

CHAPTERS 5. MULTIVARIATE PROBABILITY DISTRIBUTIONS

EXAMPLE OF RANDOM VECTORS

1. Toss coin n times, $X_i = 1$ if the i -th toss yields heads, and 0 otherwise. Random vector (X_1, X_2, \dots, X_n) . Specify sample space, and express the total number of heads in terms of (X_1, X_2, \dots, 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. $(X_1, X_2, \dots, X_{10})$. Use sample average

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_{10}}{10}.$$

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 $A \subset \mathbb{R}^2$,

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

We should have

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

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. Two cards are drawn randomly from a 52-card deck. Let X denote the number of Aces and Y the number of Kings. Find the joint distribution of (X, Y) , and compute $P(X \geq Y)$.

2. Suppose (X, Y) has a density

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

Find probability $P(X + Y \geq 1)$.

3. The probability that (X, Y) is in the unit disk $\{(x, y) : \sqrt{x^2 + y^2} \leq 1\}$ if

(a) (X, Y) is randomly selected from the disk $\{(x, y) : \sqrt{x^2 + y^2} \leq 2\}$.

(b) (X, Y) is randomly selected from the disk $\{(x, y) : \sqrt{x^2 + y^2} \leq 2\}$ according to a joint density function which is inverse proportional to the distance of (X, Y) from the center of the disk.

MARGINAL DISTRIBUTIONS

Consider a random vector (X, Y) . The joint distribution of (X, Y) contains the distributional information of X or Y alone.

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

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

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

COMMENT: When marginal = 0, the definition can be arbitrary.

UNDERSTANDING CONDITIONAL DENSITY FUNCTION

Consider $P(Y \leq y | X = x)$. Cannot understand this as the usual

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

which requires $P(B) > 0$.

Rather, think about it as the limit, when $\varepsilon \rightarrow 0$, of

$$P(Y \leq y \mid x - \varepsilon \leq X \leq x + \varepsilon)$$

Law of total probability

Recall

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

where $\{B_i\}$ is a partition of the sample space.

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

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

Remark. Law of total probability holds for all random vectors.

EXAMPLES

1. Toss a coin with probability p of heads. Suppose the second heads occurs at the 5th flip. Find the distribution of the time of the first heads.

2. (X, Y) is randomly chosen from the unit disk $\{(x, y) : \sqrt{x^2 + y^2} \leq 1\}$.
Find $f_X(x)$ and the conditional density function of Y given $X = 0$.

3. 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) $P(X - Y \geq 2 \mid X = 10)$,

(b) $P(X - Y \geq 2 \mid X \leq 10)$.

4. (A mixture distribution). Let Y be an 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 for any $x, y \in \mathbb{R}$

$$F(x, y) = F_X(x)F_Y(y)$$

DEFINITION 3: 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: 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.

Remark: Independence if and only if

Conditional distribution \equiv Marginal distribution.

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

EXAMPLES

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

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

Let

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

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) has the joint density below. Is X and Y independent?

(a)

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

(b)

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

(c)

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

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

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

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

4. 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 . Compute $P(X < Y)$.

Remark: Two ways (they are equivalent!). One approach is to write out the joint distribution of (X, Y) and direct computation, the other is to use law of total probability.

Suppose $\{X_1, X_2, \dots\}$ is a sequence of iid (independent identically distributed) random variables exponentially distributed with rate λ . What is the probability

$$P(X_1 + X_2 + \dots + X_n < Y).$$

5. Revisit some old examples where we have already implicitly and intuitively used independence before. (Triangle area less than $1/4$, Buffon's needle)

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) :

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

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 5 cards from a deck of 52 cards. Let

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

- (a) Are X_1, X_2, \dots, X_5 independent?
- (b) Are X_1, X_2, \dots, X_5 identically distributed?
- (c) Find the expected number of Aces in these 5 cards.

2. Suppose the wholesale price X of a certain commodity and the total sales Y have a joint density

$$f(x, y) = \begin{cases} xe^{-xy} & ; \theta < x < \theta + 1, y > 0 \\ 0 & ; \text{otherwise} \end{cases} .$$

Here θ is some positive parameter. Find the average amount of money spent on this commodity.

3. (Revisit the absented minded professor). A professor sign n letters of recommendation for one of his student and put them randomly into n pre-addressed envelopes. Find the expected number of letters that are put into the right envelope.

VARIANCE, COVARIANCE, AND CORRELATION

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

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

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

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

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

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

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

Theorem. For any random variables X and Y ,

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

Corollary. If X and Y are independent, we have

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

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

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

Proposition.

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

Proposition.

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

Proposition. For any constant a ,

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

Theorem.

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

Corollary. Suppose X and Y are independent. Then

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

Corollary. For any constants a, b , we have

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

Read Section 5.8 of the textbook for more general versions.

EXAMPLES

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

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

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

(Estimate of σ^2). Define

$$S^2 \doteq \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Find the expected value of S^2 .

3. 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. $\rho = 0$)?

4. (Revisit the 5-card example). Randomly draw 5 cards from a deck of 52 cards. Let

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

- (a) Are X_1, X_2, \dots, X_5 independent? (DONE)
- (b) Are X_1, X_2, \dots, X_5 identically distributed? (DONE)
- (c) Find the expected number of Aces in these 5 cards. (DONE)
- (d) Find the variance of the total number of Aces in these 5 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 $Y = (Y_1, \dots, Y_k)$. Note that

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

THE PROBABILITY FUNCTION

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

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

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

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

(b) What is the (marginal) 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 $Y = (Y_1, Y_2, \dots, Y_k)$ has multinomial distribution with parameters $(n; p_1, p_2, \dots, p_k)$. Then

1.

$$E[Y_i] = np_i, \quad \text{Var}[Y_i] = np_i(1 - p_i)$$

2.

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

CONDITIONAL EXPECTATION

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

1. Discrete case:

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

2. Continuous case:

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

- The definition of conditional expectation can be easily generalized to cases such as

$$E[h(Y)|X = x, Z = z, \dots]$$

or

$$E[Y + Z|X = x].$$

-

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

- If X and Y are [independent](#), then

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

for any value of X .

An important point of view toward conditional expectation.

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

Explanation:

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

⇓

Write $E[Y|X = x] = g(x)$

⇓

$E[Y|X] = g(X)$ 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 ?