ТЕМА 7. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СИСТЕМ СВ.

7.1 Смешанные моменты двумерной случайной величины

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

$$\alpha_{k,s}(x,y) = M\left[X^k Y^s\right]. \qquad (13.1)$$

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

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

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

Расчетные формулы:

$$\alpha_{k,s}(x,y) = \begin{cases} \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_i^k y_j^s p_{ij}, & ДСВ \\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} x^k y^s f(x,y)dxdy. & НСВ \end{cases} \qquad (13.3)$$

$$\mu_{k,s}(x,y) = \begin{cases} \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} (x_i-m_x)^k (y_j-m_y)^s p_{ij}, & ДСВ \\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x-m_x)^k (y-m_y)^s f(x,y)dxdy. & НСВ \end{cases} \qquad (13.4)$$

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:

$$a_{0,0}(x,y) = m_{0,0}(x,y) = 1;$$

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

$$a_{1,0}(x,y) = m_x; \qquad a_{0,1}(x,y) = m_y. \qquad (13.5)$$

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

$$m_{1,0}(x,y) = M[X - m_x] = 0 \qquad m_{0,1}(x,y) = M[Y - m_y) = 0; \qquad (13.6)$$
$$a_{2,0}(x,y) = a_2(x) \qquad a_{0,2}(x,y) = a_2(y)$$

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

$$m_{2,0}(x,y) = M[X - m_x)^2 = D[X] = D_x; \qquad m_{0,2}(x,y) = M[Y - m_y)^2] = D[Y] = D_y; \qquad (13.7)$$

7.2. Ковариация. Коэффициент корреляции.

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

$$m_{1,1}(x,y) = K_{xy} = M[\overset{\circ}{X} \cdot \overset{\circ}{Y}] \qquad (13.8)$$

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

$$K_{xy} = \mu_{xy} = M[(X-m_x)\cdot(Y-m_y)], \qquad (13.9)$$

Или

$$K_{xy} = M[(X-m_x)\cdot(Y-m_y)] = M[XY - Xm_y - Ym_x + m_x m_y] =$$
$$M[XY] - M[Xm_y] - M[Ym_x] + m_x m_y = M[XY] - m_x m_y. \qquad (13.10)$$

Расчетные формулы для определения ковариации:

$$K_{xy} = \begin{cases} \sum\limits_{i}\sum\limits_{j}(x_i-m_x)(y_j-m_y)p_{ij} = \sum\limits_{i}\sum\limits_{j} x_i y_j - m_x m_y. \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}(x_i-m_x)(y_j-m_y)f(x,y)dxdy = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} xyf(x,y)dxdy - m_x m_y. \end{cases}$$

(13.11)

Свойства корреляции:

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

2. *Корреляционный момент двух независимых случайных величин X и Уравен нулю.*

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

$$K_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy - m_x m_y = \left(\int_{-\infty}^{\infty} xf_x(x)dx\right)\cdot\left(\int_{-\infty}^{\infty} yf_y(y)dy\right) - m_x m_y = 0.$$

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

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

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

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

и

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

Отсюда $-\sigma_x\sigma_y \le K_{xy} \le \sigma_x\sigma_y$ Þ $|K_{xy}| \le \sigma_x\sigma_y$.

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} характеризуетстепень линейной зависимости величин:

$$r_{xy} = \frac{K_{xy}}{\sqrt{D_x D_y}} = \frac{K_{xy}}{\sigma_x\sigma_y} \tag{13.12}$$

Свойства коэффициента корреляции:

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

2. $|r_{xy}| = 1$ *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* : $D(Y) = M[Y - M(Y)]^2 = a^2 M[X - M(X)]^2 = a^2\sigma_x^2$, т.е.

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

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 $r_{xy} = 0$.

7 .3. Числовые характеристики системы случайных величин

Как и в случае одномерных случайных величин, при невозможности точно установить законы распределения, применяют приближенное описание системы случайных величин с помощью минимального количества числовых характеристик.

В качестве основных числовых характеристик используем следующие:

1. *Вектор математических ожиданий* M=(m_1, m_2,… m_n) – характеризующих средние значения величин

$$m_i = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} x_i f(x_1,\dots,x_n)dx_1\dots dx_n$$

2. *Вектор дисперсий* $\boldsymbol{D} = (D_1, D_2, \dots D_n)$– характеризующих их рассеивание

$$D_i = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} (x_i - m_i)^2 f(x_1,\dots,x_n)dx_1\dots dx_n$$

3. *Корреляционная матрица*

$$K_{ij} = M\left[\overset{\circ}{X}_i \overset{\circ}{X}_j\right],$$

где $\overset{\circ}{X}_i = X_i - m_i \quad \overset{\circ}{X}_j = X_j - m_j$ или матрица коэффициентов корреляции

$$r_{ij} = \frac{K_{ij}}{\sigma_x \sigma_y},$$

характеризующих попарную корреляцию всех величин, входящих в систему.

