Chapters 5. Multivariate Probability Distributions

Random vectors are collection of random variables defined on the same sample space.

Whenever a collection of random variables are mentioned, they are ALWAYS assumed to be defined on the same sample space.

Example of random vectors

1. Toss coin n times, $X_i = 1$ if the i-th toss yields heads, and 0 otherwise. Random variables $X_1, X_2, \ldots, X_n$. Specify sample space, and express the total number of heads in terms of $X_1, X_2, \ldots, X_n$. Independence?

2. Tomorrow's closing stock price for Google.com and Yahoo.com, say (G, Y). Independence?

3. Want to estimate the average SAT score of Brown University Freshmen? Draw a random sample of 10 Freshmen. X_i the SAT for the i-th student. Use sample average
$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{10}}{10}.$$

Description of multivariate distributions

- Discrete Random vector. The joint distribution of (X, Y) can be described by the joint probability function $\{p_{ij}\}$ such that

$$p_{ij} \doteq P(X = x_i, Y = y_j).$$

We should have $p_{ij} \geq 0$ and

$$\sum_i \sum_j p_{ij} = 1.$$

- Continuous Random vector. The joint distribution of (X, Y) can be described via a nonnegative joint density function $f(x, y)$ such that for any subset $A \subset \mathbb{R}^2$,

$$P((X, Y) \in A) = \iint_A f(x, y)dxdy.$$

We should have

$$\iint_{\mathbb{R}^2} f(x, y)dxdy = 1.$$

A general description

The joint cumulative distribution function (cdf) for a random vector (X, Y) is defined as

$$F(x, y) \doteq P(X \leq x, Y \leq y)$$

for $x, y \in \mathbb{R}$.

1. Discrete random vector:

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j)$$

2. Continuous random vector:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du dv$$

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Examples

1. Suppose (X, Y) has a density

$$f(x,y) = \begin{cases} cxye^{-(x+y)} & , \quad \text{if } x > 0, y > 0 \\ 0 & , \quad \text{otherwise} \end{cases}$$

Determine the value of c and calculate $P(X + Y \geq 1)$.

2. In a community, 30% are Republicans, 50% are Democrates, and the rest are indepedent. For a randomly selected person. Let

$$X = \begin{cases} 1 & , \quad \text{if Republican} \\ 0 & , \quad \text{otherwise} \end{cases}$$

$$Y = \begin{cases} 1 & , \quad \text{if Democrat} \\ 0 & , \quad \text{otherwise} \end{cases}$$

Find the joint probability function of X and Y.

3. (X, Y) is the coordinates of a randomly selected point from the disk $\{(x, y) : \sqrt{x^2 + y^2} \leq 2\}$. Find the joint density of (X, Y). Calcualte $P(X < Y)$ and the probability that (X, Y) is in the unit disk $\{(x, y) : \sqrt{x^2 + y^2} \leq 1\}$.

Marginal Distributions

Consider a random vector (X, Y).

1. Discrete random vector: The marginal distribution for X is given by

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}$$

2. Continuous random vector: The marginal density function for X is given by

$$f_X(x) \doteq \int_{\mathbb{R}} f(x, y)\, dy$$

3. General description: The marginal cdf for X is

$$F_X(x) = F(x, \infty).$$

Joint distribution determines the marginal distributions. Not vice versa.

	x_1	x_2	x_3	
y_1	0.1	0.2	0.2	
y_2	0.1	0.1	0.0	
y_3	0.1	0.0	0.2	

	x_1	x_2	x_3	
y_1	0.1	0.2	0.2	
y_2	0.1	0.0	0.1	
y_3	0.1	0.1	0.1	

Example. Consider random variables X and Y with joint density

$$f(x,y) = \begin{cases} e^{-x} & , \quad 0 < y < x \\ 0 & , \quad \text{otherwise} \end{cases}$$

Calculate the marginal density of X and Y respectively.

Conditional Distributions

1. Discrete random vector: Conditional distribution of Y given $X = x_i$ can be described by

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{\text{joint}}{\text{marginal}}$$

2. Continuous random vector: Conditional density function of Y given $X = x$ is defined by

$$f(y|x) \doteq \frac{f(x,y)}{f_X(x)} = \frac{\text{joint}}{\text{marginal}}$$

Remark on conditional probabilities

Suppose X and Y are continuous random variables. One must be careful about the distinction between conditional probability such as

$$P(Y \leq a | X = x)$$

and conditional probability such as

$$P(Y \leq a | X \geq x).$$

For the latter, one can use the usual definition of conditional probability and

$$P(Y \leq a | X \geq x) = \frac{P(X \geq x, Y \leq a)}{P(X \geq x)}$$

But for the former, this is not valid anymore since $P(X = x) = 0$. Instead

$$P(Y \leq a | X = x) = \int_{-\infty}^{a} f(y|x) dy$$

Law of total probability

Law of total probability. When $\{B_i\}$ is a partition of the sample space.

