11. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ДВУМЕРНЫХ ВЕЛИЧИН

11.1 Смешанные моменты двумерной случайной величины

Смешанным начальным моментом порядка $k+s$ называется математическое ожидание произведения X^k и Y^s:

$$\alpha_{k,s}(x,y) = M\left[X^k Y^s\right]. \tag{11.1}$$

Смешанным центральным моментом порядка $k+s$ называется математическое ожидание произведения центрированных величин $\overset{\circ}{X}{}^k$ и $\overset{\circ}{Y}{}^s$:

$$\mu_{k,s}(x,y) = M\left[\overset{\circ}{X}{}^k \overset{\circ}{Y}{}^s\right], \tag{11.2}$$

где $\overset{0}{X} = M[(X - m_x)]$, $\overset{0}{Y} = M[(Y - m_y)]$- центрированные случайные величины.

Расчетные формулы:

$$\alpha_{k,s}(x,y) = \begin{cases} \sum_{i-1}^{n}\sum_{j=1}^{m} x_i^k y_j^s p_{ij}, & ДСВ \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^s f(x,y)dxdy. & НСВ \end{cases} \tag{11.3}$$

$$\mu_{k,s}(x,y) = \begin{cases} \sum_{i-1}^{n}\sum_{j=1}^{m} (x_i - m_x)^k (y_j - m_y)^s p_{ij}, & ДСВ \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_x)^k (y - m_y)^s f(x,y)dxdy. & НСВ \end{cases} \tag{11.4}$$

где p_{ij} - элементы матрицы вероятностей дискретной величины (X, Y);

$f(x, y)$-совместная плотность вероятности непрерывной величины (X, Y).

Рассмотрим наиболее часто используемые начальные и центральные моменты:

$$\alpha_{0,0}(x, y) = \mu_{0,0}(x, y) = 1;$$

Первые начальные моменты представляют собой уже известные нам ***математические ожидания*** величин *X и У*, входящих в систему:

$$\alpha_{1,0}(x, y) = m_X; \qquad \alpha_{0,1}(x, y) = m_Y. \tag{11.5}$$

Совокупность математических ожиданий представляет собой характеристику положения системы. Геометрически это координаты средней точки на плоскости, вокруг которой происходит рассеивание случайных точек $(X,У)$.

$$\mu_{1,0}(x, y) = M[X-m_x]=0 \qquad \mu_{0,1}(x, y)= M[Y-m_y)=0; \tag{11.6}$$

$$\alpha_{2,0}(x, y) = \alpha_2(x) \qquad \alpha_{0,2}(x, y) = \alpha_2(y)$$

На практике широко используются вторые центральные моменты системы. Два из них представляют собой ***дисперсии***, которые характеризуют ***рассеивание*** случайной точки в направлении осей $0X$ и $0Y$:

$\mu_{2,0}(x, y)$ = M[$(X\text{-}m_x)^2$]=D[X]=D_X; $\mu_{0,2}(x, y)$ = M[$(Y\text{-}m_y)^2$]=D[Y]=D_y;
(11.7)

Особую роль играет центральный момент порядка 1+1 или второй смешанный центральный момент, который называется ***ковариацией или корреляционным моментом***

$$\mu_{1,1}(x, y) = K_{xy} = M[\overset{\circ}{X} \cdot \overset{\circ}{Y}] \tag{11.8}$$

Ковариация представляет собой математическое ожидание произведения центрированных случайных величин X и Y и характеризует степень линейной статистической зависимости величин X и Y и рассеивание относительно точки (m_X, m_y):

$$K_{xy} = \mu_{xy} = M[(X - m_x) \cdot (Y - m_y)],$$

(11.9)

Или

$$K_{xy} = M[(X - m_x) \cdot (Y - m_y)] = M[XY - Xm_y - Ym_x + m_x m_y] = M[XY] - M[Xm_y] - M[Ym_x] + m_x m_y = M[XY] - m_x m_y. \tag{11.10}$$

Расчетные формулы для определения ковариации:

$$K_{xy} = \begin{cases} \sum_i \sum_j (x_i - m_x)(y_j - m_y)p_{ij} = \sum_i \sum_j x_i y_j p_{ij} - m_x m_y. \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_i - m_x)(y_j - m_y) f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy - m_x m_y. \end{cases} \tag{11.11}$$

Свойства корреляции:

1. $K_{xy}=K_{yx}$. Это свойство очевидно.

