Chapter 2 Multivariate Distributions

2.1 Distributions of Two Random Variables

Bivariate random vector

Definition

A **random variable** is a function from a sample space $\mathcal{C}$ to $\mathcal{R}$.

Definition

An n-dim **random vector** is a function from $\mathcal{C}$ to $\mathcal{R}^n$.

- A 2-dim random vector is also called a bivariate random variable.

Remark: $X = (X_1, X_2)'$ assigns to each element c of the sample space $\mathcal{C}$ exactly one ordered pair of numbers $X_1(c) = x_1$ and $X_2(c) = x_2$.

Example

1. Height and weight of respondent.
2. Fuel consumption and hours on an engine.

Discrete Random Variables

Joint probability mass function

Definition

A **joint probability mass function** $p_{X_1,X_2}(x_1,x_2) = p(X_1 = x_1, X_2 = x_2)$ (or $p(x_1,x_2)$) with space $(x_1,x_2) \in S$ has the properties that

(a) $0 \leq p(x_1,x_2) \leq 1$,

(b) $\sum_{(x_1,x_2)\in S} p(x_1,x_2) = 1$,

(c) $P[(X_1,X_2) \in A] = \sum_{(x_1,x_2)\in A} p(x_1,x_2)$.

Example

A restaurant serves three fixed-price dinners costing $\$7$, $\$9$, and $\$10$. For a randomly selected couple dinning at this restaurant, let
$X_1 =$ the cost of the man's dinner and
$X_2 =$ the cost of the woman's dinner.
The joint pmf of X_1 and X_2 is given in the following table:

		x_1		
		7	9	10
	7	0.05	0.05	0.10
x_2	9	0.05	0.10	0.35
	10	0.00	0.20	0.10

- What is the probability of $P(X_1 \geq 9, X_2 \leq 9)$? $\underline{0.60}$.
- Does man's dinner cost more?

Marginal probability mass function

Definition

Suppose that X_1 and X_2 have the joint pmf $p(x_1, x_2)$. Then the pmf for X_i, denoted by $p_i(\cdot)$, $i = 1, 2$ is the **marginal pmf**.

Note $p_1(x_1) = \sum_{x_2} p(x_1, x_2)$ and $p_2(x_2) = \sum_{x_1} p(x_1, x_2)$.

Example Find the marginal pmf of the previous example.

	x_1	
7	9	10
0.10	0.35	0.55

	x_2	
7	9	10
0.20	0.50	0.30

Example

Let X_1 =Smaller die face, X_2 =Larger die face, when rolling a pair of two dice. The following table shows a partition of the sample space into 21 events.

		x_1					
		1	2	3	4	5	6
	1	$1/36$	0	0	0	0	0
	2	$2/36$	$1/36$	0	0	0	0
x_2	3	$2/36$	$2/36$	$1/36$	0	0	0
	4	$2/36$	$2/36$	$2/36$	$1/36$	0	0
	5	$2/36$	$2/36$	$2/36$	$2/36$	$1/36$	0
	6	$2/36$	$2/36$	$2/36$	$2/36$	$2/36$	$1/36$

Find the marginal pmf's.

Definition

Let $Y = u(X_1, X_2)$. Then, Y is a random variable and

$$E[u(X_1, X_2)] = \sum_{x_1} \sum_{x_2} u(x_1, x_2) p(x_1, x_2)$$

under the condition that

$$\sum_{x_1} \sum_{x_2} |u(x_1, x_2)| p(x_1, x_2)| < \infty$$

Example

Find $E(\max\{X_1, X_2\})$ for the restaurant problem. $\underline{9.65}$.

Continuous Random Variables

Joint density function

A **joint density function** $f_{X_1,X_2}(x_1, x_2)$ (or $f(x_1, x_2)$) with space $(x_1, x_2) \in S$ has the properties that

(a) $f(x_1, x_2) > 0$,

(b) $\int_{(x_1,x_2)\in S} f(x_1, x_2) dx_1 dx_2 = 1$,

(c) $P[(X_1, X_2) \in A] = \int_{(x_1,x_2)\in A} f(x_1, x_2) dx_1 dx_2$.

Example

Let X_1 and X_2 be continuous random variables with joint density function

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 & \text{for } 0 < x_1, x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

1. Find $P(1/4 < X_1 < 3/4; 1/2 < X_2 < 1)$.
2. Find $P(X_1 < X_2)$.
3. Find $P(X_1 + X_2 < 1)$.

Solution:

$$\int_{1/2}^{1}\int_{1/4}^{3/4} 4x_1x_2dx_1dx_2 = 3/8 = 0.375.$$

$$\int_{0}^{1}\int_{0}^{x_2} 4x_1x_2dx_1dx_2 = 1/2 = 0.5.$$

$$\int_{0}^{1}\int_{0}^{1-x_2} 4x_1x_2dx_1dx_2 = 1/6 = 0.167.$$

Marginal probability density function

Suppose that X_1 and X_2 have the joint pdf $f(x_1, x_2)$. Then the pdf for X_i, denoted by $f_i(\cdot)$, $i = 1, 2$ is the **marginal pdf**.

Note: $f_1(x_1) = \int_{x_2} f(x_1, x_2)dx_2$ and $f_2(x_2) = \int_{x_1} f(x_1, x_2)dx_1$.

Example

Find the marginal pdf from the previous problem.

Solution:

$f_1(x) = f_2(x) = 2x.$

Example

Let X_1 and X_2 be continuous random variables with joint density function

$$f(x_1, x_2) = \begin{cases} cx_1x_2 & \text{for } 0 < x_1 < x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

1. Find c.
2. Find $P(X_1 + X_2 < 1)$.
3. Find marginal probability density function of X_1 and X_2.

Solution:

We have $c = 8$ because

$$\int_0^1 \int_{x_1}^1 x_1 x_2 dx_1 dx_2 = 1/8 = 0.125.$$

$$\int_0^{1/2} \int_{x_1}^{1-x_1} 8x_1 x_2 dx_1 dx_2 = 1/6 = 0.167.$$

For the marginal pdf, we have

$$f_{X_1}(x_1) = \int_{x_1}^1 8x_1 x_2 dx_2 = 4x_1 - 4x_1^3,$$

$$f_{X_2}(x_2) = \int_0^{x_2} 8x_1 x_2 dx_1 = 4x_2^3.$$

Let X_1 and X_2 be continuous random variables with joint pdf

$$f(x_1, x_2) = \begin{cases} cx_1x_2 & \text{for } 0 < x_1 < x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

What is $P\{[X_1 < X_2] \cap [X_2 > 4(X_1 - 1/2)^2]\}$?

Solution:
We see $1/4$ is the solution of $x = 4(x - \frac{1}{2})^2$ on $0 < x < 1$. The range of X_2 is $(1/4, 1)$. When $X_2 = x_2$ is given, we next get the range of X_1. By $X_2 = 4(X_1 - 1/2)^2$, we have

$$X_1 = \frac{1}{2} \pm \sqrt{\frac{X_2}{4}}.$$

We determine the lower bound of X_1 is $\frac{1}{2} \pm \sqrt{\frac{X_2}{4}}$ because the intersection of $X_1 = X_2$ and $X_2 = 4(X_1 - 1/2)^2$ is less than $1/2$ when $X_1 \in (0, 1)$. We also have $X_1 < 1$, so the probability is

$$\int_{\frac{1}{4}}^{1} \int_{\frac{1}{2} - \sqrt{\frac{x_2}{4}}}^{x_1} 8x_1x_2 dx_1 dx_2 = 0.974.$$

Expectation – continuous random variables

Let $Y = u(X_1, X_2)$. Then, Y is a random variable and

$$E[u(X_1, X_2)] = \int_{x_1} \int_{x_2} u(x_1, x_2) f(x_1, x_2) dx_2 dx_1$$

under the condition that

$$\int_{x_1} \int_{x_2} |u(x_1, x_2)| f(x_1, x_2) dx_2 dx_1 < \infty$$

Example

Let X_1 and X_2 be continuous random variables with joint density function

$$f(x_1, x_2) = \begin{cases} (36/5)x_1x_2(1 - x_1x_2) & \text{for } 0 < x_1, x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $E(X_1X_2)$.

Solution:

$$\int_0^1 \int_0^1 \frac{36}{5}(x_1^2x_2^2(1 - x_1x_2))dx_1dx_2 = 0.35.$$

Theorem

Let (X_1, X_2) be a random vector. Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ be random variables whose expectations exist. Then for all real numbers k_1 and k_2,

$$E(k_1Y_1 + k_2Y_2) = k_1E(Y_1) + k_2E(Y_2).$$

We also note that

$$Eg(X_2) = \int_{-\infty}^{\infty} g(x_2)f(x_1, x_2)dx_1dx_2 = \int_{-\infty}^{\infty} g(x_2)f_{X_2}(x_2)dx_2.$$

Example 2.1.5 & 2.1.6

Let (X_1, X_2) be a random vector with pdf

$$f(x_1, x_2) = \begin{cases} 8x_1x_2 & 0 < x_1 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Let $Y_1 = 7X_1X_2^2 + 5X_2$ and $Y_2 = X_1/X_2$. Determine $E(Y_1)$ and $E(Y_2)$.

Discrete & Continuous R.V.

