CHAPTER IV
JOINT PROBABILITY DISTRIBUTIONS

Department of Foundation Year
Institute of Technology of Cambodia

LIN Mongkolsery
sery@itc.edu.kh

2021-2022

Contents

Contents

Two Discrete Random Variables

1. $D_x =$ the set of all possible values of rv X.
2. $D = \{(x, y) : x \in D_x,\ y \in D_y\}$

Definition 1

Let X and Y be two drv's defined on the sample space $\mathcal{S}$. The probability that $X = x$ and $Y = y$ is denoted by

$$p(x, y) = P(X = x, Y = y).$$

The function $p(x, y)$ is called **the joint probability mass function** (joint pmf) of X and Y and has the following properties:

1. $0 \leq p(x, y) \leq 1$.
2. $\sum\sum_{(x,y)\in D} p(x, y) = 1$.
3. $P[(X, Y) \in A] = \sum\sum_{(x,y)\in A} p(x, y)$, where $A \subseteq D$.

Two Discrete Random Variables

Example 2

Roll a pair of fair dice. For each of the 36 sample points with probability $1/36$, let X denote the smaller and Y the larger outcome on the dice. The joint pmf of X and Y is given by the probabilities

$$p(x,y) = \begin{cases} \frac{1}{36}, & 1 \le x = y \le 6 \\ \frac{2}{36}, & 1 \le x < y \le 6. \end{cases}$$

Two Discrete Random Variables

Definition 3

The **marginal probability mass function of** X, denoted by $p_X(x)$, is given by

$$p_X(x) = \sum_{y \in D_y} p(x, y) \qquad \text{for each possible value } x.$$

Similarly, the **marginal probability mass function of** Y is

$$p_Y(y) = \sum_{x \in D_x} p(x, y) \qquad \text{for each possible value } y.$$

Two Discrete Random Variables

Example 4

Let the joint pmf of X and Y be defined by

$$p(x, y) = \frac{x+y}{21}, \quad x = 1, 2, 3, \qquad y = 1, 2.$$

Then

$$p_X(x) = \frac{2x+3}{21}, \qquad x = 1, 2, 3,$$

and

$$p_Y(y) = \frac{2+y}{7}, \qquad y = 1, 2.$$

Two Continuous Random Variables

Definition 5

Let X and Y be two crv's. A **joint probability density function** (joint pdf) for these two variables is a function $f(x, y)$ satisfying $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$. Then for any (measurable) set $A \subseteq \mathbb{R}^2$,

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

In particular, if $A = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$, then

$$P\left((X, Y) \in A\right) = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)dxdy.$$

Two Continuous Random Variables

Definition 6

The **marginal probability density functions of** X **and** Y, denoted by $f_X(x)$ and $f_Y(y)$, respectively, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \text{for } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \quad \text{for } -\infty < y < \infty$$

Example 7

Let X and Y have the joint pdf

$$f(x,y) = \frac{4}{3}(1-xy), \quad 0 \le x \le 1, \quad 0 \le y \le 1.$$

Find the marginal pdfs of X and of Y.

Independent Random Variables

Definition 8

Two random variables X and Y are said to be **independent** if for every pair of x and y values

$$p(x, y) = p_X(x).p_Y(y) \quad X, Y \text{ are discrete}$$

or

$$f(x, y) = f_X(x).f_Y(y) \quad X, Y \text{ are continuous.}$$

Otherwise, X and Y are said to be **dependent**.