Дисперсия каждой из случайных величин X_i есть, по существу, не что иное, как корреляционный момент X_i и той же величины X_i : $D_i = K_{ii} = M[\overset{\circ}{X_i^2}] = M[\overset{\circ}{X_i} \cdot \overset{\circ}{X_i}]$, поэтому корреляционная матрица и матрица коэффициентов корреляции принимает вид:

$$\|K_{ij}\| = \begin{vmatrix} D_1 & K_{12} & \dots & K_{1n} \\ K_{21} & D_2 & \dots & K_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \cdots & D_n \end{vmatrix},$$

где

$$K_{ij} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - m_i)(x_j - m_j) f(x_1, \dots, x_n) dx_1 \dots dx_n$$

В случае, когда случайные величины не коррелированы, все элементы, кроме диагональной, равны 0 и корреляционная матрица принимает вид диагональной матрицы:

$$\|K_{ij}\| = \begin{vmatrix} D_1 & 0 & \dots & 0 \\ 0 & D_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n \end{vmatrix}$$

Помимо корреляционной матрицы для описания систем случайных величин может быть использована матрица коэффициентов корреляции. Это матрица, составленная не из корреляционных моментов, а из коэффициентов корреляции (нормированная корреляционная матрица):

$$\|r_{ij}\| = \begin{vmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{vmatrix},$$

где

$$R_{ij} = \frac{K_{ij}}{\sigma_i \sigma_j} = \frac{K_{ij}}{\sqrt{D_i D_j}}$$

Для некоррелированных случайных величин матрица коэффициентов корреляции вырождается в единичную матрицу:

$$\|r_{ij}\| = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix}.$$

7 .4. Условные числовые характеристики

Пусть (X, У) – 2-мерная СВс 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 Yприняла определенное значение Y= yи 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 наусловные.

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

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 a и b - неизвестные величины.

Теорема. *Линейная средняя квадратическаярегрессия Yна X имеет вид*

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

здесь $\beta = r_{xy} \dfrac{\sigma_x}{\sigma_y}$ ***- коэффициент регрессии Yна X ,***

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

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

Пример 13.1. Определить коэффициент корреляции величин Xи Y(См. пример 13.1).
Решение. Определим математические ожидания величин Xи Yпо формуле (13.3):

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

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

Найдем значение K_{xy} по формуле (13.11):

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

Определим дисперсии величин Xи Yпо формуле (11.4):

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

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

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 13.2.*Двумерная случайная величина равномерно распределена в области D, ограниченной прямыми X= 0, Y= 0 и X+ Y= 4. Определить коэффициент корреляции величин Xи 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 Xпо формулам (10.3) и (10.4) соответственно:

$$m_x = {}_{1,0}(x, y) = \int_0^4 \int_0^{4-x} x\frac{1}{8}dx\,dy = \frac{1}{8}\int_0^4 x\,dx \int_0^{4-x} dy = \frac{1}{8}\int_0^4 x(4 - x)dx = \frac{4}{3};$$

$$D_x = {}_{2,0}(x, y) = \int_0^4 \int_0^{4-x} (x - m_x)^2 \frac{1}{8}dx\,dy = \frac{1}{8}\int_0^4 (x - \frac{4}{3})^2 (4 - x)dx = \frac{8}{9}.$$

Так как область Dсимметрична относительно осей координат, то величины Xи Yбудут иметь одинакровыечисловые характеристики: $m_x = m_y = 4/3$, $D_x = D_y = 8/9$.
Определим корреляционный момент K_{xy} по формуле (11.11):

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

Коэффициент корреляции величин Xи Yбудет равен (2.6):

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

7 .5. Нормальный закон распределения на плоскости

Непрерывная двумерная случайная величина (X, Y) имеет нормальное распределение, если ее совместная плотность вероятности имеет вид

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - r_{xy}^2}} e^{-\frac{1}{2(1-r_{xy}^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2r_{xy}(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]},$$

m_x – математическое ожидание случайной величины X .
m_y – математическое ожидание случайной величины Y .
σ_x – средне квадратичное отклонение СВ X.
σ_y – средне квадратичное отклонение СВ Y .
r_{xy} -- коэффициент корреляции.

Условные законы распределения СВ Xи Yтакже являются нормальными:

$$f_1(x/y) = \frac{1}{\sigma_x\sqrt{2\pi(1 - r_{xy}^2)}} e^{-\frac{1}{2(1-r_{xy}^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{r_{xy}(y-m_y)}{\sigma_y}\right]^2};$$

$$f(y/x) = \frac{1}{\sigma_y\sqrt{2\pi(1 - r_{xy}^2)}} e^{-\frac{1}{2(1-r_{xy}^2)}\left[\frac{(y-m_y)}{\sigma_y^2} - \frac{r_{xy}(x-m_x)}{\sigma_x}\right]^2}.$$

Условные числовые характеристики имеют вид:

$$m[x/y] = m_x + r_{xy}\sigma_x(y - m_y)/\sigma_y; \quad D[x/y] = \sigma_x^2(1 - r_{xy}^2);$$
$$m[y/x] = m_y + r_{xy}\sigma_y(x - m_x)/\sigma_x; \quad D[y/x] = \sigma_y^2(1 - r_{xy}^2).$$

Для системы нормально распределенных случайных величин линии регрессии m[x/ y] и m[y/ x] представляют собой *прямые линии*, т.е. регрессия *линейна*.
Для двумернойнормально распределенной с.в. – если составляющие некоррелированы, то они и независимы, т.е. $r_{xy} = o$ Þ

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}e^{-\frac{1}{2}\left[\frac{(x-m_x)^2}{\sigma_x^2}+\frac{(y-m_y)^2}{\sigma_y^2}\right]}=$$

$$=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}\cdot\frac{1}{\sigma_y\sqrt{2\pi}}e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}=f_1(x)f_2(y)$$

Итак, для нормально распределенных составляющих двумерной случайной величины понятия независимости и некоррелированности равносильны.

Можно показать, что двумерная случайная величина (Х,У) распределена нормально, то Х и Усвязаны линейной крреляционнойзависимостью.