2. *Корреляционный момент двух независимых случайных величин Х и У равен нулю.*

Доказательство: т.к. случайные величины X и Y – независимы, то и их совместная плотность распределения представляется произведением плотностей распределения случайных величин $f_x(x)$ и $f_y(y)$. Тогда:

$$K_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy - m_x m_y$$
$$= \left(\int_{-\infty}^{\infty} x f_x(x) dx \right) \cdot \left(\int_{-\infty}^{\infty} y f_y(y) dy \right) - m_x m_y = 0.$$

3. *Абсолютная величина корреляционного момента двух случайных величин не превышает среднего геометрического их дисперсий*

$$|K_{xy}| \le \sigma_x \cdot \sigma_y \text{ или } |K_{xy}| \le \sqrt{D_x \cdot D_y}$$

Доказательство: Введем в рассмотрение случайные величины $Z_1 = \sigma_y X - \sigma_x Y$, $Z_2 = \sigma_y X + \sigma_x Y$, и вычислим их дисперсии

$$D(Z_1) = M[Z - m_z]^2 = 2\sigma_x^2 \sigma_y^2 - 2\sigma_x \sigma_y \cdot K_{xy}$$

$$D(Z_2) = M[Z - m_z]^2 = 2\sigma_x^2 \sigma_y^2 + 2\sigma_x \sigma_y \cdot K_{xy}$$

Т. к. дисперсия всегда неотрицательна, то

$$2\sigma_x^2\sigma_y^2 - 2\sigma_x\sigma_y \cdot K_{xy} \ge 0 \Rightarrow \sigma_x\sigma_y \ge K_{xy}$$

и

$$2\sigma_x^2\sigma_y^2 + 2\sigma_x\sigma_y \cdot K_{xy} \ge 0 \Rightarrow -\sigma_x\sigma_y \le K_{xy}.$$

Отсюда $-\sigma_x\sigma_y \le K_{xy} \le \sigma_x\sigma_y \Rightarrow \left|K_{xy}\right| \le \sigma_x\sigma_y$.

Если $K_{xy} \ne 0$, случайные величины X и Y называются *коррелированными.* Если $K_{xy} = 0$, то необязательно, что X и Y независимы. В этом случае они называются *некоррелированными.* Итак, из коррелированности двух случайных величин следует их зависимость, но из зависимости еще не вытекает их коррелированность. Из независимости двух случайных величин следует их некоррелированность, но из некоррелированности еще нельзя заключить о независимости этих величин.

Величина ковариации зависит единиц измерения каждой из случайных величин, входящих в систему и от того, насколько каждая из случайных величин отклоняется от своего математического ожидания (одна – мало, вторая – сильно, все равно K_{xy} будет мал).

Поэтому для характеристики связи между X и Y в чистом виде переходят к безразмерной характеристике, которая называется ***Коэффициент корреляции r_{xy}*** характеризует степень линейной зависимости величин:

$$r_{xy} = \frac{K_{xy}}{\sqrt{D_x D_y}} = \frac{K_{xy}}{\sigma_x\sigma_y} \qquad (11.12)$$

Свойства коэффициента корреляции:

1. *Абсолютная величина коэффициента корреляции двух случайных величин не превышает единицы:* $\left|r_{xy}\right| \le 1$

2. $\left|r_{xy}\right| = 1$ *если* $Y=aX+b$

Доказательство: $M(Y) = M[aX+b] = aM(X)+b$

Подставим в выражение

$$K_{xy} = M\{[X - M(X)][Y - M(Y)]\} = M\{[X - M(X)][aX + b - aM(X) - b]\} =$$

$$= aM\{[X - M(X)][X - M(X)]\} = aM[X - M(X)]^2 = aD(X) = a\sigma_x^2$$

т.к. $[Y - M(Y)] = [aX + b - aM(X) - b]\} == a[X - M(X)]$

Найдем дисперсию Y: $D(Y) = M[Y - M(Y)]^2 = a^2M[X - M(X)]^2 = a^2\sigma_x^2$,

т.е.