Joint cumulative distribution function

Definition

The **joint cumulative distribution function** of (X_1, X_2) is

$$F_{X_1,X_2}(x_1, x_2) = P\left[\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}\right] \quad \text{for all } (x_1, x_2) \in R^2.$$

Relationship with pmf and pdf:

1. Discrete random variables:

$$F_{X_1,X_2}(x_1, x_2) = \sum_{X_1 \leq x_1} \sum_{X_2 \leq x_2} p(x_1, x_2).$$

2. Continuous random variables:

$$F_{X_1,X_2}(x_1, x_2) = \int_0^{x_1} \int_0^{x_2} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2.$$

Joint cumulative distribution function (cont'd)

Definition

The **joint cumulative distribution function** of (X_1, X_2) is

$$F_{X_1,X_2}(x_1, x_2) = P\left[\{X_1 \le x_1\} \cap \{X_2 \le x_2\}\right] \quad \text{for all } (x_1, x_2) \in R^2.$$

Properties:

1. $F(x_1, x_2)$ is nondecreasing in x_1 and x_2.
2. $F(-\infty, x_2) = F(x_1, -\infty) = 0$.
3. $F(\infty, \infty) = 1$.
4. For a rectangle $(a_1, b_1] \times (a_2, b_2]$, we have

$$\begin{aligned} &P\{\,(X_1, X_2) \in (a_1, b_1] \times (a_2, b_2]\,\} \\ =&F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \end{aligned}$$

Example 2.1.1

Consider the discrete random vector (X_1, X_2). Its pmf is given in the following table:

$X_1 \backslash X_2$	0	1	2	3
0	$1/8$	$1/8$	0	0
1	0	$2/8$	$2/8$	0
2	0	0	$1/8$	$1/8$

Find the value of the joint cdf $F(x_1, x_2)$ at $(1, 2)$.
Solution: $\underline{3/4}$.

Example

1. Find the joint cdf of

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 2e^{-x_1-x_2} & 0 < x_1, x_2 < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Solution:

$$F_{X_1,X_2}(x_1,x_2) = \int_0^{x_1}\int_0^{x_2} 2e^{-t_1-t_2}dt_1dt_2 = 2(1-e^{-x_1})(1-e^{-x_2}).$$

2. Find the joint cdf of

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 2e^{-x_1-x_2} & 0 < x_1 < x_2 < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Solution:

$$F_{X_1,X_2}(x_1,x_2) = \int_0^{\min(x_1,x_2)}\int_{t_1}^{x_2} 2e^{-t_1-t_2}dt_2dt_1.$$

Moment generating function (mgf)

Definition

Let $\mathbf{X} = (X_1, X_2)^\top$ be a random vector. If

$$M(t_1, t_2) = \mathsf{E}\left(e^{t_1 X_1 + t_2 X_2}\right)$$

exists for $|t_1| < h_1$ and $|t_2| < h_2$, where h_1 and h_2 are positive, then we call $M(t_1, t_2)$ the **moment generating function** (mgf) of $\mathbf{X} = (X_1, X_2)^\top$.

We may write

$$M(t_1, t_2) = \mathsf{E}\left(e^{t_1 X_1 + t_2 X_2}\right) = \mathsf{E}\left(e^{\mathbf{t}^\top \mathbf{X}}\right)$$

where $\mathbf{t}^\top$ is a row vector (t_1, t_2) and $\mathbf{X}$ is a column vector $(X_1, X_2)^\top$.

Example 2.1.7

Let the continuous-type random variables X and Y have the joint pdf

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

Determine the joint mgf.

Solution:

$$M_{X,Y}(t_1,t_2) = \int_0^\infty \int_x^\infty \exp(t_1 x + t_2 y - y) dy dx = \frac{1}{(1-t_1-t_2)(1-t_2)},$$

provided that $t_1 + t_2 < 1$ and $t_2 < 1$.

Marginal mgf

Recall that

$$M_{X_1,X_2}(t_1,t_2) = \mathsf{E}\left(e^{t_1X_1+t_2X_2}\right).$$

The marginal mgf is given by

$$M_{X_1}(t_1) = \mathsf{E}\left(e^{t_1X_1}\right) = M_{X_1,X_2}(t_1,0),$$

$$M_{X_2}(t_2) = \mathsf{E}\left(e^{t_2X_2}\right) = M_{X_1,X_2}(0,t_2).$$

Let the continuous-type random variables X and Y have the joint pdf

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

Determine the marginal mgf.

Solution:

$$M_{X,Y}(t_1,t_2) = \int_0^\infty \int_x^\infty \exp(t_1 x + t_2 y - y) dy dx = \frac{1}{(1-t_1-t_2)(1-t_2)},$$

provided that $t_1 + t_2 < 1$ and $t_2 < 1$.

$$M_X(t_1) = M_{X,Y}(t_1, 0) = \frac{1}{1-t_1},\ t_1 < 1,$$

$$M_Y(t_2) = M_{X,Y}(0, t_2) = \frac{1}{(1-t_2)^2},\ t_2 < 1.$$

Example 2.1.7 (cont'd)

Let the continuous-type random variables X and Y have the joint pdf

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

Determine the marginal mgf.

Solution:

$$M_X(t_1) = M_{X,Y}(t_1, 0) = \frac{1}{1-t_1},\ t_1 < 1,$$

$$M_Y(t_2) = M_{X,Y}(0, t_2) = \frac{1}{(1-t_2)^2},\ t_2 < 1.$$

Note that

$$f_1(x) = \int_x^\infty e^{-y} dy = e^{-x}, 0 < x < \infty,$$

$$f_2(x) = \int_0^y e^{-y} dx = ye^{-y}, 0 < y < \infty.$$

Fact: It can be shown that

$$\mathsf{E}(XY) = \frac{dM_{X,Y}(t_1, t_2)}{dt_1 dt_2}\Big|_{t_1=0, t_2=0}.$$

Example: Method 1: In the previous example,

$$\mathsf{E}(XY) = \int_0^\infty \int_0^y xye^{-y} dx dy = 3.$$

Method 2:

$$M_{X,Y}(t_1, t_2) = \frac{1}{(1 - t_1 - t_2)(1 - t_2)},$$

$$\frac{dM_{X,Y}(t_1, t_2)}{dt_1 dt_2} = -\frac{t_1 + 3t_2 - 3}{(t_2 - 1)^2(-t_1 - t_2 + 1)^3},$$

where we see $\frac{dM_{X,Y}(t_1, t_2)}{dt_1 dt_2}\Big|_{t_1=0, t_2=0} = 3$ as well.

Chapter 2 Multivariate Distributions

2.2 Transformation: Bivariate Random Variables

Transformation of discrete random vectors

- Assume there is a one to one mapping between $X = (X_1, X_2)^\top$ and $Y = (Y_1, Y_2)^\top$:

$$Y_1 = u_1(X_1, X_2), \qquad X_1 = w_1(Y_1, Y_2),$$
$$Y_2 = u_2(X_1, X_2), \qquad X_2 = w_2(Y_1, Y_2).$$

- Transformation of **discrete** random variable:

$$p_{Y_1,Y_2}(y_1, y_2) = p_{X_1,X_2}(w_1(y_1, y_2), w_2(y_1, y_2)).$$

Example 2.2.1

Let X and Y be independent random variables such that

$$p_X(x) = \frac{\mu_1^x}{x!} e^{-\mu_1}, \qquad x = 0, 1, 2, \ldots.$$

and

$$p_Y(y) = \frac{\mu_2^y}{y!} e^{-\mu_2}, \qquad y = 0, 1, 2, \ldots.$$

- Find the pmf of $U = X + Y$.
- Determine the mgf of U.

Transformation of continuous random variables

Let J denote the **Jacobian** of the transformation. This is the determinant of the 2×2 matrix

$$\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix}$$

The determinant is $J(y_1, y_2) = \frac{\partial x_1}{\partial y_1} \cdot \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \cdot \frac{\partial x_2}{\partial y_1}$.

Transformation formula: The joint pdf of the **continuous** random vector $Y = (Y_1, Y_2)^\top$ is

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2)}(w_1(y_1, y_2), w_2(y_1, y_2)) \cdot |J(y_1, y_2)|.$$

Notice the bars around the function J, denoting absolute value.

Example

A device containing two key components fails when, and only when, both components fail. The lifetimes, X_1 and X_2, of these components have a joint pdf $f(x_1, x_2) = e^{-x_1 - x_2}$, where $x_1, x_2 > 0$ and zero otherwise. The cost Y_1, of operating the device until failure is $Y_1 = 2X_1 + X_2$.

1. Find the joint pdf of Y_1, Y_2 where $Y_2 = X_2$.
2. Find the marginal pdf for Y_1 (Ans: $e^{-y_1/2} - e^{-y_1}$, for $y_1 > 0$)

Example 2.2.5

Suppose (X_1, X_2) has joint pdf

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 10x_1x_2^2 & 0 < x < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Let $Y_1 = X_1/X_2$ and $Y_2 = X_2$. Find the joint and marginal pdf's of Y_1 and Y_2.

Solution sketch

1. One to one transformation:

$$y_1 = x_1/x_2, \quad y_2 = x_2, \qquad 0 < x_1 < x_2 < 1$$
$$x_1 = y_1 y_2, \quad x_2 = y_2, \quad 0 < y_1 < 1,\ 0 < y_2 < 1.$$

2. Give the joint pdf:

$f_{Y_1,Y_2}(y_1, y_2) = 10 y_1 y_2 y_2^2 |y_2|$, where y is defined above or 0 elsewhere.

3. Give the marginal pdf of Y_1:

$$f_{Y_1}(y_1) = \int_0^1 f_{Y_1,Y_2}(y_1, y_2) dy_2 = 2y_1,\ 0 < y_1 < 0.$$

Example 2.2.4

Suppose (X_1, X_2) has joint pdf

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} \frac{1}{4}\exp(-\frac{x_1+x_2}{2}) & 0 < x_1 < \infty, 0 < x_2 < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Let $Y_1 = 1/2(X_1 - X_2)$ and $Y_2 = X_2$. Find the joint and marginal pdf's of Y_1 and Y_2.