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

Law of total probability. Suppose X is a discrete random variable. For any $G \subset \mathbb{R}^2$, we have

$$P((X,Y) \in G) = \sum_i P((x_i, Y) \in G | X = x_i) P(X = x_i)$$

Law of total probability. Suppose X is a continuous random variable. For any $G \subset \mathbb{R}^2$, we have

$$P((X,Y) \in G) = \int_{\mathbb{R}} P((x, Y) \in G | X = x) f_X(x) dx$$

Examples

1. Toss a coin with probability p of heads. Given that the second heads occurs at the 5th flip, find the distribution, the expected value, and the variance of the time of the first heads.

2. Consider a random vector (X, Y) with joint density

$$f(x,y) = \begin{cases} e^{-x} & , \quad 0 < y < x \\ 0 & , \quad \text{otherwise} \end{cases}$$

Compute

(a) $P(X - Y \geq 2 \,|\, X \geq 10)$,

(b) $P(X - Y \geq 2 \,|\, X = 10)$,

(c) Given $X = 10$, what is the expected value and variance of Y.

3. Let Y be a exponential random variable with rate 1. And given $Y = \lambda$, the distribution of X is Poisson with parameter λ. Find the marginal distribution of X.

Independence

Below are some equivalent definitions.

Definition 1: Two random variables X, Y are said to be independent if for any subsets $A, B \subset \mathbb{R}$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Definition 2: Two random variables X, Y are said to be independent if

1. when X, Y are discrete

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j).$$

2. when X, Y are continuous

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

Remark: Independence if and only if

$$\text{conditional distribution} \equiv \text{marginal distribution}.$$

Remark: Suppose X, Y are independent. Then for any functions g and h, $g(X)$ and $h(Y)$ are also independent

Remark: Two continuous random variables are independent if and only if its density $f(x, y)$ can be written in split-form of

$$f(x, y) = g(x)h(y).$$

See Theorem 5.5 in the textbook. Be VERY careful on the region!

Example. Are X and Y independent, with f as the joint density?

1.

$$f(x, y) = \begin{cases} (x + y) & , \quad 0 < x < 1, 0 < y < 1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

2.

$$f(x, y) = \begin{cases} 6x^2 y & , \quad 0 \le x \le 1, 0 \le y \le 1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

3.

$$f(x, y) = \begin{cases} 8xy & , \quad 0 \le y \le x \le 1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

Examples

1. Suppose $X_1, X_2, \ldots, X_n$ are independent identically distributed (iid) Bernoulli random variables with

$$P(X_i = 1) = p, \qquad P(X_i = 0) = 1 - p.$$

Let

$$Y = X_1 + X_2 + \cdots + X_n.$$

What is the distribution of Y?

Suppose X is distributed as $B(n; p)$ and Y is distributed as $B(m; p)$. If X and Y are independent, what is the distribution of $X + Y$?

2. Suppose X, Y are independent random variables with distribution

$$P(X = k) = P(Y = k) = \frac{1}{5}, \quad k = 1, 2, \ldots, 5$$

Find $P(X + Y \leq 5)$.

3. Suppose X and Y are independent random variables such that X is exponentially distributed with rate λ and Y is exponentially distributed with rate μ. Find out the joint density of X and Y and compute $P(X < Y)$.

4. A unit-length branch is randomly splitted into 3 pieces. What is the probability that the 3 pieces can form a triangle?

5. Suppose X and Y are independent Poisson random variables with parameter λ and μ respectively. What is the distribution of $X+Y$?

Expected values

- Discrete random vector (X, Y):

$$E[g(X,Y)] = \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j).$$

- Continuous random vector (X, Y):

$$E[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y) f(x,y)\, dxdy.$$

Properties of expected values

Theorem. Given any random variables X, Y and any constants a, b,

$$E[aX + bY] = aE[X] + bE[Y].$$

Theorem. Suppose X and Y are independent. Then

$$E[XY] = E[X]E[Y].$$

‘

Examples

1. Randomly draw 4 cards from a deck of 52 cards. Let

$$X_i \doteq \begin{cases} 1 & , \quad \text{if the } i\text{-th card is an Ace} \\ 0 & , \quad \text{otherwise} \end{cases}$$

(a) Are $X_1, \ldots, X_4$ independent?

(b) Are $X_1, \ldots, X_4$ identically distributed?

(c) Find the expected number of Aces in these 4 cards.

2. Let X be the total time that a customer spends at a bank, and Y the time she spends waiting in line. Assume that X and Y have joint density

$$f(x,y) = \lambda^2 e^{-\lambda x}, \quad 0 \le y \le x < \infty, \quad \text{and 0 elsewhere.}$$

Find out the mean service time.

Variance, covariance, and correlation

Two random variables X, Y with mean μ_X, μ_Y respectively. Their Covariance is defined as

$$\mathrm{Cov}(X, Y) \doteq E[(X - \mu_X)(Y - \mu_Y)].$$

Let σ_X and σ_Y be the standard deviation of X and Y. The correlation coefficient of X and Y is defined as

$$\rho \doteq \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- What does correlation mean? [(height, weight), (house age, house price)]
- The correlation coefficient satisfies

$$-1 \leq \rho \leq 1.$$

- $\mathrm{Var}[X] = \mathrm{Cov}(X, X)$.