More Than Two Random Variables

Definition 9

If $X_1, X_2, \ldots, X_n$ are all discrete rv's, the joint pmf of the variables is the function

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n).$$

If the variables are continuous, the joint pdf of $X_1, \ldots, X_n$ is the function $f(x_1, x_2, \ldots, x_n)$ such that for any n intervals $[a_1, b_1], \ldots, [a_n, b_n]$,

$$P(a_1 \leq x_1 \leq b_1, \ldots, a_n \leq x_n \leq b_n) = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n) dx_n \ldots dx_1.$$

More Than Two Random Variables

Example 10

Consider an experiment consisting of n independent and identical trials, in which each trial can result in any one of r possible outcomes. Let $p_i = P(\text{outcome } i \text{ on any particular trial})$, and define random variables by $X_i =$ the number of trials resulting in outcome $i (i = 1, \ldots, r)$. Such an experiment is called a **multinomial experiment**, and the joint pmf of $X_1, \ldots, X_r$ is called the **multinomial distribution**. By using a counting argument analogous to the one used in deriving the binomial distribution, the joint pmf of $X_1, \ldots, X_r$ can be shown to be

$$p(x_1, \ldots, x_r) = \begin{cases} \dfrac{n!}{(x_1!)(x_2!)\ldots(x_r)!} p_1^{x_1} \ldots p_r^{x_r}, & x_i = 0, 1, 2, \ldots \\ 0, & \text{otherwise}, \end{cases}$$

where $x_1 + \cdots + x_r = n$.

Definition 11

The random variables $X_1, X_2, \ldots, X_n$ are said to be **independent** if for every subset $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ of the variables (each pair, each triple, and so on), the joint pmf or pdf of the subset is equal to the product of the marginal pmf's or pdf's.

Example 12

An electronic device runs until one of its three components fails. The lifetimes (in weeks), X_1, X_2, X_3, of these components are independent, and each has the Weibull pdf

$$f(x) = \frac{2x}{25} e^{(-\frac{x}{5})^2}, \qquad 0 < x < \infty.$$

The probability that the device stops running in the first three weeks is equal to $1 - P(X_i > 3, i = 1, 2, 3) = 1 - \prod_{i=1}^{3} P(X_i > 3) = 0.66.$

Conditional Distributions

Definition 13

Let X and Y be two crv's with joint pdf $f(x,y)$ and marginal pdf $f_X(x)$. Then for any x for which $f_X(x) > 0$, the **conditional probability density function of Y given that $X = x$** is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \quad -\infty < y < \infty.$$

If X and Y are discrete, replacing pdf's by pmf's in this definition gives the **conditional probability mass function of Y when $X = x$**.

Example 14

Two components of a minicomputer have the following joint pdf for their useful lifetimes X and Y:

$$f(x) = \begin{cases} xe^{-x(1+y)}, & x \geq 0 \text{ and } y \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

1. What is the probability that the lifetime X of the first component exceeds 3?
2. What are the marginal pdf's of X and Y? Are the two lifetimes independent? Explain.
3. What is the probability that the lifetime of at least one component exceeds 3?
4. What is the probability that the lifetime X exceeds 3, knowing that the lifetime Y is 4?

Contents

Expected Values, Covariance, and Correlation

Definition 15

Let X and Y be jointly distributed rv's with pmf $p(x, y)$ or pdf $f(x, y)$ according to whether the variables are discrete or continuous. Then the expected value of a function $h(X, Y)$, denoted by $E[h(X, Y)]$ or $\mu_{h(X,Y)}$, is given by

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)p(x,y), & X, Y \text{are discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)dxdy, & X, Y \text{are continuous} \end{cases}$$

Covariance

Note: $\mu_X = E(X)$

Definition 16

The **covariance between two rv's** X **and** Y is

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

$$= \begin{cases} \displaystyle\sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y), & X, Y \text{ discrete} \\ \displaystyle\int_{\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)p(x, y)dxdy, & X, Y \text{ continuous} \end{cases}$$

Theorem 1

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

Correlation

Definition 17

The **correlation coefficient of** X **and** Y, denoted by $\mathrm{Corr}(X, Y), \rho_{X,Y}$, or just ρ, is defined by

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X . \sigma_Y}.$$

Theorem 2

1. If a and c are either both positive or both negative,
$$\mathrm{Corr}(aX + b, cY + d) = \mathrm{Corr}(X, Y).$$
2. For any two rv's X and Y, $-1 \leq \mathrm{Corr}(X, Y) \leq 1$.
3. If X and Y are independent, then $\rho = 0$.
4. $\rho = 1$ or -1 iff $Y = aX + b$ for some numbers a and b with $a \neq 0$.