Solution sketch

1. One to one transformation:

$$y_1 = \frac{1}{2}(x_1 - x_2), \quad y_2 = x_2, \quad 0 < x_1 < \infty, \;\; 0 < x_2 < \infty.$$
$$x_1 = 2y_1 + y_2, \qquad x_2 = y_2, \qquad -2y_1 < y_2, \; 0 < y_2 < \infty.$$

2. Give the joint pdf:

$f_{Y_1,Y_2}(y_1, y_2) = e^{-y_1 - y_2}/4 \times |2|$, where y is defined above or 0 elsewhere.

3. Give the marginal pdf of Y_1:

$$f_{Y_1}(y_1) = \begin{cases} \int_{-2y_1}^{\infty} f_{Y_1,Y_2}(y_1, y_2) dy_2 = e^{y_1}/2, \; -\infty < y_1 < 0, \\ \int_{0}^{\infty} f_{Y_1,Y_2}(y_1, y_2) dy_2 = e^{-y_1}/2, \; 0 \le y_1 < \infty, \end{cases}$$

which gives $f_{Y_1}(y_1) = e^{-|y_1|}, \; -\infty < y < \infty$.

Example 2.2.7

Suppose (X_1, X_2) has joint pdf

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} \frac{1}{4}\exp(-\frac{x_1+x_2}{2}) & 0 < x_1 < \infty, 0 < x_2 < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Let $Y_1 = 1/2(X_1 - X_2)$. What is the mgf of Y_1?

Solution sketch

$$
\begin{aligned}
\mathsf{E}(e^{tY}) &= \int_0^\infty \int_0^\infty e^{t(x_1-x_2)/2}\frac{1}{4}e^{-(x_1+x_2)/2}dx_1dx_2 \\
&= \left[\int_0^\infty \frac{1}{2}e^{-x_1(1-t)/2}dx_1\right]\left[\int_0^\infty \frac{1}{2}e^{-x_2(1+t)/2}dx_2\right] \\
&= \left[\frac{1}{1-t}\right]\left[\frac{1}{1+t}\right] \\
&= \frac{1}{1-t^2},
\end{aligned}
$$

provided that $1-t>0$ and $1+t>0$. This is equivalent to

$$
\int_{-\infty}^\infty e^{tx}\frac{e^{-|x|}}{2} = \frac{1}{1-t^2},\ -1<t<1,
$$

which is the mgf of double exponential distribution.

Chapter 2 Multivariate Distributions

2.3 Conditional Distributions and Expectations

Conditional probability for discrete r.v.

Motivating example

Let X_1 =Smaller die face, X_2 =Larger die face, when rolling a pair of two dice. The following table shows a partition of the sample space into 21 events.

		x_1					
		1	2	3	4	5	6
	1	$1/36$	0	0	0	0	0
	2	$2/36$	$1/36$	0	0	0	0
x_2	3	$2/36$	$2/36$	$1/36$	0	0	0
	4	$2/36$	$2/36$	$2/36$	$1/36$	0	0
	5	$2/36$	$2/36$	$2/36$	$2/36$	$1/36$	0
	6	$2/36$	$2/36$	$2/36$	$2/36$	$2/36$	$1/36$

Recalling our definition of conditional probability for events, we have (for example)

$$P(X_2 = 4|X_1 = 2) = \frac{P[\{X_1 = 2\} \cap \{X_2 = 4\}]}{P(X_1 = 2)} = \frac{2/36}{9/36} = \frac{2}{9}.$$

- Recall that for two events A_1 and A_2 with $P(A_1) > 0$, the conditional probability of A_2 given A_1 is

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}.$$

- Let X_1 and X_2 denote **discrete** random variables with joint pmf $p_{X_1,X_2}(x_1, x_2)$ and marginal pmfs $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$. Then for every x_1 such that $p_{X_1}(x_1) > 0$, we have

$$P(X_2 = x_2|X_1 = x_1) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)}.$$

We use a simple notation:

$$p_{X_2|X_1}(x_2|x_1) = p_{2|1}(x_2|x_1) = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)}.$$

- We call $p_{X_2|X_1}(x_2|x_1)$ the **conditional pmf** of X_2, given that $X_1 = x_1$.

Verify $p_{X_2|X_1}(x_2|x_1)$ satisfies the condition of being a pmf.

1. $p_{X_2|X_1}(x_2|x_1) \geq 0$.
2. $$\begin{aligned}\sum_{x_2} p_{X_2|X_1}(x_2|x_1) &= \sum_{x_2} \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)} \\ &= \frac{1}{p_{X_1}(x_1)} \sum_{x_2} p_{X_1,X_2}(x_1,x_2) \\ &= \frac{p_{X_1}(x_1)}{p_{X_1}(x_1)} = 1.\end{aligned}$$

Conditional expectation of discrete random variables:

$$\mathsf{E}(X_1|X_2 = x_2) = \sum_{x_1} x_1 p_{X_1|X_2}(x_1|x_2).$$

Example

Returning to the previous example, it is straightforward to work out the conditional pmf as well as associated functions like expectations. For instance,

$$p_{X_1|X_2}(x_1|X_2 = 3) = \begin{cases} 2/5 & \text{if } x_1 = 1, 2 \\ 1/5 & \text{if } x_1 = 3 \\ 0 & \text{if } x_1 = 4, 5, 6. \end{cases}$$

and $E(X_1|X_2 = 3) = 9/5$.

Conditional probability for continuous r.v.

- Let X_1 and X_2 denote **continuous** random variables with joint pdf $f_{X_1,X_2}(x_1,x_2)$ and marginal pmfs $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$. Then for every x_1 such that $f_{X_1}(x_1) > 0$, we define

$$f_{X_2|X_1}(x_2|x_1) = f_{2|1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}.$$

- Verify that $f_{X_2|X_1}$ satisfies the conditions of being a pdf.

$$(1) \qquad f_{X_2|X_1}(x_2|x_1) \geq 0.$$

$$\begin{aligned}(2) \quad \int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1)dx_2 &= \int_{-\infty}^{\infty} \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} dx_2 \\ &= \frac{1}{f_{X_1}(x_1)} \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)dx_2 \\ &= \frac{f_{X_1}(x_1)}{f_{X_1}(x_1)} = 1.\end{aligned}$$

Conditional expectation of continuous random variables

- If $u(X_2)$ is a function of X_2, the conditional expectation of $u(X_2)$, given that $X_1 = x_1$, if it exists, is given by

$$\mathsf{E}[u(X_2)|x_1] = \int_{-\infty}^{\infty} u(x_2) f_{2|1}(x_2|x_1)\, dx_2.$$

- If they do exist, then $\mathsf{E}(X_2|x_1)$ is the **conditional mean** and

$$\mathsf{Var}(X_2|x_1) = \mathsf{E}\{[X_2 - E(X_2|x_1)]^2|x_1\}$$

is the **conditional variance** of X_2, given $X_1 = x_1$.

Example

Find the conditionals $f_{X_2|X_1}$ and $f_{X_1|X_2}$ for (X_1, X_2) with joint cdf

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} 2e^{-x_1-x_2} & 0 < x_1 < x_2 < \infty \\ 0 & \text{otherwise.} \end{cases}$$

- Calculate $P\left(a < X_2 \leq b | X_1 = x_1\right)$.
- Calculate the expectation $\mathsf{E}\left[u(X_2) | X_1 = x_1\right]$.
- Calculate the variance $\mathsf{Var}\left(X_2 | X_1 = x_1\right)$.

Example (2.3.1)

Let X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} 2 & 0 < x_1 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find $P\left(0 < X_1 < \frac{1}{2} \,|\, X_2 = \frac{3}{4}\right)$ and Var $(X_1|x_2)$.

Example (2.3.2)

Let X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} 6x_2 & 0 < x_2 < x_1 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

1. Compute $\mathsf{E}(X_2)$.
2. Compute the function $h(x_1) = \mathsf{E}\,(X_2|x_1)$. Then compute $\mathsf{E}\,[h(X_1)]$ and Var $[h(X_1)]$.

Theorem 2.3.1

Let (X_1, X_2) be a random vector. Then

(a) $\mathsf{E}\left[\mathsf{E}\left(X_2|X_1\right)\right] = \mathsf{E}\left(X_2\right)$,

(b) $\mathsf{Var}(X_2) = \mathsf{Var}\left[\,\mathsf{E}(X_2|X_1)\,\right] + \mathsf{E}\left[\mathsf{Var}(X_2|X_1)\right]$.

Interpretation:

- Both X_2 and $\mathsf{E}(X_2|X_1)$ are unbiased estimator of $\mathsf{E}(X_2) = \mu_2$.
- The part (b) shows that $\mathsf{E}(X_2|X_1)$ is more reliable.
- We will talk more about this when studying sufficient statistics in Chapter 7, Rao and Blackwell Theorem.

$$\mathsf{E}\left[\mathsf{E}\left(X_2|X_1\right)\right] = \mathsf{E}\left(X_2\right).$$

Proof.

The proof is for the continuous case. The discrete case is proved by using summations instead of integrals. We see

$$\begin{aligned}
\mathsf{E}(X_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1, x_2)dx_2dx_1 \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x_2 \frac{f(x_1, x_2)}{f_1(x_1)}dx_2\right] f_1(x_1)dx_1 \\
&= \int_{-\infty}^{\infty} \mathsf{E}(X_2|x_1) f_1(x_1)dx_1 \\
&= \mathsf{E}[\mathsf{E}(X_2|X_1)].
\end{aligned}$$

□

$$\mathsf{Var}(X_2) = \mathsf{Var}\left[\,\mathsf{E}(X_2|X_1)\,\right] + \mathsf{E}\left[\mathsf{Var}(X_2|X_1)\right].$$

Proof.