Theorem. For any random variables X and Y,

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y].$$

Corollary. If X and Y are independent, we have

$$\mathrm{Cov}(X, Y) = 0.$$

Proposition. For any constants a, b, c, d and random variables X, Y,

$$\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y).$$

Proposition.

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$$

Proposition.

$$\mathrm{Cov}(X, Y + Z) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z).$$

Proposition. For any constant a,

$$\mathrm{Cov}(X, a) = 0$$

Theorem.

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y)$$

Corollary. Suppose X and Y are independent. Then

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

Corollary. For any constants a, b, we have

$$\mathrm{Var}[aX+bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] + 2ab\mathrm{Cov}(X,Y).$$

Read Section 5.8 of the textbook for more general versions.

Examples

1. Suppose that X and Y are independent, and $E[X] = 1$, $E[Y] = 4$. Compute the following quantities. (i) $E[2X - Y + 3]$; (ii) $\mathrm{Var}[2X - Y + 3]$; (iii) $E[XY]$; (iv) $E[X^2 Y^2]$; (v) $\mathrm{Cov}(X, XY)$.

2. Find the expected value and variance of $B(n; p)$.

3. (Estimate of μ) Let $X_1, X_2, \ldots, X_n$ are iid random variables. Find the expected value and variance of

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

4. Let X and Y be the coordinates of a point randomly selected from the unit disk $\{(x, y) : x^2 + y^2 \leq 1\}$. Are X and Y independent? Are they uncorrelated (i.e. $\rho = 0$)?

5. (Revisit the 4-card example). Randomly draw 4 cards from a deck of 52 cards. Find the variance of the total number of Aces in these 4 cards.

The multinomial probability distribution

Just like Binomial distribution, except that every trial now has k outcomes.

1. The experiment consists of n identical trials.
2. The outcome of each trial falls into one of k categories.
3. The probability of the outcome falls into category i is p_i, with
$$p_1 + p_2 + \cdots + p_k = 1.$$
4. The trials are independent.
5. Let Y_i be the number of trials for which the outcome falls into category i.

Multinomial random vector $Y = (Y_1, \ldots, Y_k)$. Note that

$$Y_1 + Y_2 + \ldots + Y_k = n.$$

The probability function

For non-negative integers $y_1, \ldots, y_k$ such that $y_1 + \cdots + y_k = n$,

$$P(Y_1 = y_1, \ldots, Y_k = y_k) = \binom{n}{y_1 \; y_2 \; \cdots \; y_k} p_1^{y_1} \cdots p_k^{y_k}$$

Examples

1. Suppose 10 points are chosen uniformly and independently from $[0, 1]$. Find the probability that four points are less than 0.5 and four points are bigger than 0.7?

2. Suppose $Y = (Y_1, Y_2, \ldots, Y_k)$ is multinomial with parameters $(n; p_1, p_2, \ldots, p_k)$.

(a) What is the distribution of, say Y_1?

(b) What is the joint distribution of, say (Y_1, Y_2)?

3. Suppose the students in a high school are 16% freshmen, 14% sophomores, 38% juniors, and 32% seniors. Randomly pick 15 students. Find the probability that exactly 8 students will be either freshmen or sophomores?

Properties of Multinomial Distribution

Theorem. Suppose $Y = (Y_1, Y_2, \ldots, Y_k)$ has multinomial distribution with parameters $(n; p_1, p_2, \ldots, p_k)$. Then

1. $$E[Y_i] = np_i, \qquad \mathrm{Var}[Y_i] = np_i(1 - p_i)$$

2. $$\mathrm{Cov}(Y_i, Y_j) = -np_i p_j, \quad i \neq j.$$

Conditional Expectation

The conditional expectation $E[h(Y)|X = x]$ is

1. When X is discrete:

$$E[h(Y)|X = x_i] = \sum_j h(y_j)P(Y = y_j|X = x_i)$$

2. When X is continuous:

$$E[h(Y)|X = x] = \int_{\mathbb{R}} h(y)f(y|x)dy.$$

Remark. For any constants a and b,

$$E[aY + bZ|X = x] = aE[Y|X = x] + bE[Z|X = x]$$

Remark. If X and Y are independent, then for any x,

$$E[Y|X = x] \equiv E[Y]$$

An important point of view toward conditional expectation.

$E[Y|X]$ is a random variable!

Explanation:

$$E[Y|X = x] \text{ can be viewed as a function of } x$$

$$\Downarrow$$

$$\text{Write } E[Y|X = x] = g(x)$$

$$\Downarrow$$

$$E[Y|X] = g(X) \text{ is a random variable.}$$

Theorem: For any random variables X and Y

$$E[E[Y|X]] = E[Y].$$

Examples

1. Suppose the number of automobile accidents in a certain intersection in one week is Poisson distributed with parameter Λ. Further assume that Λ varies week from week and is assumed to be exponential distributed with rate λ. Find the average number of accidents in a week.

2. Consider a particle jumping around at 3 locations, say A, B, and C. Suppose at each step, the particle has 50% chance to jump to each of the other two locations. If the particle is currently at location A, then on average how many steps it needs to arrive at location B?