10. Совместные распределения, корреляция, независимые случайные величины

Определение. Рассмотрим две дискретные случайные величины: ξ принимает значения $x_1, x_2, \dots, x_N$, η принимает значения $y_1, y_2, \dots, y_M$ (N и M могут быть и бесконечными). Совместное распределение этих случайных величин есть набор вероятностей

$$p_{ik} = P\{\xi = x_i, \eta = y_k\}$$

(сокращенная запись $\{\xi = x_i, \eta = y_k\}$ означает совместное наступление двух событий $\{\xi = x_i\} \cap \{\eta = y_k\}$).

Основные свойства совместного распределения непосредственно следуют из общих свойств вероятности. Обозначим события $A_i = \{\xi = x_i\}$, $B_k = \{\eta = y_k\}$, которые попарно между собой несовместны, $A_i \cap A_j = \emptyset$, $B_k \cap B_l = \emptyset$ и составляют разбиения множества Ω: $\sum_{i=1}^{N} A_i = \Omega$, $\sum_{k=1}^{M} B_k = \Omega$, тогда

$$\sum_{k=1}^{m} p_{ik} = \sum_{k=1}^{m} P(A_i \cap B_k) = P\left(A_i \cap \sum B_k\right) = P(A_i) = P\{\xi = x_i\} \triangleq p_i$$

и точно также

$$\sum_{i=1}^{N} p_{ik} = P\{\eta = y_k\} \triangleq q_k,$$

кроме того и сумма всех вероятностей удовлетворяет условию нормировки,

$$\sum_{i=1}^{N} \sum_{k=1}^{M} p_{ik} = 1 .$$

Рассмотрим теперь случайные величины ξ , η, принимающие значения в некоторых интервалах вещественной оси.

Определение. Плотностью совместного распределения случайных величин ξ , η называется неотрицательная функция $p_{\xi,\eta}(x, y)$ такая, что для любых $a < b, c < d$

$$P\{a < \xi \le b, c < \eta \le d\} = \int_a^b \int_c^d p_{\xi,\eta}(x, y) dx dy.$$

Для совместной плотности справедливы свойства, аналогичные доказанным выше для дискретного совместного распределения: свертка по одной из переменных дает плотность распределения второй случайной величины. Действительно, функцию распределения можно выразить через совместную плотность,

$$F_\xi(x) = P\{\xi \le x\} = P\{\xi \le x, -\infty < \eta \le \infty\} = \int_{-\infty}^{x} \int_{-\infty}^{\infty} p_{\xi,\eta}(u, y) dy du\,;$$

взяв производную от интеграла с переменным верхним пределом, получим

$$p_\xi(x) = \frac{d}{dx} F_\xi(x) = \int_{-\infty}^{\infty} p_{\xi,\eta}(x, y) dy. \qquad (1)$$

Таким же образом выводим формулу для второй плотности:

$$p_\eta(y) = \int_{-\infty}^{\infty} p_{\xi,\eta}(x,y)dx. \tag{2}$$

Имеет место также условие нормировки:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{\xi,\eta}(x,y)dxdy = 1.$$

Определение. Случайные величины ξ, η называются независимыми, если они принимают свои значения независимо друг от друга, иначе говоря, для любых борелевских множеств A и B события $\{\xi \in A\}$, $\{\eta \in B\}$ независимы:

$$P\{\xi \in A, \eta \in B\} = P\{\xi \in A\}P\{\eta \in B\}.$$

Дискретные случайные величины являются независимыми, если

$$p_{ik} = P\{\xi = x_i, \eta = y_k\} = P\{\xi = x_i\}P\{\eta = y_k\} = p_i q_k;$$

непрерывные случайные величины будут независимыми, если

$$p_{\xi,\eta}(x,y) = p_\xi(x)p_\eta(y).$$

Используя конструкцию совместного распределения, построим распределение для суммы случайных величин, и в частности, для суммы независимых случайных величин.