The proof is for both the discrete and continuous cases:

$$\begin{aligned}\mathsf{E}[\mathsf{Var}(X_2|X_1)] &= \mathsf{E}[\mathsf{E}(X_2^2|X_1) - (\mathsf{E}(X_2|X_1))^2] \\ &= \mathsf{E}[\mathsf{E}(X_2^2|X_1)] - \mathsf{E}[\mathsf{E}(X_2|X_1)^2] \\ &= \mathsf{E}(X_2^2) - \mathsf{E}[\mathsf{E}(X_2|X_1)^2];\end{aligned}$$

$$\begin{aligned}\mathsf{Var}[\mathsf{E}(X_2|X_1)] &= \mathsf{E}[\mathsf{E}(X_2|X_1)^2] - \{\mathsf{E}[\mathsf{E}(X_2|X_1)]\}^2 \\ &= \mathsf{E}[\mathsf{E}(X_2|X_1)^2] - [\mathsf{E}(X_2)]^2.\end{aligned}$$

Thus,

$$\mathsf{E}[\mathsf{Var}(X_2|X_1)] + \mathsf{Var}[\mathsf{E}(X_2|X_1)] = \mathsf{E}(X_2^2) - [\mathsf{E}(X_2)]^2 = \mathsf{Var}(X_2).$$

□

We further see that

$$\mathsf{Var}\left[\,\mathsf{E}(X_2|X_1)\,\right] \le \mathsf{Var}(X_2).$$

Example 2.3.3

Let $X_1$and X_2 be discrete random variables. Suppose the conditional pmf of X_1 given X_2 and the marginal distribution of X_2 are given by

$$p(x_1|x_2) = \binom{x_2}{x_1}\left(\frac{1}{2}\right)^{x_2},\ x_1 = 0, 1, \ldots, x_2,$$
$$p(x_2) = \frac{2}{3}\left(\frac{1}{3}\right)^{x_2-1},\ x_2 = 1, 2, 3 \ldots.$$

Determine the mgf of X_1.

Example

Assume that the joint pdf for $X_2|X_1 = x_1$ on the support $\mathcal{S} = \{0 < x_1 < 1, 0 < x_2 < 2, x_1 + x_2 < 2\}$ is

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} \dfrac{2x_1}{2 - x_1} & \text{in } \mathcal{S}, \\ 0 & \text{otherwise.} \end{cases}$$

Find $\mathrm{E}(X_2)$ through $\mathsf{E}(X_2) = \mathsf{E}[\mathsf{E}(X_2|X_1)]$.

Solution:

The conditional pdf for $X_2|X_1 = x_1$, $0 < x_1 < 1$ is

$$f_{X_2|X_1}(x_2|x_1) = \begin{cases} 1/(2-x_1) & \text{if } 0 < x_2 < 2 - x_1 \\ 0 & \text{otherwise.} \end{cases}$$

and the marginal pdf for X_1 is $f_{X_1}(x_1) = 2x_1$ for $0 < x_1 < 1$ and zero otherwise.

$$\mathrm{E}(X_2|X_1) = \int_0^{2-x_1} x_2 \frac{1}{2-x_1} dx_2 = \frac{2-x_1}{2},$$

$$\mathrm{E}(\mathrm{E}(X_2|X_1)) = \int_0^1 \frac{2-x_1}{2} 2x_1 dx_1 = 2/3.$$

We can verify this by

$$\mathrm{E}(X_2) = \int_0^1 \int_0^{2-x_1} x_2 \frac{2x_1}{2-x_1} dx_2 dx_1 = 2/3.$$

Chapter 2 Multivariate Distributions

2.4 The Correlation Coefficient

Recall the definition of the variance of X:

$$\mathsf{Var}(X) = \mathsf{E}[(X-\mu)^2].$$

Definition

Let X and Y be two random variables with expectations $\mu_1 = \mathsf{E}X$ and $\mu_2 = \mathsf{E}Y$, respectively. The **covariance** of X and Y, if it exists, is defined to be

$$\mathsf{Cov}(X, Y) = \mathsf{E}\left[(X-\mu_1)(Y-\mu_2)\right].$$

Computation shortcut:

$$\mathsf{E}\left[(X-\mu_1)(Y-\mu_2)\right] = \mathsf{E}(XY) - \mu_1\mu_2.$$

Example 2.4.1

Let X and Y be two random variables with joint pdf

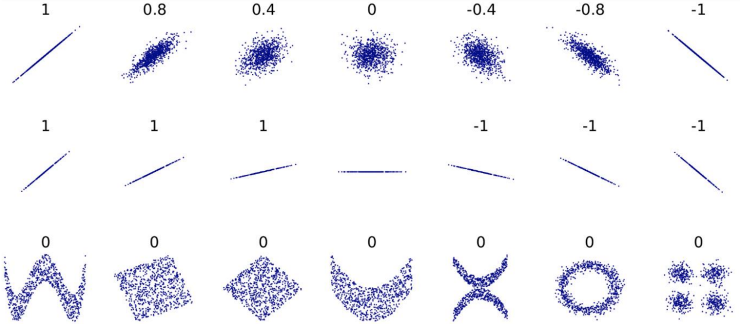
$$f(x,y) = \begin{cases} x+y & 0<x,y<1 \\ 0 & \text{elsewhere.} \end{cases}$$

Determine the covariance of X and Y.

Definition

The **correlation coefficient** of X and Y is defined to be

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$



Example

What is the correlation coefficient of the previous example?

The plot is from Wikipedia https://en.wikipedia.org/wiki/Correlation_and_dependence

Linear conditional mean

For two random variables X and Y, write $u(x) = E(Y|x)$:

$$\mathsf{E}(Y|x) = \int_{-\infty}^{\infty} y f_{2|1}(y|x) dy = \frac{\int_{-\infty}^{\infty} y f_{X,Y}(x,y) dy}{f_1(x)}.$$

If $u(x)$ is a linear function of x, say

$$u(x) = E(Y|x) = a + bx,$$

then we say that the conditional mean of Y is linear in x. The following theorem gives the values of a and b.

Theorem 2.4.1

Let X and Y be two random variables , with means μ_1, μ_2, variances σ_1^2, σ_2^2, and correlation coefficient ρ. If the conditional mean of Y is linear in x, then

$$\mathsf{E}(Y|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}\left(X - \mu_1\right),$$

$$\mathsf{E}\left[\mathsf{Var}\left(Y|X\right)\right] = \sigma_2^2\left(1 - \rho^2\right).$$

Example 2.4.2

Let X and Y have the linear conditional means

$$\mathsf{E}(Y|x) = 4x + 3$$

and

$$\mathsf{E}(X|y) = \frac{1}{16}y - 3.$$

What are the values of μ_1, μ_2, ρ, and σ_2/σ_1?

Recall that the mgf of the random vector (X, Y) is defined to be $M(t_1, t_2) = E\left[e^{t_1 X + t_2 Y}\right]$. It can be shown that

$$\frac{\partial^{k+m}}{\partial t_1^k \partial t_2^m} M(t_1, t_2) = E\left[X^k Y^m e^{t_1 X + t_2 Y}\right].$$

$$\left.\frac{\partial^{k+m}}{\partial t_1^k \partial t_2^m} M(t_1, t_2)\right|_{t_1 = t_2 = 0} = E\left[X^k Y^m\right].$$

- $\mu_1 = \mathsf{E}(X) = \frac{\partial M(0,0)}{\partial t_1}$
- $\mu_2 = \mathsf{E}(Y) = \frac{\partial M(0,0)}{\partial t_2}$
- $\mathsf{Var}(X) = \mathsf{E}(X^2) - \mu_1^2 = \frac{\partial^2 M(0,0)}{\partial t_1^2} - \mu_1^2$
- $\mathsf{Var}(Y) = \mathsf{E}(Y^2) - \mu_2^2 = \frac{\partial^2 M(0,0)}{\partial t_2^2} - \mu_2^2$
- $\mathsf{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y) = \frac{\partial^2 M(0,0)}{\partial t_1 \partial t_2} - \mu_1 \mu_2$
- $\rho = \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{Var}(X)}\sqrt{\mathsf{Var}(Y)}}$

Example 2.4.4

Let X and Y be two random variables with joint pdf

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

Determine the correlation coefficient of X and Y.

Solution:
The mgf is

$$M(t_1,t_2) = \frac{1}{(1-t_1-t_2)(1-t_2)},\ t_1+t_2<1,\ t_2<1.$$

We have $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, Cov$(X,Y) = 1$.

Chapter 2 Multivariate Distributions

2.5 Independent Random Variables

Motivation

Suppose the bivariate random variables (X_1, X_2) is continuously distributed, and for all $x_1 \in S_{X_1}$, and $x_2 \in S_{X_2}$,

$$f_{X_1|X_2}(x_1|x_2) = f_{X_1}(x_1). \tag{1}$$

Since, by the definition of conditional pdf,

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)},$$

it follows that

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \text{ for all } x_1 \in S_{X_1}, x_2 \in S_{X_2}. \tag{2}$$

Clearly (1) and (2) are equivalent. Exactly the same logic applies for a discrete random variable.

