Stat 121 (Mathematical Statistics I)

Lesson 4.1 Joint Distribution for Two Random Variables

Learning Outcomes

At the end of the lesson, students must be able to

1. Explain joint probability distribution of random variables, and
2. Derive the joint distribution of two discrete and continuous random variables.

Introduction

In the Modules 2 and 3, we were interested in **univariate** random variables (of the discrete and continuous type, respectively). However, in many problems, there are two or more random variables of interest and the goal is to understand the probabilistic behavior of them together. For example,

- Researchers would like to use a student's pretest score (Y_1) and his/her posttest score (Y_2) to assess the effectiveness of an educational program
- In clinical trials, physicians want to characterize the concentration of a drug in a patient's body (Y) as a function of the patient's body weight (X)
- In a variety trial, we might be interest how new varieties respond to a fertilizer treatments in terms of agronomic characteristics such plant height (Y_1), leaf area index (Y_2), yield (Y_3), etc

In each example, it is natural to posit a relationship between or among the random variables involved. This relationship can be described mathematically using a **joint probability distribution**. This distribution, in turn, allows us to make probability statements involving the random variables- just as univariate distributions allow us to do with a single variable.

Joint distribution of two discrete random variables

Definition

Let Y_1 and Y_2 be discrete random variables. We call $\mathbf{Y} = (Y_1, Y_2)$ a **discrete random vector**. The **joint probability mass function** of Y_1 and Y_2 is

$$p_{(Y_1,Y_2)}(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$$

which is nonzero for all $(y_1, y_2) \in R$. The set $R \subset \mathbb{R}^2$ is the two-dimensional support of $\mathbf{Y} = (Y_1, Y_2)$. The joint PMF $p_{(Y_1,Y_2)}(y_1, y_2)$ has the following properties:

1. $0 \leq p_{(Y_1,Y_2)}(y_1, y_2) \leq 1, \forall y_1, y_2$, and
2. $\sum\sum p_{(Y_1,Y_2)}(y_1, y_2) = 1, \forall (y_1, y_2) \in R$.

Note that these are direct equivalents of the properties of the PMF of the univariate random variables.

Example 4.1.1

Let Y_1 be the number of earthquakes and Y_2 be the number of typhoons recorded in Eastern Visayas in July 2025. Suppose the joint PMF of Y_1 and Y_2 is given by

$p(y_1, y_2)$	$y_2 = 0$	$y_2 = 1$	$y_2 = 2$
$y_1 = 0$	0.64	0.08	0.04
$y_1 = 1$	0.12	0.06	0.02
$y_1 = 2$	0.02	0.01	0.01

a. What is the probability that there is no earthquake and no typhoon recorded in Eastern Visayas in July 2025?

b. What is the probability that there are 2 earthquakes and less than 2 typhoons in Eastern Visayas in July 2025?

c. What is the probability that there are at most 1 earthquake recorded in Eastern Visayas in July 2025?

SOLUTION

a. $P(Y_1 = 0, Y_2 = 0) = 0.64$

b.
$$\begin{aligned} P(Y_1 = 2, Y_2 < 2) &= P(Y_1 = 2, Y_2 = 1) + P(Y_1 = 2, Y_2 = 0) \\ &= 0.01 + 0.02 \\ &= 0.03 \end{aligned}$$

c.

$$
\begin{aligned}
P(Y_1 \le 1) &= P(Y_1 = 1, Y_2 = 0) + P(Y_1 = 1, Y_2 = 1) + P(Y_1 = 1, Y_2 = 2) \\
&\quad + P(Y_1 = 0, Y_2 = 0) + P(Y_1 = 0, Y_2 = 1) + P(Y_1 = 0, Y_2 = 2) \\
&= 0.12 + 0.06 + 0.02 + 0.64 + 0.08 + 0.04 \\
&= 0.96
\end{aligned}
$$

Example 4.1.2

A local supermarket has three checkout counters. Two customers arrive at the counters at different times when the counters are serving no other customers. Each customer chooses a counter at random, independently of the other. Let Y_1 denote the number of customers who choose counter 1, and Y_2 the number who select counter 2. Find the joint probability function of Y_1 and Y_2.

SOLUTION

Let the pair (i, j) denote the simple event that the first customer chose counter i and the second customer chose counter j, where $i, j = 1, 2, 3$.

The sample space associated with the experiment is

$$S = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}$$

The element $(1, 1)$ in the sample space means that both customers go to Counter 1, and $(2, 3)$ means that the first customer is in Counter 2 and the second is in Counter 3.

