Lecture 7
Multivariate Distributions

Chao Song

College of Ecology

Lanzhou University

September 30, 2024

Multivariate distribution

In many practical cases, it is possible, and often desirable, to take more than one measurement of a random observation. Moreover, we sometimes want to use these measurements to predict a third one. For example, we measure the GPA and extracurriculum activities of a student, and we give each of them a comprehensive evaluation score.

Definition: Let X and Y be two discrete random variables. Let S denote the two-dimensional space of X and Y. The probability that $X = x$ and $Y = y$ is denoted by $f(x, y) = P(X = x, Y = y)$. The function $f(x, y)$ is called the joint probability mass function.

Joint probability mass function

Example: Roll a pair of fair dice. For each of the 36 sampling points with probability $1/36$, let X denote the smaller and Y the larger outcome on the dice. For example, if the outcome is $(3, 2)$, then the observed values are $X = 2,\ Y = 3$. What is the joint PMF of X and Y?

The event $X = 3,\ Y = 3$ can happen in one of two ways $(2, 3)$ or $(3, 2)$. So its probability is $2/36$. However, for event such as $X = 2,\ Y = 2$, it can only happen in one way. Thus, in general, the joint probability mass function is

$$f(x, y) = \begin{cases} \frac{1}{36} & x = y \\ \frac{1}{18} & x \neq y \end{cases}$$

Multinomial distribution

Suppose we have three mutually exclusive and exhaustive ways for an experiment to end: perfect, seconds, and defective. We repeat the experiment n independent times and the probability p_X, p_Y, $1 - p_X - p_Y$ of the three type of results. Let X and Y be the number of perfect and seconds. What is the joint probability mass function of X and Y?

The probability of having x perfects, y seconds, and $n - x - y$ defective is

$$p_X^x p_Y^y (1 - p_X - p_Y)^{n-x-y}$$

And it can be achieved in

$$\mathbf{C}_n^x \mathbf{C}_{n-x}^y = \frac{n!}{x!(n-x)!} \frac{(n-x)!}{y!(n-x-y)!} = \frac{n!}{x!y!(n-x-y)!}$$

Thus, the joint PMF is

$$f(x, y) = \frac{n!}{x!y!(n-x-y)!} p_X^x p_Y^y (1 - p_X - p_Y)^{n-x-y}$$

Marginal probability mass function

Let X and Y have the joint probability function $f(x, y)$ with space S. The probability mass function of X alone is called the marginal probability mass function of X and is defined by

$$f_X(x) = \sum_y f(x, y) \quad x \in S_X$$

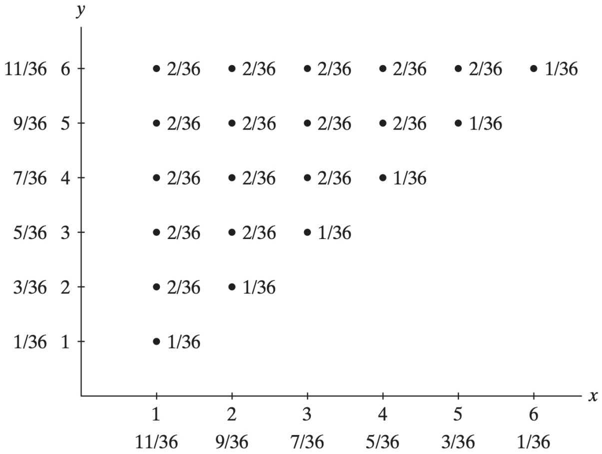
The random variables X and Y are independent if and only if, for every $x \in S_X$ and $y \in S_Y$,

$$f(x, y) = f_X(x) f_Y(y)$$

Otherwise, X ad Y are said to be dependent.

Marginal probability mass function

Example: In the dice rolling example mentioned above, what is the marginal probability mass function of X and Y? Are X and Y independent?



Marginal probability mass function

If X and Y has a multinomial distribution, are they independent?

It is easy to see by logic that X and Y both have a binomial distribution.

$$f_X(x) = \mathbf{C}_n^x p_X^x (1 - p_X)^{n-x}$$

$$f_Y(y) = \mathbf{C}_n^y p_Y^y (1 - p_Y)^{n-y}$$

Therefore,

$$f_X(x) f_Y(y) = \mathbf{C}_n^x \mathbf{C}_n^y p_X^x (1 - p_X)^{n-x} p_Y^y (1 - p_Y)^{n-y} \neq f(xy)$$

Thus, X and Y are not indepenent.

Mathematical expectation

Let X_1 and X_2 be random variables of the discrete type with the joint PMF $f(x_1, x_2)$ on the space S. If $u(X_1, X_2)$ is a function of these two random variables, then

$$E\left[u(X_1, X_2)\right] = \sum_{(x_1, x_2) \in S} \sum u(x_1, x_2) f(x_1, x_2)$$

if it exists, is called the mathematical expectation of $u(X_1, X_2)$.