Definition of independence

We say two random variables X_1 and X_2 are **independent** if

- (Continuous case) their joint pdf is equal to the product of their marginal pdf's:

$$f(x_1, x_2) \equiv f_1(x_1) f_2(x_2).$$

- (Discrete case) their joint pmf is equal to the product of their marginal pmf's:

$$p(x_1, x_2) \equiv p_1(x_1) p_2(x_2).$$

Immediate indicators of dependency

Suppose that X_1 and X_2 have a joint support $\mathcal{S} = \{(x_1, x_2)\}$ and marginal supports $\mathcal{S}_1 = \{x_1\}$ and $\mathcal{S}_2 = \{x_2\}$. If X_1 and X_2 are independent, then

$$\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2.$$

In other words,

- (Continuous case) If the joint support $\mathcal{S}$ is not a rectangle, then X_1 and X_2 are dependent.
- (Discrete case) If there is a zero entry in the table of pmf, then X_1 and X_2 are dependent.

Example 2.5.1

Let the joint pdf of X_1 and X_2 be

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Are they independent?

Solution:
No, because $f(x_1, x_2) \neq f_1(x_1) f_2(x_2)$:

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 = \int_0^1 (x_1 + x_2) dx_2 = x_1 + 1/2,\ 0 < x_1 < 1,$$
$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 = \int_0^1 (x_1 + x_2) dx_1 = x_2 + 1/2,\ 0 < x_2 < 1,$$

Theorem 2.5.1

Two random variables X_1 and X_2 are independent **if and only if**

- (Continuous case) their joint pdf can be written as a product of a nonnegative function of x_1 and a nonnegative function of x_2:

$$f(x_1, x_2) \equiv g(x_1)h(x_2) \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2$$

- (Discrete case) their joint pmf can be written as a product of a nonnegative function of x_1 and a nonnegative function of x_2:

$$p(x_1, x_2) \equiv g(x_1)h(x_2).$$

Sketch of proof

- **Only if:** Independence $\Rightarrow f(x_1, x_2) \equiv g(x_1)h(x_2)$:
  This can be seen as $g(x_1) = f_1(x_1)$ and $h(x_2) = f_2(x_2)$.
- **If:** Independence $\Leftarrow f(x_1, x_2) \equiv g(x_1)h(x_2)$:
  If we have $f(x_1, x_2) \equiv g(x_1)h(x_2)$, we have

$$f_1(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_2 = g(x_1)\left[\int_{-\infty}^{\infty} h(x_2)dx_2\right] = c_1 g(x_1),$$

$$f_2(x_2) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1 = h(x_2)\left[\int_{-\infty}^{\infty} g(x_1)dx_1\right] = c_2 h(x_2),$$

where c_1 and c_2 are constants. We see $c_1 c_2 = 1$ because

$$1 = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1 dx_2 = \left[\int_{-\infty}^{\infty} g(x_1)dx_1\right]\left[\int_{-\infty}^{\infty} h(x_2)dx_2\right] = c_2 c_1.$$

Thus, $f(x_1, x_2) = g(x_1)h(x_2) = c_1 g(x_1) c_2 h(x_2) = f_1(x_1) f_2(x_2)$.

Independence in terms of CDF

Theorem 2.5.2 Let (X_1, X_2) have the joint cdf $F(x_1, x_2)$ and let X_1 and X_2 have the marginal cdf $F_1(x_1)$ and $F_2(x_2)$, respectively. Then X_1 and X_2 are independent **if and only if**

$$F(x_1, x_2) = F_1(x_1)F_2(x_2), \ \forall (x_1, x_2) \in \mathbb{R}^2.$$

Theorem 2.5.3 The random variables X_1 and X_2 are independent random variables **if and only if** the following condition holds

$$P(a < X_1 \leq b, c < X_2 \leq d) = P(a < X_1 \leq b)P(c < X_1 \leq d),$$

for every $a < b$ and $c < d$, where a, b, c, d are constants.

Example 2.5.3

Let the joint pdf of X_1 and X_2 be

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Are they independent?

Solution:
No, because

$$P(0 < X_1 < \frac{1}{2}, 0 < X_2 < \frac{1}{2}) \neq P(0 < X_1 < \frac{1}{2})P(0 < X_2 < \frac{1}{2}):$$

$$P(0 < X_1 < \frac{1}{2}, 0 < X_2 < \frac{1}{2}) = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} (x_1 + x_2) dx_1 dx_2 = 1/8,$$

$$P(0 < X_1 < \frac{1}{2}) = \int_0^{\frac{1}{2}} (x_1 + \frac{1}{2}) dx_1 = 3/8,$$

$$P(0 < X_2 < \frac{1}{2}) = \int_0^{\frac{1}{2}} (x_2 + \frac{1}{2}) dx_2 = 3/8.$$

Theorem 2.5.4

If X_1 and X_2 are independent and that $\mathsf{E}[u(X_1)]$ and $\mathsf{E}[v(X_2)]$ exist. Then

$$\mathsf{E}[u(X_1)v(X_2)] = \mathsf{E}[u(X_1)]\,\mathsf{E}[v(X_2)].$$

Proof.

$$\begin{aligned}
\mathsf{E}[u(X_1)v(X_2)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x_1)v(x_2)f(x_1,x_2)dx_1dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x_1)v(x_2)f_1(x_1)f_2(x_2)dx_1dx_2 \\
&= \left[\int_{-\infty}^{\infty} u(x_1)f_1(x_1)dx_1\right]\left[\int_{-\infty}^{\infty} f_2(x_2)v(x_2)dx_2\right] \\
&= \mathsf{E}[u(X_1)]\,\mathsf{E}[v(X_2)].
\end{aligned}$$

□

Two special cases

- For independent random variable:

$$\mathsf{E}(X_1 X_2) = \mathsf{E}(X_1)\mathsf{E}(X_2).$$

- Independence implies that covariance $\mathsf{Cov}(X_1, X_2) = 0$:

$$\mathsf{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \mathsf{E}(X_1 - \mu_1)\mathsf{E}(X_2 - \mu_2).$$

Independence always implies zero covariance (correlation).
Zero covariance (correlation) does NOT always imply independence:

Example

Assume that

$$p_{X,Y}(-1,1) = p_{X,Y}(1,1) = 1/4; \quad p_{X,Y}(0,-1) = 1/2.$$

X and Y are not independent because (for example) $p_{Y|X}(-1|0) = 1 \neq p_Y(-1) = 1/2$ but $\text{Cov}(X,Y) = 0$ (check).

Theorem 2.5.5

Suppose that (X_1, X_2) have the joint mgf $M(t_1, t_2)$ and marginal mgf's $M_1(t_1)$ and $M_2(t_2)$, respectively. Then, X_1 and X_2 are independent **if and only if**

$$M(t_1, t_2) \equiv M_1(t_1)M_2(t_2).$$

Example 2.5.5

Let X and Y be two random variables with joint pdf

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

Are they independent?

Solution:
The mgf is

$$M(t_1, t_2) = \frac{1}{(1 - t_1 - t_2)(1 - t_2)}, \ t_1 + t_2 < 1, \ t_2 < 1.$$

Because

$$M(t_1, t_2) \neq M_1(t_1) M_2(t_2) = M(t_1, 0) M(0, t_2),$$

they are dependent.

Chapter 2 Multivariate Distributions

2.6 Extension to Several Random Variables

Examples

1. Random experiment consists of drawing an individual c from a population $\mathcal{C}$.
   Characteristics: height, weight, age, test scores, ...
2. Random experiments consists of the U.S economy at time t.
   Characteristics: consumer prices, unemployment rate, Dow Jones Industrial Average, Gross Domestic Product, ...

A note on notation. We will often use boldface letters to denote vectors. For example, we use $\boldsymbol{X}$ to denote the random vector $(X_1, \ldots, X_n)$, and $\boldsymbol{x}$ to denote the observed values $(x_1, \ldots, x_n)$.

Pmf and cdf for the discrete case

- The joint pmf of a discrete random vector $\boldsymbol{X}$ is defined to be

$$p_{\mathbf{X}}(\mathbf{x}) = P[X_1 = x_1, \ldots, X_n = x_n].$$

- The joint cdf of a discrete random vector $\mathbf{X}$ is defined to be

$$F_{\mathbf{X}}(\mathbf{x}) = P[X_1 \leq x_1, \ldots, X_n \leq x_n].$$

- For the discrete case, $p_{\boldsymbol{X}}(\boldsymbol{x})$ can be used to calculate $P(\boldsymbol{X} \in A)$ for $A \subset \mathbb{R}^n$:

$$P(\boldsymbol{X} \in A) = \sum_{\boldsymbol{x} \in A} p_{\boldsymbol{X}}(\boldsymbol{x}) \,.$$

Pdf and cdf for the continuous case

- The joint cdf of a continuous random vector $\mathbf{X}$ is defined to be

$$F_{\mathbf{X}}(\mathbf{x}) = P[X_1 \leq x_1, \ldots, X_n \leq x_n].$$

- The joint pdf of a continuous random vector $\boldsymbol{X}$ is a function $f_{\boldsymbol{X}}(\boldsymbol{x})$ such that for any $A \subset \mathbb{R}^n$

$$\begin{aligned} P(\boldsymbol{X} \in A) &= \int_A f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int \ldots \int_A f_{X_1, \cdots, X_n}(x_1, \cdots, x_n)\, dx_1 \cdots dx_n. \end{aligned}$$

- For the continuous case, we have

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}).$$

Example

Let

$$f(x_1, x_2, x_3) = \begin{cases} 8x_1x_2x_3 & \text{for } 0 < x_1, x_2, x_3 < 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

Verify that this is a legitimate pdf.