$\sigma_y = |a|\sigma_x$, коэффициент корреляции: $r_{xy} = \dfrac{K_{xy}}{\sigma_x\sigma_y} = \dfrac{a\sigma_x^2}{\sigma_x(|a|\sigma_x)} = \dfrac{a}{|a|} \Rightarrow \quad \left|r_{xy}\right| = 1$

Коэффициент корреляции служит для оценки тесноты линейной связи между X и Y: чем ближе абсолютная величина коэффициента корреляции к 1, тем связь сильнее, чем ближе к 0, тем слабее.

3. Если величины X и Y независимы, то $r_{xy} = 0$.

11.2. Условные числовые характеристики

Пусть (X, У) – 2-мерная СВ с известным законом распределения F(X,Y) или f(x,y). *Условным математическим ожиданием компоненты X называется математическое ожидание СВ X, вычисленное при условии, что СВ Y приняла определенное значение Y=y и обозначается M(X/Y).* Аналогично определяется условное математическое ожидание и для СВ *Y*.

Используя формулы для вычисления числовых характеристик случайных величин можно вычислить и условные числовые характеристики, заменив безусловные законы распределения на условные.

$$M(X|Y)=\begin{cases}\sum_{i=1}^{n} x_i\cdot p(x_i/Y=y_j)\\ \int_{-\infty}^{+\infty} x\cdot f_x(x/y)dx\end{cases} \quad \text{и} \quad M(Y|X)=\begin{cases}\sum_{j=1}^{m} y_j\cdot p(y_j/X=x_i)\\ \int_{-\infty}^{+\infty} y\cdot f_y(y/x)dy\end{cases} \qquad (11.13)$$

Условное математическое ожидание СВ *Y* при заданном $X=x$: $M[Y/x]=m_{y/x}$ называется регрессией *Y* на *x*; аналогично $M[X/y]=m_{x/y}$ называется регрессией *X* на *y*. Графики этих зависимостей от *x* и y называются ***линиями регрессии*** или ***«кривыми регрессии».***

Регрессионный анализ позволяет выявить характер связи между величинами. Представим СВ *Y* в виде линейной функции *X*:

$$Y\cong g(x)=\alpha X+\beta,$$

где α и β - неизвестные величины.

Теорема. *Линейная средняя квадратическая регрессия Y на X имеет вид*

$$g(x)=m_y+r_{xy}\frac{\sigma_x}{\sigma_y}(X-m_x),$$

здесь $\beta=r_{xy}\frac{\sigma_x}{\sigma_y}$ - ***коэффициент регрессии Y на X,***

а прямую $y-m_y=r_{xy}\frac{\sigma_x}{\sigma_y}(X-m_x)$ - называют ***прямой среднеквадратической регрессии Y на X.*** Аналогично можно получить прямою среднеквадратической регрессии *X* на *Y*.

Обе прямые проходят через точку (m_x, m_y), которую называют центром совместного распределения величин *X* и *Y*.

Пример 11.1. Определить коэффициент корреляции величин *X* и *Y* (См. пример 10.1).
Решение. Определим математические ожидания величин *X* и *Y* по формуле (11.3):

$$m_X=\alpha_{1,0}(x,y)=\sum_{i=1}^{2}\sum_{j=1}^{3}x_i p_{ij} = 0\cdot 0{,}1+0\cdot 0{,}2+0\cdot 0+1\cdot 0{,}2+$$
$$+1\cdot 0{,}3+1\cdot 0{,}2=0{,}7,$$

$$m_y = \alpha_{0,1}(x,y) = \sum_{i=1}^{2}\sum_{j=1}^{3} y_j p_{ij} = -1\cdot 0{,}1 - 1\cdot 0{,}2 + 0\cdot 0{,}2 + 0\cdot 0{,}3 +$$
$$+ 1\cdot 0 + 1\cdot 0{,}2 = -0{,}1.$$