Теорема 1. Плотность распределения суммы не прерывных случайных величин имеет вид

$$p_{\xi+\eta}(x) = \int_{-\infty}^{\infty} p_{\xi,\eta}(u, x-u)du \tag{3}$$

для независимых случайных величин

$$p_{\xi+\eta}(x) = \int_{-\infty}^{\infty} p_\xi(u)p_\eta(x-u)du \tag{4}$$

Доказательство. Функция распределения случайной величины $\xi + \eta$:

$$F_{\xi+\eta}(x) = P\{\xi + \eta \le x\} = \iint_{u+v\le x} p_{\xi,\eta}(u,v)dudv,$$

или, преобразуя двойной интеграл в повторный (см. Рисунок 1):

$$F_{\xi+\eta}(x) = \int_{-\infty}^{\infty} du \int_{-\infty}^{x-u} p(u,v)dv;$$

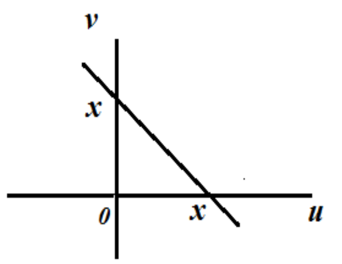

Рисунок 1. К вычислению интеграла

вычислив производную по x, получаем плотность

$$p_{\xi+\eta}(x) = \int\limits_{-\infty}^{\infty} p_{\xi,\eta}(u, x-u)du$$

Для независимых случайных величин получаем выражение (4), этот интеграл (4) называется сверткой распределений.■

Теорема 2. Если случайные величины ξ,η независимы, то имеет место следующее важное свойство математического ожидания:

$$E(\xi\eta) = E\xi \cdot E\eta.$$

Доказательство. Действительно, для дискретных случайных величин ξ,η по определению, имеем

$$E(\xi\eta) = \sum_{i,k} x_i y_k\, p_{ik} = \sum_{i,k} x_i y_k\, p_i q_k = \sum_i x_i\, p_i \sum_k y_k\, q_k = E\xi \cdot E\eta.$$

Для непрерывных же ξ,η:

$$E(\xi\eta) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} xyp_{\xi,\eta}(x,y)dxdy = \int\limits_{-\infty}^{\infty} xp_\xi(x)dx \int\limits_{-\infty}^{\infty} yp_\eta(y)dy = E\xi \cdot E\eta. \blacksquare$$

Следствие из этого свойства: для независимых случайных величин дисперсия суммы равна сумме дисперсий,

$$D(\xi+\eta) = D(\xi) + D(\eta).$$

Действительно, обозначая $a_1 = E\xi$, $a_2 = E\eta$, получаем

$$D(\xi+\eta) = E(\xi+\eta - E(\xi+\eta))^2 = E((\xi - a_1) + (\eta - a_2))^2 =$$

$$= E(\xi - a_1)^2 + 2E(\xi - a_1)(\eta - a_2) \quad + E(\eta - a_2)^2 = D(\xi) + D(\eta),$$

поскольку $E(\xi - a_1)(\eta - a_2) = E(\xi - a_1)E(\eta - a_2) = 0$ по причине независимости ξ и η. Как будет видно в дальнейшем, из этого равенства можно вывести очень существенные следствия.

Отметим еще свойство математического ожидания, необходимое при выполнении операций над двумя случайными величинами: если функция $f(x,y)$ такова, что $f(\xi,\eta)$ является случайной величиной, то

$$Ef(\xi,\eta) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)p_{\xi,\eta}(x,y)dxdy$$

для непрерывных ξ,η и

$$Ef(\xi,\eta) = \sum_{i,k} f(x_i, y_k)\, p_{ik}$$

если ξ,η имеют дискретное распределение. В дискретном варианте это равенство фактически является определением математического ожидания, а на непрерывный случай его можно продолжить, используя свойство σ −непрерывности вероятности.

Пример 1. Применим свойство дисперсии суммы независимых случайных величин для вычисления дисперсии биномиального распределения .

Случайную величину ξ с биномиальным распределением можно представить как сумму n независимых бинарных случайных величин:

$$\xi = \xi_1 + \xi_2 + \cdots + \xi_n,$$
$$P(\xi_i = 1) = p, P(\xi_i = 0) = 1 - p;$$

тогда дисперсия ξ равна сумме дисперсий,

$$D\xi = nD(\xi_1) = np(1 - p).$$

Операция суммирования случайных величин очень часто применяется в приложениях и в теории. Особенно это важно в математической статистике для построения оценок и критериев проверки гипотез.

Совместное распределение двух случайных величин в непрерывном случае мы определили через плотность совместного распределения. Но можно построить также и функцию совместного распределения, как для непрерывных, так и дискретных случайных величин.