Solution:

$$\int_{x_1=0}^{1} \int_{x_2=0}^{1} \int_{x_3=0}^{1} 8x_1x_2x_3 dx_3 dx_2 dx_1 = 1.$$

Expectation

- For the discrete case, the expectation of $Y = u(X_1, \ldots, X_n)$, if it exists, is defined to be

$$E(Y) = \sum_{x_1, \ldots, x_n} \cdots \sum u(x_1, \ldots, x_n) p_{\mathbf{X}}(x_1, \ldots, x_n).$$

- For the continuous case, the expectation of $Y = u(X_1, \ldots, X_n)$, if it exists, is defined to be

$$E(Y) = \int_{x_1, \ldots, x_n} \cdots \int u(x_1, \ldots, x_n) f_{\mathbf{X}}(x_1, \ldots, x_n) dx_1 \cdots dx_n.$$

As before, E is a linear operator. That is,

$$E\left[\sum_{j=1}^{m} k_j Y_j\right] = \sum_{j=1}^{m} k_j E\left[Y_j\right].$$

Example

Find $E(5X_1X_2^2 + 3X_2X_3^4)$.
Solution:

$$\begin{aligned}
E(X_1X_2^2) &= \int_0^1\int_0^1\int_0^1 (x_1x_2^2)8x_1x_2x_3dx_3dx_2dx_1 = \frac{1}{3}, \\
E(X_2X_3^4) &= \int_0^1\int_0^1\int_0^1 (x_2x_3^4)8x_1x_2x_3dx_3dx_2dx_1 = \frac{2}{9}, \\
E(5X_1X_2^2 + 3X_2X_3^4) &= 5\cdot\frac{4}{15} + 3\cdot\frac{2}{9} = \frac{4}{3} + \frac{2}{3} = 2
\end{aligned}$$

In an obvious way, we may extend the concepts of marginal pmf and marginal pdf for the multidimensional case. For the discrete case, the **marginal pmf** of (X_1, X_2) is defined to be

$$p_{12}(x_1, x_2) = \sum_{x_3} \cdots \sum_{x_n} p_{\mathbf{X}}(x_1, x_2, \ldots, x_n).$$

For the continuous case, the **marginal pdf** of (X_1, X_2) is defined to be

$$f_{12}(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) dx_3 \cdots dx_n.$$

We then extend the concept of conditional pmf and conditional pdf. For the discrete case, suppose $p_1(x_1) > 0$. We define the the **conditional pmf** of $(X_2, \ldots, X_n)$ given $X_1 = x_1$ to be

$$p_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1) = \frac{p(x_1, x_2, \ldots, x_n)}{p_1(x_1)}.$$

For the continuous case, suppose $f_1(x_1) > 0$. We define the **conditional pdf** of $(X_2, \ldots, X_n)$ given $X_1 = x_1$ to be

$$f_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1) = \frac{f(x_1, x_2, \ldots, x_n)}{f_1(x_1)}.$$

For the discrete case, suppose $p_1(x_1) > 0$. Then we define the **conditional expectation** of $u(X_2, \ldots, X_n)$ given $X_1 = x_1$ to be

$$\mathsf{E}\left[u(X_2, \ldots, X_n)|x_1\right] = \sum_{x_2} \cdots \sum_{x_n} u(x_2, \ldots, x_n) p_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1).$$

For the continuous case, suppose $f_1(x_1) > 0$. Then we define the **conditional expectation** of $u(X_2, \ldots, X_n)$ given $X_1 = x_1$ to be

$$\begin{aligned} &\mathsf{E}\left[u(X_2, \ldots, X_n)|x_1\right] \\ = &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_2, \ldots, x_n) f_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1) dx_2 \cdots dx_n. \end{aligned}$$

Mutual Independence

We say that the n random variables $X_1, \ldots, X_n$ are **mutually independent** if, for the discrete case,

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n), \quad \text{for all } (x_1, \cdots, x_n) \in \mathbb{R}^n,$$

or, for the continuous case,

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n) \quad \text{for all } (x_1, \cdots, x_n) \in \mathbb{R}^n.$$

If the n random variables $X_1, \ldots, X_n$ are **mutually independent**, then

$$\begin{aligned} & P\left(a_1 < X_1 < b_1, \ldots, a_n < X_n < b_n\right) \\ = & P\left(a_1 < X_1 < b_1\right) \cdots P\left(a_n < X_n < b_n\right). \end{aligned}$$

We may rewrite the above equation as

$$P\left(\bigcap_{j=1}^{n} \left(a_j < X_j < b_j\right)\right) = \prod_{j=1}^{n} P\left(a_j < X_j < b_j\right).$$

If the n random variables X_1, X_2, ..., X_n are **mutually independent**, then

$$\mathsf{E}\left[u_1(X_1)u_2(X_2)\cdots u_n(X_n)\right] = \mathsf{E}\left[u_1(X_1)\right] E\left[u_2(X_2)\right]\cdots \mathsf{E}\left[u_n(X_n)\right],$$

$$\mathsf{E}\left[\prod_{j=1}^{n} u_j(X_j)\right] = \prod_{j=1}^{n} \mathsf{E}\left[u_j(X_j)\right].$$

As a special case of the above, if the n random variables X_1, X_2, ..., X_n are mutually independent, then for mgf,

$$M(t_1, t_2, \cdots, t_n) = \prod_{j=1}^{n} M_j(t_j),$$

which can be seen from

$$\begin{aligned} M(t_1, t_2, \cdots, t_n) &= \mathsf{E}[\exp(t_1X_1 + t_2X_2 + \ldots + t_nX_n)] \\ &= \mathsf{E}\left[\prod_{j=1}^{n} \exp(t_jX_j)\right] \\ &= \prod_{j=1}^{n} \mathsf{E}\left[\exp(t_jX_j)\right] = \prod_{j=1}^{n} M_j(t_j). \end{aligned}$$

Mutual independence v.s. pairwise independence

- We say the n random variables X_1, X_2, $\ldots$, X_n are **pairwise independent** if for all pairs (i, j) with $i \neq j$, the random variables X_i and X_j are independent.
- Unless there is a possible misunderstanding between mutual independence and pairwise independence, we usually drop the modifier **mutual**.
- If the n random variables X_1, X_2, $\ldots$, X_n are independent and have the same distribution, then we say that they are **independent and identically distributed**, which we abbreviate as **i.i.d.**.

Compare "**mutual** independence" and "**pairwise** independence".

Example (from S. Bernstein)

Consider a random vector (X_1, X_2, X_3) that has joint pmf $p(x_1, x_2, x_3)$

$$= \begin{cases} \frac{1}{4} & \text{for } (x_1, x_2, x_3) \in \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}. \\ 0 & \text{otherwise.} \end{cases}$$

Solution:

$$p_{ij}(x_i, x_j) = \begin{cases} \frac{1}{4} & \text{for } (x_i, x_j) \in \{(0,0), (1,0), (0,1), (1,1)\}. \\ 0 & \text{otherwise.} \end{cases}$$

$$p_i(x_i) = \begin{cases} \frac{1}{2} & \text{for } (x_i) \in \{0,1\}. \\ 0 & \text{otherwise.} \end{cases}$$

pairwise independence : $\quad p_{ij}(x_i, x_j) = p_i(x_i)p_j(x_j).$

not mutual independence : $\quad p(x_1, x_2, x_3) \neq p_1(x_1)p_2(x_2)p_3(x_3).$

Multivariate Variance-Covariance Matrix

1. Let $\boldsymbol{X} = (X_1, \cdots, X_n)^\top$ be a random vector.
2. We define the expectation of $\boldsymbol{X}$ as $\mathsf{E}\boldsymbol{X} = (\mathsf{E}X_1, \cdots, \mathsf{E}X_n)^\top$.
3. Let $\mathbf{W} = [W_{ij}]$ be a $m \times n$ matrix, where W_{ij} are random variables. That is,

$$\mathbf{W} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & & W_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ W_{m1} & W_{m2} & \cdots & W_{mn} \end{bmatrix} = [W_{ij}]_{m \times n}.$$

4. We define the **expectation of this random matrix** as $\mathsf{E}[\mathbf{W}] = [\mathsf{E}(W_{ij})]$. That is,

$$\mathsf{E}[\mathbf{W}] = \begin{bmatrix} \mathsf{E}(W_{11}) & \mathsf{E}(W_{12}) & \cdots & \mathsf{E}(W_{1n}) \\ \mathsf{E}(W_{21}) & \mathsf{E}(W_{22}) & & \mathsf{E}(W_{2n}) \\ \cdots & \cdots & \cdots & \cdots \\ \mathsf{E}(W_{m1}) & \mathsf{E}(W_{m2}) & \cdots & \mathsf{E}(W_{mn}) \end{bmatrix} = [\mathsf{E}(W_{ij})]_{m \times n}.$$

Theorem 2.6.2

Let $\mathbf{W}$ and $\mathbf{V}$ be $m \times n$ random matrices, and let $\mathbf{A}$ and $\mathbf{B}$ be $k \times m$ constant matrices, and let $\mathbf{C}$ be a $n \times l$ constant matrix. Then,

$$\mathsf{E}\left[\mathbf{AW} + \mathbf{BV}\right] = \mathbf{A}\mathsf{E}\left[\mathbf{W}\right] + \mathbf{B}\mathsf{E}\left[\mathbf{V}\right]$$

and

$$\mathsf{E}\left[\mathbf{AWC}\right] = \mathbf{A}\mathsf{E}\left[\mathbf{W}\right]\mathbf{C}.$$

Proof sketch:

The (i, j) of the first equation:

$$\mathsf{E}\left[\sum_{s=1}^{m} A_{is}W_{sj} + \sum_{s=1}^{m} B_{is}V_{sj}\right] = \sum_{s=1}^{m} A_{is}\mathsf{E}[W_{sj}] + \sum_{s=1}^{m} B_{is}\mathsf{E}[V_{sj}].$$

