемся к вариационному ряду (1.3), построенному по выборке из распределения $N(a_x, a_y, \sigma_x^2, \sigma_y^2, \rho)$. Как уже отмечалось в §1, линейные преобразования (1.1) связывают ряды (1.3) и (3.2). Следовательно, имеется простая связь между числовыми характеристиками вариационных рядов (1.3) и (3.2). Например, из теорем 2-3 находим:

$$E\tilde{Y}_i = \sigma_y E\tilde{V}_i + a_x = \rho \sigma_y EU_{(i)} + a_y,$$

$$\text{cov}(\tilde{Y}_i, \tilde{Y}_j) = \sigma_y^2 \text{cov}(\tilde{V}_i, \tilde{V}_j) = \rho^2 \sigma_y^2 \text{cov}(U_{(i)}, U_{(j)}), \quad i, j = \overline{1, n}.$$

Примечательно, что числовые характеристики ряда (1.3) выражаются через подобные числовые характеристики одномерной выборки из стандартного нормального распределения.

Определение 2. Двумерную выборку $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ назовем упорядоченной, если в ее вариационном ряде $(X_{(1)}, \tilde{Y}_1), (X_{(2)}, \tilde{Y}_2), \dots, (X_{(n)}, \tilde{Y}_n)$ вторые координаты также возрастают: $\tilde{Y}_1 \leq \tilde{Y}_2 \leq \dots \leq \tilde{Y}_n$.

Результаты § 2 позволяют находить вероятность появления упорядоченной выборки из двумерного нормального распределения. Покажем это, ради простоты, на примере выборки объема 2.

Теорема 4. Выборка $(U_1, V_1), (U_2, V_2) \in N(0, 0, 1, 1, \rho)$ упорядочена с вероятностью:

$$P(\tilde{V}_1 \leq \tilde{V}_2) = 1 - \frac{1}{\pi} \arctg\left(\sqrt{1 - \rho^2} / \rho\right).$$

Доказательство. Так как совместная плотность \tilde{V}_1 и \tilde{V}_2 равна:

$$f(v_1, v_2) = 2! \int_{-\infty}^{\infty} \varphi(u_2, v_2) du_2 \int_{-\infty}^{u_2} \varphi(u_1, v_1) du_1,$$

то из лемм 1-2 получим выражение:

$$\begin{aligned} P(\tilde{V}_1 \leq \tilde{V}_2) &= \int_{-\infty}^{\infty} dv_2 \int_{-\infty}^{v_2} f(v_1, v_2) dv_1 = \\ &= 2! \int_{-\infty}^{\infty} \varphi(v_2) dv_2 \int_{-\infty}^{v_2} \varphi(v_1) \Phi\left(\rho(v_2 - v_1) / \sqrt{2(1 - \rho^2)}\right) dv_1. \end{aligned}$$

Внутренний интеграл после интегрирования по частям примет вид:

$$\int_{-\infty}^{v_2} \Phi\left(\rho(v_2 - v_1) / \sqrt{2} \sqrt{1 - \rho^2}\right) d\Phi(v_1) = \Phi(v_2) / 2 + \int_{-\infty}^0 \varphi(t) \Phi\left(\rho t / \sqrt{2(1 - \rho^2)} + v_2\right) dt.$$

Тогда из последних двух соотношений и леммы 2 окончательно находим:

$$P(\tilde{V}_1 \leq \tilde{V}_2) = \frac{1}{2} + 2 \int_{-\infty}^0 \varphi(t) \Phi\left(t \sqrt{1 - \rho^2} / \rho\right) dt = 1 - \frac{1}{\pi} \arctg\left(\sqrt{1 - \rho^2} / \rho\right).$$

Здесь мы воспользовались равенством:

$$\int_{-\infty}^0 \varphi(t) \Phi(at) dt = \frac{1}{4} - \frac{\arctg(a)}{2\pi},$$

справедливость которого устанавливается также, как и равенства (1.6).

СПИСОК ЛИТЕРАТУРЫ

1. Watterson, G.A. Linear estimation in censored samples from multivariate normal populations // Ann. Math. Statist. 1959. V. 30. С. 814-824.
2. Дейвид Г. Порядковые статистики. М.: Наука, 1979. 336 с.
3. Сархан А.Е., Гринберг Б.Г. Введение в теорию порядковых статистик. М.: Статистика, 1970. 414 с.