Найдем значение K_{xy} по формуле (11.11):

$$K_{xy} = \sum_{i=1}^{2}\sum_{j=1}^{3} x_i y_j p_{ij} - m_x m_y = 0\cdot(-1)\cdot 0{,}1 + 0\cdot 0\cdot 0{,}2 + 0\cdot 1\cdot 0 +$$
$$1\cdot(-1)\cdot 0{,}2 + 1\cdot 0\cdot 0{,}3 + 1\cdot 1\cdot 0{,}2 - 0{,}7\cdot(-0{,}1) = 0{,}07.$$

Определим дисперсии величин X и Y по формуле (11.4):

$$D_x = \mu_{2,0}(x,y) = \sum_{i=1}^{2}\sum_{j=1}^{3} (x_i - m_x)^2 p_{ij} = (-0{,}7)\cdot 2\cdot 0{,}1 + (-0{,}7)\cdot 2\cdot 0{,}2 +$$
$$+ (-0{,}7)\cdot 2\cdot 0 + (0{,}3)\cdot 2\cdot 0{,}2 + (0{,}3)\cdot 2\cdot 0{,}3 + (0{,}3)\cdot 2\cdot 0{,}2 = 0{,}21,$$

$$D_y = \mu_{0,2}(x,y) = \sum_{i=1}^{2}\sum_{j=1}^{3} (y_j - m_y)^2 p_{ij} = (-1{,}1)\cdot 2\cdot 0{,}1 + (-1{,}1)\cdot 2\cdot 0{,}2 +$$
$$+ (0{,}1)\cdot 2\cdot 0{,}2 + (0{,}1)\cdot 2\cdot 0{,}3 + (0{,}9)\cdot 2\cdot 0 + (0{,}9)\cdot 2\cdot 0{,}2 = 0{,}49.$$

Значение коэффициента корреляции R_{xy} вычислим по формуле

$$r_{xy} = \frac{K_{XY}}{\sqrt{D_X D_Y}} = \frac{0{,}07}{\sqrt{0{,}21\cdot 0{,}49}} \approx 0{,}22 \qquad (11.14)$$

Пример 11.2. Двумерная случайная величина равномерно распределена в области D, ограниченной прямыми $X = 0$, $Y = 0$ и $X + Y = 4$. Определить коэффициент корреляции величин X и Y.

Решение. Запишем в аналитической форме совместную плотность вероятности:

$$f(x,y) = \begin{cases} c, 0 \le x \le 4, 0 \le y \le 4 - x \\ 0, x \notin [0,4], y \notin [0, 4-x] \end{cases}.$$

Определим c, используя свойства плотности распределения (условие нормировки):

$$\int_0^4 \int_0^{4-x} c\,dx\,dy = c\int_0^4 (4-x)dx = c\cdot 8 = 1 \Rightarrow c = \frac{1}{8}.$$

Найдем математическое ожидание и дисперсию величины X по формулам (10.3) и (10.4) соответственно:

$$m_X = \alpha_{1,0}(x,y) = \int_0^4 \int_0^{4-x} x\frac{1}{8}dxdy = \frac{1}{8}\int_0^4 xdx \int_0^{4-x} dy = \frac{1}{8}\int_0^4 x(4-x)dx = \frac{4}{3};$$

$$D_X = \mu_{2,0}(x,y) = \int_0^4 \int_0^{4-x} (x - m_x)^2 \frac{1}{8}dxdy = \frac{1}{8}\int_0^4 \left(x - \frac{4}{3}\right)^2 (4-x)dx = \frac{8}{9}.$$

Так как область D симметрична относительно осей координат, то величины X и Y будут иметь одинакровые числовые характеристики: $m_X = m_y = 4/3$, $D_X = D_y = 8/9$.

Определим корреляционный момент K_{xy} по формуле (11.11):

$$K_{xy} = \int_0^4 \int_0^{4-x} xy\frac{1}{8}dxdy - m_x\cdot m_y = \frac{1}{8}\int_0^4 xdx\int_0^{4-x} ydy - \left(\frac{4}{3}\right)^2 = \frac{1}{16}\int_0^4 x(4-x)^2dx - \left(\frac{4}{3}\right)^2 = -\frac{4}{9}$$

Коэффициент корреляции величин X и Y будет равен (2.6):

$$r_{xy} = \frac{K_{XY}}{\sqrt{D_X D_Y}} = -\frac{1}{2}$$