Let $\mathbf{X} = (X_1, \dots X_n)^\top$ be an n-dimensional random vector with mean vector $\boldsymbol{\mu}$. Then the **variance-covariance matrix** of $\boldsymbol{X}$ is defined to be

$$
\begin{aligned}
& \text{Cov}(\mathbf{X}) \\
= \quad & \text{E}\left[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^\top\right] \\
= \quad & \text{E}\begin{bmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_n - \mu_n) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)(X_2 - \mu_2) & & (X_2 - \mu_2)(X_n - \mu_n) \\ \dots & \dots & \dots & \dots \\ (X_n - \mu_n)(X_1 - \mu_1) & (X_n - \mu_n)(X_2 - \mu_2) & \cdots & (X_n - \mu_n)(X_n - \mu_n) \end{bmatrix} \\
= \quad & \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}
\end{aligned}
$$

Example of a covariance matrix

Let X and Y be two random variables with joint pdf

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

We have $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_{1,2} = \text{Cov}(X,Y) = 1$.
Let $\boldsymbol{Z} = (X,Y)^\top$, then

$$\text{E}(\boldsymbol{Z}) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \text{Cov}(\boldsymbol{Z}) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Theorem 2.6.3 – Two properties of covariance matrix

Let $\boldsymbol{X} = (X_1, \dots X_n)^\top$ be an n-dimensional random vector with mean vector $\boldsymbol{\mu}$. Then,

$$\text{Cov}\left(\boldsymbol{X}\right) = \text{E}\left[\boldsymbol{X}\boldsymbol{X}^\top\right] - \boldsymbol{\mu}\boldsymbol{\mu}^\top. \quad (3)$$

If further let $\mathbf{A}$ be an $m \times n$ constant matrix, then we have

$$\text{Cov}\left(\boldsymbol{A}\boldsymbol{X}\right) = \boldsymbol{A}\text{Cov}\left(\boldsymbol{X}\right)\boldsymbol{A}^\top.$$

Proof.

$$\begin{aligned}
\text{Cov}(\boldsymbol{X}) &= \text{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top] \\
&= \text{E}[(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{\mu}\boldsymbol{X}^\top - \boldsymbol{X}\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu})^\top] \\
&= \text{E}[\boldsymbol{X}\boldsymbol{X}^\top] - \boldsymbol{\mu}\text{E}[\boldsymbol{X}^\top] - \text{E}[\boldsymbol{X}]\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top. \\
\text{Cov}(\boldsymbol{A}\boldsymbol{X}) &= \text{E}\left[(\boldsymbol{A}\boldsymbol{X})(\boldsymbol{A}\boldsymbol{X})^\top\right] - (\boldsymbol{A}\boldsymbol{\mu})(\boldsymbol{A}\boldsymbol{\mu})^\top \\
&= \text{E}\left[\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{A}^\top\right] - \boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^\top\boldsymbol{A}^\top \\
&= \boldsymbol{A}\text{E}\left[\boldsymbol{X}\boldsymbol{X}^\top\right]\boldsymbol{A}^\top - \boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^\top\boldsymbol{A}^\top
\end{aligned}$$

Proof without matrix notation

$$
\begin{aligned}
& \mathsf{Cov}(\mathbf{X}) \\
= \quad & \mathsf{E}\left[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^\top\right] \\
= \quad & \mathsf{E}\begin{bmatrix}
(X_1-\mu_1)(X_1-\mu_1) & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_n-\mu_n) \\
(X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)(X_2-\mu_2) & & (X_2-\mu_2)(X_n-\mu_n) \\
\cdots & \cdots & \cdots & \cdots \\
(X_n-\mu_n)(X_1-\mu_1) & (X_n-\mu_n)(X_2-\mu_2) & \cdots & (X_n-\mu_n)(X_n-\mu_n)
\end{bmatrix} \\
= \quad & \begin{bmatrix}
\mathsf{E}(X_1X_1)-\mu_1\mu_1 & \mathsf{E}(X_1X_2)-\mu_1\mu_2 & \cdots & \mathsf{E}(X_1X_n)-\mu_1\mu_n \\
\mathsf{E}(X_2X_1)-\mu_2\mu_1 & \mathsf{E}(X_2X_2)-\mu_2\mu_2 & & \mathsf{E}(X_2X_n)-\mu_2\mu_n \\
\cdots & \cdots & \cdots & \cdots \\
\mathsf{E}(X_nX_1)-\mu_n\mu_1 & \mathsf{E}(X_nX_2)-\mu_n\mu_2 & \cdots & \mathsf{E}(X_nX_n)-\mu_n\mu_n
\end{bmatrix} \\
= \quad & \mathsf{E}\begin{bmatrix}
(X_1X_1) & (X_1X_2) & \cdots & (X_1X_n) \\
(X_2X_1) & (X_2X_2) & & (X_2X_n) \\
\cdots & \cdots & \cdots & \cdots \\
(X_nX_1) & (X_nX_2) & \cdots & X_nX_n
\end{bmatrix} - \begin{bmatrix}
\mu_1\mu_1 & \mu_1\mu_2 & \cdots & \mu_1\mu_n \\
\mu_2\mu_1 & \mu_2\mu_2 & & \mu_2\mu_n \\
\cdots & \cdots & \cdots & \cdots \\
\mu_n\mu_1 & \mu_n\mu_2 & \cdots & \mu_n\mu_n
\end{bmatrix} \\
= \quad & \mathsf{E}\left[\mathbf{X}\mathbf{X}^\top\right] - \mu\mu^\top.
\end{aligned}
$$

- All variance-covariance matrices are **positive semi-definite**, that is $\boldsymbol{a}^\top \text{Cov}(\boldsymbol{X})\boldsymbol{a} \geq 0$ for any $\boldsymbol{a} \in \mathbb{R}^n$.
- This is because

$$\boldsymbol{a}^\top \text{Cov}(\boldsymbol{X})\boldsymbol{a} = \text{Var}(\boldsymbol{a}^\top \boldsymbol{X}) \geq 0,$$

where we note that $\boldsymbol{a}^\top \boldsymbol{X}$ is a univariate random variable.

Chapter 2 Multivariate Distributions

2.7 Transformation for Several Random Variables

One to one transformation

- Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector with pdf $f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n)$ with support $\mathcal{S}$. Let

$$\begin{cases} y_1 = g_1(x_1, x_2, \ldots, x_n) \\ y_2 = g_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ y_n = g_n(x_1, x_2, \ldots, x_n) \end{cases}$$

  be a multivariate function that maps $(x_1, x_2, \ldots, x_n) \in \mathcal{S}$ to $(y_1, y_2, \ldots, y_n) \in \mathcal{T}$. Suppose that it is a one-to-one correspondence.
- Suppose that the inverse functions are given by

$$\begin{cases} x_1 = h_1(y_1, y_2, \ldots, y_n) \\ x_2 = h_2(y_1, y_2, \ldots, y_n) \\ \vdots \\ x_n = h_n(y_1, y_2, \ldots, y_n) \end{cases}.$$

- Let the Jacobian be

$$J = \left| \frac{\partial (x_1, x_2, \ldots, x_n)}{\partial (y_1, y_2, \ldots, y_n)} \right| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}.$$

- Then, the joint pdf of Y_1, Y_2, ..., Y_n determined by the mapping above is

$$\begin{aligned} &f_{\boldsymbol{Y}}(y_1, y_2, \ldots, y_n) \\ &= |J| \, f_{\boldsymbol{X}} \left[h_1(y_1, y_2, \ldots, y_n), h_2(y_1, y_2, \ldots, y_n), \ldots, h_n(y_1, y_2, \ldots, y_n) \right], \end{aligned}$$

for $(y_1, y_2, \ldots, y_n) \in \mathcal{T}$.

Example 2.7.1

Suppose X_1, X_2, and X_3 have joint pdf

$$f(x_1, x_2, x_3) = \begin{cases} 48x_1x_2x_3 & 0 < x_1 < x_2 < x_3 < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

and let

$$\begin{cases} Y_1 = X_1/X_2 \\ Y_2 = X_2/X_3 \\ Y_3 = X_3. \end{cases}$$

Determine the joint pdf of Y_1, Y_2 and Y_3.