Определение. Функцией совместного распределения случайных величин ξ, η называется

$$F_{\xi,\eta}(x, y) = P\{\xi \le x, \eta \le y\}.$$

Для дискретных случайных величин функция распределения очевидным образом выражается через вероятности $p_{ik} = P\{\xi = x_i, \eta = y_k\}$. Для непрерывных случайных величин задана плотность совместного распределения $p_{\xi,\eta}(x, y)$, через которую можно записать

$$F_{\xi,\eta}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_{\xi,\eta}(u, v) du dv.$$

Отметим некоторые свойства совместной функции распределения, прямо следующие из свойств вероятности:

$$F_{\xi,\eta}(-\infty, y) = F_{\xi,\eta}(x, -\infty) = 0 \text{ при всех } x,y;$$

$$F_{\xi,\eta}(\infty, \infty) = 1;$$

$$F_{\xi,\eta}(x, \infty) = F_\xi(x),\ F_{\xi,\eta}(\infty, y) = F_\eta(y);$$

$$p_{\xi,\eta}(u, v) = \frac{\partial^2}{\partial x \partial y} F_{\xi,\eta}(x, y);$$

имеет также место свойство, согласно которому совместная функция распределения $F_{\xi,\eta}(x, y)$ зависит от x,y только через функции распределения, $F_\xi(x)$, $F_\eta(y)$:

$$F_{\xi,\eta}(x, y) = c(F_\xi(x), F_\eta(y)),$$

где $c(u, v)$ – функция распределения, заданная на единичном квадрате; в частности, для независимых случайных величин

$$F_{\xi,\eta}(x, y) = F_\xi(x) F_\eta(y),$$

то есть здесь $c(u, v) = uv$.

Для измерения взаимной зависимости между случайными величинами часто бывает полезен следующий показатель.

Определение. Ковариацией случайных величин ξ, η называется

$$cov(\xi, \eta) = E\{(\xi - E\xi)(\eta - E\eta)\}.$$

Теорема 3. Ковариация обладает следующими свойствами:

1) $cov(\xi, \eta) = E\xi\eta - E\xi \cdot E\eta$;
2) $cov(c\xi, \eta) = cE\xi\eta,\ cov(\xi, \xi) = D\xi$;
3) $D(\xi + \eta) = D\xi + D\eta + 2cov(\xi, \eta)$;
4) для независимых случайных величин $cov(\xi, \eta) = 0$;
5) $|cov(\xi, \eta)| \le \sqrt{D\xi \cdot D\eta}$

Доказательство. Первые четыре свойства вполне очевидно следуют из определения. Чтобы доказать пятое свойство, рассмотрим квадратичное по x (x - произвольное вещественное число) выражение

$$D(\xi + x\eta) = D\xi + x^2 D\eta + 2xcov(\xi, \eta);$$

поскольку оно всегда неотрицательно, то дискриминант квадратного уравнения

$$x^2 + 2x\frac{cov(\xi, \eta)}{D\xi D\eta} + \frac{D\xi}{D\eta} = 0$$

отрицателен,

$$\frac{(cov(\xi, \eta))^2}{(D\eta)^2} - \frac{D\xi}{D\eta} \le 0,$$

следовательно,

$$(cov(\xi, \eta))^2 \le D\xi D\eta,$$

откуда и следует доказываемое неравенство. ■

Это неравенство подсказывает идею ввести нормированную ковариацию: если $cov(\xi, \eta)$ разделить на произведение среднеквадратичных отклонений, максимальное значение такого коэффициента станет равным 1. Такая нормированная ковариация называется корреляцией.

Определение. Коэффициентом корреляции случайных величин ξ, η называется

$$r(\xi, \eta) = \frac{cov(\xi, \eta)}{\sqrt{D\xi D\eta}}.$$

Свойства коэффициента корреляции:

1) $|r(\xi, \eta)| \le 1$;
2) $r(\xi, \eta) = \frac{E\xi\eta - E\xi \cdot E\eta}{\sqrt{D\xi D\eta}}$
3) для независимых случайных величин $r(\xi, \eta) = 0$;
4) максимальное значение коэффициента корреляции, $|r(\xi, \eta)| = 1$, достигается для линейно связанных случайных величин, $\eta = c_1 + c_2\xi$.

Докажем последнее свойство: подставляя $E\eta = c_1 + c_2 E\xi$, $D\eta = c_2^2 D\xi$ в формулу для $r(\xi, \eta)$, находим

$$r(\xi,\eta) = \frac{E(\xi - E\xi)(\eta - E\eta)}{\sqrt{D\xi D\eta}} = \frac{E(\xi - E\xi)c_2(\xi - E\xi)}{\sqrt{c_2^2 D^2(\xi)}} = \frac{c_2 D\xi}{|c_2| D\xi} = sign(c_2). \blacksquare$$