Example 18

An instructor has given a short quiz consisting of two parts. For a randomly selected student, let X = the number of points earned on the first part and Y = the number of points earned on the second part. Suppose that the joint pmf of X and Y is given the table

	$p(x,y)$	y = 0	5	10	15
	0	.02	.06	.02	.10
x	5	.04	.15	.20	.10
	10	.01	.15	.14	.01

1. What is the total expected score $E(X+Y)$?
2. If the maximum of the two scores is recorded, what is the expected recorded score?
3. Compute the covariance and correlation for X and Y.

Contents

The Distribution of the Sample Mean

Definition 19

The rv's $X_1, X_2, \ldots, X_n$ are said to form a (simple) **random sample** of size n if

1. The X_i's are independent random variables.
2. Every X_i has the same probability distribution.

Theorem 3

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ and let $\overline{X} = (X_1 + \cdots + X_n)/n$ (sample mean). Then

1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$
2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

In addition, with $T_0 = X_1 + \cdots + X_n$ (the sample total),

$$E(T_0) = n\mu, \quad V(T_0) = n\sigma^2, \quad \text{and} \quad \sigma_{T_0} = \sqrt{n}\sigma$$

The Central Limit Theorem

Theorem 4

If $X_1, X_2, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$, then

$$\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n).$$

Theorem 5 (The Central Limit Theorem (CLT))

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then if n is sufficiently large, $\overline{X}$ has approximately a normal distribution with $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \sigma^2/n$, and T_0 also has approximately a normal distribution with $\mu_{T_0} = n\mu, \sigma^2_{T_0} = n\sigma^2$. The larger the value of n, the better the approximation.

Rule of Thumb

If $n > 30$, the CLT can be applied.

The Central Limit Theorem

Example 20

The amount of a particular impurity in a batch of a certain chemical product is a random variable with mean value $4.0g$ and standard deviation $1.5g$. If 50 batches are independently prepared, what is the (approximate) probability that the sample average amount of impurity $\overline{X}$ is between 3.5 and $3.8g$?

Contents

The Distribution of a Linear Combination

Definition 21

Given a collection of n random variables $X_1, \ldots, X_n$ and n numerical constants $a_1, \ldots, a_n$, the rv

$$Y = a_1X_1 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i$$

is called the **linear combination** of the X_i's.

The Distribution of a Linear Combination

Theorem 6

Let $X_1, X_2, \ldots, X_n$ have mean values $\mu_1, \ldots, \mu_n$, respectively, and variances $\sigma_1^2, \ldots, \sigma_n^2$, respectively.

1. Whether or not the X_i's are independent,
$$E(a_1X_1 + \cdots + a_nX_n) = a_1E(X_1) + \cdots + a_nE(X_n).$$
2. If $X_1, \ldots, X_n$ are independent,
$$V(a_1X_1 + \cdots + a_nX_n) = a_1^2V(X_1) + \cdots + a_n^2V(X_n)$$
$$\sigma_{a_1X_1+\cdots+a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}.$$
3. For any $X_1, \ldots, X_n$,
$$V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\mathrm{Cov}(X_i, X_j).$$

The Case of Normal Random Variables

Theorem 7

If $X_1, X_2, \ldots, X_n$ are independent, normally distributed rv's (with possibly different means and/or variances), then any linear combination of the X_i's also has a normal distribution.

Example 22

A gas station sells three grades of gasoline: regular, extra, and super. These are priced at \$3.00, \$3.20, and \$3.40 per gallon, respectively. Let X_1, X_2, and X_3 denote the amounts of these grades purchased (gallons) on a particular day. Suppose the X_i's are independent with $\mu_1 = 1000, \mu_2 = 500, \mu_3 = 300, \sigma_1 = 100, \sigma_2 = 80$, and $\sigma_3 = 50$. Suppose that the revenue from sales is $Y = 3.0X_1 + 3.2X_2 + 3.4X_3$. Find $P(Y > 4500)$?