If $Y_1 = X_1/X_2$, $Y_2 = X_2/X_3$, and $Y_3 = X_3$, then the inverse transformation is given by

$$x_1 = y_1 y_2 y_3, \; x_2 = y_2 y_3, \; \text{and} \; x_3 = y_3.$$

The Jacobian is given by

$$J = \begin{vmatrix} y_2 y_3 & y_1 y_3 & y_1 y_2 \\ 0 & y_3 & y_2 \\ 0 & 0 & 1 \end{vmatrix} = y_2 y_3^2.$$

Moreover, inequalities defining the support are equivalent to

$$0 < y_1 y_2 y_3, \; y_1 y_2 y_3 < y_2 y_3, \; y_2 y_3 < y_3, \; \text{and} \; y_3 < 1,$$

which reduces to the support $\mathcal{T}$ of Y_1, Y_2, Y_3 of

$$\mathcal{T} = \{(y_1, y_2, y_3) : \; 0 < y_i < 1, i = 1, 2, 3\}.$$

Hence the joint pdf of Y_1, Y_2, Y_3 is

$$\begin{aligned} g(y_1, y_2, y_3) &= 48(y_1 y_2 y_3)(y_2 y_3) y_3 |y_2 y_3^2| \\ &= \begin{cases} 48 y_1 y_2^3 y_3^5 & 0 < y_i < 1, i = 1, 2, 3 \\ 0 & \text{elsewhere.} \end{cases} \end{aligned} \tag{2.7.2}$$

The marginal pdfs are

$$\begin{aligned} g_1(y_1) &= 2y_1, 0 < y_1 < 1, \text{zero elsewhere} \\ g_2(y_2) &= 4y_2^3, 0 < y_2 < 1, \text{zero elsewhere} \\ g_3(y_3) &= 6y_3^5, 0 < y_3 < 1, \text{zero elsewhere.} \end{aligned}$$

Because $g(y_1, y_2, y_3) = g_1(y_1) g_2(y_2) g_3(y_3)$, the random variables Y_1, Y_2, Y_3 are mutually independent. ■

Multiple to one transformation

- Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector with pdf $f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n)$ with support $\mathcal{S}$. Let

$$\begin{cases} y_1 = g_1(x_1, x_2, \ldots, x_n) \\ y_2 = g_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ y_n = g_n(x_1, x_2, \ldots, x_n) \end{cases}$$

  be a multivariate function that maps $\boldsymbol{X} = (x_1, x_2, \ldots, x_n) \in \mathcal{S}$ to $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n) \in \mathcal{T}$.
- Suppose that the support $\mathcal{S}$ can be represented as the union of k mutually disjoint sets such that for each i, there is one-to-one correspondence bewteen $\boldsymbol{X}$ and $\boldsymbol{Y}$.
- Suppose that the inverse functions are given by

$$\begin{cases} x_1 = h_{1i}(y_1, y_2, \ldots, y_n) \\ x_2 = h_{2i}(y_1, y_2, \ldots, y_n) \\ \vdots \\ x_n = h_{ni}(y_1, y_2, \ldots, y_n) \end{cases}.$$

Let the Jacobian be

$$J_i = \left| \frac{\partial (x_1, x_2, \ldots, x_n)}{\partial (y_1, y_2, \ldots, y_n)} \right| = \begin{vmatrix} \frac{\partial h_{1i}}{\partial y_1} & \frac{\partial h_{1i}}{\partial y_2} & \cdots & \frac{\partial h_{1i}}{\partial y_n} \\ \frac{\partial h_{2i}}{\partial y_1} & \frac{\partial h_{2i}}{\partial y_2} & \cdots & \frac{\partial h_{2i}}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial h_{ni}}{\partial y_1} & \frac{\partial h_{ni}}{\partial y_2} & \cdots & \frac{\partial h_{ni}}{\partial y_n} \end{vmatrix}.$$

Then, the joint pdf of Y_1, Y_2, ..., Y_n determined by the mapping above is

$$f_{\boldsymbol{Y}}(y_1, y_2, \ldots, y_n) = \sum_{i=1}^{k} |J_i| \, f_{\boldsymbol{X}} \left[h_{1i}(y_1, y_2, \ldots, y_n), h_{2i}(y_1, y_2, \ldots, y_n), \ldots, h_{ni}(y_1, y_2, \ldots, y_n) \right],$$

for $(y_1, y_2, \ldots, y_n) \in \mathcal{T}$.

Example 2.7.3

Let X_1 and X_2 have the joint pdf defined over the unit circle given by

$$f(x_1, x_2) = \begin{cases} \frac{1}{\pi} & 0 < x_1^2 + x_2^2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Let

$$\begin{cases} Y_1 = X_1^2 + X_2^2 \\ Y_2 = X_1^2 / \left(X_1^2 + X_2^2\right). \end{cases}$$

Determine the joint pdf of Y_1 and Y_2.

Let $Y_1 = X_1^2 + X_2^2$ and $Y_2 = X_1^2/(X_1^2 + X_2^2)$. Thus $y_1 y_2 = x_1^2$ and $x_2^2 = y_1(1 - y_2)$. The support $\mathcal{S}$ maps onto $\mathcal{T} = \{(y_1, y_2) : 0 < y_i < 1, i = 1, 2\}$. For each ordered pair $(y_1, y_2) \in \mathcal{T}$, there are four points in $\mathcal{S}$, given by

$$\begin{aligned}
&(x_1, x_2) \quad \text{such that} \quad x_1 = \sqrt{y_1 y_2} \text{ and } x_2 = \sqrt{y_1(1 - y_2)} \\
&(x_1, x_2) \quad \text{such that} \quad x_1 = \sqrt{y_1 y_2} \text{ and } x_2 = -\sqrt{y_1(1 - y_2)} \\
&(x_1, x_2) \quad \text{such that} \quad x_1 = -\sqrt{y_1 y_2} \text{ and } x_2 = \sqrt{y_1(1 - y_2)} \\
\text{and } &(x_1, x_2) \quad \text{such that} \quad x_1 = -\sqrt{y_1 y_2} \text{ and } x_2 = -\sqrt{y_1(1 - y_2)}.
\end{aligned}$$

The value of the first Jacobian is

$$\begin{aligned}
J_1 &= \begin{vmatrix} \frac{1}{2}\sqrt{y_2/y_1} & \frac{1}{2}\sqrt{y_1/y_2} \\ \frac{1}{2}\sqrt{(1 - y_2)/y_1} & -\frac{1}{2}\sqrt{y_1/(1 - y_2)} \end{vmatrix} \\
&= \frac{1}{4}\left\{-\sqrt{\frac{1 - y_2}{y_2}} - \sqrt{\frac{y_2}{1 - y_2}}\right\} = -\frac{1}{4}\frac{1}{\sqrt{y_2(1 - y_2)}}.
\end{aligned}$$

It is easy to see that the absolute value of each of the four Jacobians equals $1/4\sqrt{y_2(1 - y_2)}$. Hence, the joint pdf of Y_1 and Y_2 is the sum of four terms and can be written as

$$g(y_1, y_2) = 4\frac{1}{\pi}\frac{1}{4\sqrt{y_2(1 - y_2)}} = \frac{1}{\pi\sqrt{y_2(1 - y_2)}}, \quad (y_1, y_2) \in \mathcal{T}.$$

Thus Y_1 and Y_2 are independent random variables by Theorem 2.5.1. ■

Chapter 2 Multivariate Distributions

2.8 Linear Combinations of Random Variables

Motivation

- We are interested in a function of $T = T(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ is a random vector.
- For example, we let each X_i denote the final percentage of STAT 4100 grade. Assume we know the distribution of each X_i, can we know the distribution of the average percentage $\bar{X}$?
- In this section, we focus on linear combination of these variables, i.e.,

$$T = \sum_{i=1}^{n} a_i X_i.$$

Expectation of linear combinations

Theorem 2.8.1. Let $T = \sum_{i=1}^{n} a_i X_i$. Provided that $\mathsf{E}[|X_i|] < \infty$, for all $i = 1, \ldots, n$, then

$$\mathsf{E}(T) = \sum_{i=1}^{n} a_i \mathsf{E}(X_i).$$

This theorem follows immediately from the linearity of the expectation operation.

Theorem 2.8.2. Let $T = \sum_{i=1}^{n} a_i X_i$ and $W = \sum_{j=1}^{m} b_j Y_j$. If $\mathsf{E}[X_i^2] < \infty$ and $\mathsf{E}[Y_j^2] < \infty$, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, then

$$\mathsf{Cov}(T, W) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathsf{Cov}(X_i, Y_j).$$

Proof:

$$\begin{aligned}\mathsf{Cov}(T, W) &= \mathsf{E}\left[\sum_{i=1}^{n} \sum_{j=1}^{m} (a_i X_i - a_i \mathsf{E}(X_i))(b_j Y_j - b_j \mathsf{E}(Y_j))\right] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \mathsf{E}[(a_i X_i - a_i \mathsf{E}(X_i))(b_j Y_j - b_j \mathsf{E}(Y_j))].\end{aligned}$$

Corollary 2.8.1. Let $T = \sum_{i=1}^{n} a_i X_i$. Provided $\mathsf{E}[X_i^2] < \infty$, for $i = 1, \ldots, n$, then

$$\mathsf{Var}(T) = \mathsf{Cov}(T, T) = \sum_{i=1}^{n} a_i^2 \mathsf{Var}(X_i) + 2 \sum_{i<j}^{m} a_i a_j \mathsf{Cov}(X_i, Y_j).$$

Corollary 2.8.2. If $X_1, \ldots, X_n$ are independent random variables with finite variances, then

$$\mathsf{Var}(T) = \sum_{i=1}^{n} a_i^2 \mathsf{Var}(X_i).$$

Special case If X_1 and X_2 have finite variances, then

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\mathsf{Cov}(X, Y).$$

If they are also independent, then

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y).$$

Note that $\mathsf{E}(X + Y) = \mathsf{E}(X) + \mathsf{E}(Y)$ regardless of independence.

Example 2.8.1 – Sample mean

Let $X_1, \dots, X_n$ be independent and identically distributed random variables with common mean μ and variance σ^2. The sample mean is defined by $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$. This is a linear combination of the sample observations with $a_i \equiv n^{-1}$; hence by Theorem 2.8.1 and Corollary 2.8.2, we have

$$\mathrm{E}(\bar{X}) = \mu \text{ and } \mathrm{Var}(\bar{X}) = \sigma^2/n.$$

Example 2.8.2 – Sample variance

Define the **sample variance** by

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

Following from the fact that $\mathsf{E}(X^2) = \sigma^2 + \mu^2$,

$$\begin{aligned}\mathsf{E}(S^2) &= (n-1)^{-1}\left(\sum_{i=1}^{n}\mathsf{E}(X_i^2) - n\mathsf{E}(\bar{X}^2)\right)\\ &= (n-1)^{-1}\{n\sigma^2 + n\mu^2 - n[(\sigma^2/n + \mu^2)]\}\\ &= \sigma^2.\end{aligned}$$