Now, we have $Y_1 = 0, 1, 2$ and $Y_2 = 0, 1, 2$. Thus,

$P(Y_1 = 0, Y_2 = 0) = \frac{1}{9}$, this refers to the pair (3,3) which means both customers are in Counter 3

$P(Y_1 = 0, Y_2 = 1) = \frac{2}{9}$, for pairs (3,2) and (2,3)

$P(Y_1 = 0, Y_2 = 2) = \frac{1}{9}$, for pair (2,2)

$P(Y_1 = 1, Y_2 = 0) = \frac{2}{9}$, for (1,3) and (3,1)

$P(Y_1 = 1, Y_2 = 1) = \frac{2}{9}$, for pairs (1,2) and (2,1)

$P(Y_1 = 1, Y_2 = 2) = 0$, impossible to happen since there are 2 customers only

$P(Y_1 = 2, Y_2 = 0) = \frac{1}{9}$, for pair (1,1)

$P(Y_1 = 2, Y_2 = 1) = 0$, impossible to happen since there are 2 customers only

$P(Y_1 = 2, Y_2 = 2) = 0$, impossible to happen since there are 2 customers only

Therefore, the joint PMF can be written as

	$Y_2 = 0$	$Y_2 = 1$	$Y_2 = 2$
$Y_1 = 0$	1/9	2/9	1/9
$Y_1 = 1$	2/9	2/9	0
$Y_1 = 2$	1/9	0	0

Joint distribution of two continuous random variables

Definition

Suppose Y_1 and Y_2 are continuous random variables. We call $\mathbf{Y} = (Y_1, Y_2)$ a **continuous random vector**. The **joint probability density function** of Y_1 and Y_2 is given by

$$f_{(Y_1,Y_2)}(y_1, y_2).$$

The joint PDF $f_{(Y_1,Y_2)}(y_1, y_2)$ is a three-dimensional function which is strictly larger than zero over $R \subset \mathbb{R}^2$, the two-dimensional support of $\mathbf{Y} = (Y_1, Y_2)$. The joint PDF $f_{(Y_1,Y_2)}(y_1, y_2)$ has the following properties:

1. $f_{(Y_1,Y_2)}(y_1, y_2) \geq 0, \forall (y_1, y_2) \in \mathbb{R}^2$
2. The function $f_{(Y_1,Y_2)}(y_1, y_2)$ integrates to 1, that is,

$$\int\int f_{(Y_1,Y_2)}(y_1, y_2) dy_1 dy_2 = 1, \forall (y_1, y_2) \in \mathbb{R}^2$$

Example 4.1.3

Suppose $\mathbf{Y} = (Y_1, Y_2)$ is a continuous random vector wirh joint PDF

$$f_{(Y_1,Y_2)}(y_1, y_2) = \begin{cases} cy_1 y_2, \ 0 < y_2 < y_1 < 1 \\ 0, \ \text{otherwise} \end{cases}$$

Find the value of c that makes $f_{(Y_1,Y_2)}(y_1, y_2)$ a valid PDF.

SOLUTION

To find the value of c we set the following equation based on the second property of the joint PDF

$$\int_{y_2=0}^{1} \int_{y_1=y_2}^{1} cy_1 y_2 \, dy_1 dy_2 = 1$$

We have

$$
\begin{aligned}
\int_{y_2=0}^{1}\int_{y_1=y_2}^{1} cy_1y_2\,dy_1dy_2 &= \int_{y_2=0}^{1}\left[cy_2\left(\frac{y_1^2}{2}\right)\Bigg|_{y_1=y_2}^{1}\right]dy_2 \\
&= \int_{y_2=0}^{1}\left[\frac{c}{2}y_2\left(1^2-y_2^2\right)\right]dy_2 \\
&= \frac{c}{2}\int_{y_2=0}^{1}(y_2-y_2^3)dy_2 \\
&= \frac{c}{2}\left[\frac{y_2^2}{2}-\frac{y_2^4}{4}\right]\Bigg|_0^1 \\
&= \frac{c}{2}\left[\frac{1^2}{2}-\frac{1^4}{4}\right]-0 \\
&= \frac{c}{2}\left[\frac{1}{2}-\frac{1}{4}\right] \\
&= \frac{c}{2}\left(\frac{1}{4}\right) \\
&= \frac{c}{8}
\end{aligned}
$$

Equating to 1, we get $c = 8$. Thus,

$$
f_{(Y_1,Y_2)}(y_1,y_2) = \begin{cases} 8y_1y_2, & 0 < y_2 < y_1 < 1 \\ 0, & \text{otherwise} \end{cases}
$$

<u>Example 4.1.4</u>

Let Y_1 and Y_2 have the joint probability density function given by

$$
f_{(Y_1,Y_2)}(y_1,y_2) = \begin{cases} 4y_1y_2, & 0 < y_1 < 1; 0 < y_2 < 1 \\ 0, & \text{otherwise} \end{cases}
$$

Calculate $P(Y_1 \leq 0.5, Y_2 < 0.75)$.

SOLUTION: Left as a classroom exercise.

Learning Task

Instruction: Answer the following as indicated.

1. Of nine executives in a business firm, four are married, three have never married, and two are divorced. Three of the executives are to be selected for promotion. Let Y_1 denote the number of married executives and Y_2 denote the number of never-married executives among the three selected for promotion. Assuming that the three are randomly selected from the nine available, find the joint probability function of Y_1 and Y_2.

2. A bank operates a drive-up and a walk-up window. Let Y_1 be the proportion of time the drive-up facility is in use, and Y_2 the proportion of time the walk-up window in use. Say the manager has given us the joint PDF based on his experience:

$$f_{(Y_1,Y_2)}(y_1, y_2) = \begin{cases} \frac{6}{5}(y_1 + y_2^2), \ 0 \le y_1 \le 1; 0 \le y_2 \le 1 \\ 0, \ \text{elsewhere} \end{cases}$$

a. Verify that the above function is a valid PDF.

b. Calculate $P(Y_1 \le 0.25, Y_2 \le 0.25)$.