If $u(X_1, X_2) = X_i$, then $E[u(X_1, X_2)] = E(X_i) = \mu_i$; if $u(X_1, X_2) = (X_i - \mu_i)^2$, then $E[u(X_1, X_2)] = E[(X_i - \mu_i)^2] = Var(X_i)$

Mathematical expectation

Example: There are eight chips in a bow: three marked $(0, 0)$, two marked $(1, 0)$, two marked $(0, 1)$, and one marked $(1, 1)$. A player selects a chip at random and is given the sum of the two coordinates in dollars as a prize. What is the expected prize money a play can get?

Let X_1 and X_2 denote the two coordinates. Their joint PMF is

$$f(x, y) = \frac{3 - x_1 - x_2}{8}, \quad x_1 = 0, 1 \;\; and \;\; x_2 = 0, 1$$

Thus,

$$\begin{aligned} E(X_1 + X_2) &= \sum_{x_2=0}^{1} \sum_{x_1=0}^{1} (x_1 + x_2) \frac{3 - x_1 - x_2}{8} \\ &= (0)(\frac{3}{8}) + (1)(\frac{2}{8}) + (1)(\frac{2}{8}) + (2)(\frac{1}{8}) = \frac{3}{4} \end{aligned}$$

Correlation coefficient

Let $u(X, Y) = (X - \mu_X)(Y - \mu_Y)$, then

$$E[u(X, Y)] = E[(X - \mu_X)(Y - \mu_Y)] = Cov(X, Y) = \sigma_{XY}$$

is called the covariance of X and Y.

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

is called the correlation coefficient of X and Y.

A commonly used formula to calculate covariance:

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y \end{aligned}$$

Correlation coefficient

Example: Let X and Y have the joint PMF

$$f(x,y) = \frac{x+2y}{18}, \quad x = 1,2 \;\; and \;\; y = 1,2$$

What is the correlation coefficient of X and Y?

The marginal PMF are respectively

$$f_X(x) = \sum_{y=1}^{2} \frac{x+2y}{18} = \frac{x+3}{9}$$

$$f_Y(y) = \sum_{x=1}^{2} \frac{x+2y}{18} = \frac{3+4y}{18}$$

The mean and variance of X are

$$\mu_X = \sum_{x=1}^{2} x\frac{x+3}{9} = (1)\frac{4}{9} + (2)\frac{5}{9} = \frac{14}{9}$$

Correlation coefficient

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \sum_{x=1}^{2} x^2 \frac{x+3}{9} - \left(\frac{14}{9}\right)^2 = \frac{20}{81}$$

Similarly, we get the mean and variance of Y

$$\mu_Y = \frac{29}{18} \quad \sigma_Y^2 = \frac{77}{324}$$

The covariance of X and Y

$$\begin{aligned} Cov(X,Y) &= \sum_{x=1}^{2}\sum_{y=1}^{2} xy\frac{x+2y}{18} - \frac{14}{9}\frac{29}{18} \\ &= (1)(1)\frac{3}{18} + (2)(1)\frac{4}{18} + (1)(2)\frac{5}{18} + (2)(2)\frac{6}{18} - \frac{14}{9}\frac{29}{18} \\ &= -\frac{1}{162} \end{aligned}$$

$$\rho = \frac{Cov(X,Y)}{\sigma_X\sigma_Y} = -0.025$$

Correlation coefficient

Proposition: If X and Y are independent, $Cov(X, Y) = 0$.

$$\begin{aligned} E(XY) &= \sum_{S_X} \sum_{S_Y} xyf(x, y) \\ &= \sum_{S_X} \sum_{S_Y} xyf_X(x)f_Y(y) \\ &= \sum_{S_X} xf_X(x) \sum_{S_Y} yf_Y(y) \\ &= \mu_X \mu_Y \end{aligned}$$

Thus, we have

$$Cov(X, Y) = E(XY) - \mu_X \mu_Y = 0$$

Correlation coefficient

If $Cov(X, Y) = 0$, are X and Y necessarily independent?

Example: Let X and Y have the joint PMF

$$f(x, y) = \frac{1}{3}, \quad (x, y) = (0, 1), (1, 0), (2, 1).$$

It is easy to get the marginal PMF of X and Y:

$$f_X(x) = \frac{1}{3},\ x = 0, 1, 2; \quad f_Y(y) = \begin{cases} \frac{1}{3},\ y = 0 \\ \frac{2}{3},\ y = 1 \end{cases}$$

Thus, $\mu_X = 1$ amd $\mu_Y = 2/3$. Then

$$\begin{aligned} Cov(X, Y) &= E(XY) - \mu_X\mu_Y \\ &= (0)(1)\frac{1}{3} + (1)(0)\frac{1}{3} + (2)(1)\frac{1}{3} - (1)\frac{2}{3} \\ &= 0 \end{aligned}$$

It is obvious that $f(x, y) \neq f_X(x)f_Y(y)$. Thus, X and Y are dependent.