Statistics and Probabilities for Computing (CEN1602)

Unit-5

Joint Probability Distributions

Prepared by

Dr. Ahmad Alzahrani

Outline

1. From one random variable to two random variables
   - Why we study two variables together
2. Joint probability distributions
   - Joint PMF for discrete random variables
   - Joint PDF for continuous random variables
3. Marginal distributions
   - Reducing a joint distribution to one variable
   - Marginal PMFs and marginal PDFs
4. Conditional probability from a joint distribution
   - Probability of one variable value when another variable value is known
5. Independence of two random variables
   - When knowing X does not change what we know about Y
6. Covariance and correlation
   - Measuring how two variables move together
7. Multinomial distribution
   - Extension of binomial distribution to more than two categories
8. Bivariate normal distribution
   - A joint PDF for two normally distributed continuous variables

Learning Objectives for Unit 5

By the end of this unit, students should be able to:

1. Explain why we sometimes study two random variables together instead of one variable only.
2. Define a joint probability distribution for two random variables.
3. Use a joint PMF to describe two discrete random variables.
4. Use a joint PDF to describe two continuous random variables.
5. Find marginal distributions from a joint distribution.
6. Explain conditional probability from a joint distribution.
7. Determine whether two random variables are independent.
8. Interpret covariance and correlation as measures of relationship between two variables.
9. Recognize and apply the multinomial distribution for outcomes with more than two categories.
10. Recognize the bivariate normal distribution as a model for two normally distributed continuous variables.

Main Idea

- **In Unit 4**, we asked questions about **one variable**. Example:

$$X = \textit{battery usage (\%) in one hour}$$

$$P(20 < X < 40)$$

What is the probability that battery usage is between 20% and 40%?

- **In Unit 5**, we ask questions about **two variables** **at the same time**. Example:

$$X = \textit{number of opened apps}$$

$$Y = \textit{battery usage (\%) in one hour}$$

$$P(X > 5 \textit{ and } Y > 40)$$

What is the probability that more than 5 apps are open and battery usage is more than 40%?

- Sometimes two variables are related.
- For example, opening more apps may be associated with higher battery usage.
- **Joint probability** helps us describe and analyze this relationship.

Main Idea

- Other questions can be answered by **joint probability distribution**:
  - What is the probability that a user has 2 failed attempts and logs in within 10 seconds?
  - What is the probability that a server has high CPU usage and high memory usage at the same time?
  - Are download size and download time related?
  - Can knowing one variable help us understand another?

- **Examples of 2 related variables on computing:**
  - Number of active users **vs.** number of server requests per minute
  - Number of opened apps **vs.** battery consumption
  - Number of browser tabs open **vs.** memory consumption
  - User session duration **vs.** number of clicks
  - Number of detected errors **vs.** system recovery time
  - Number of running processes **vs.** processor temperature
  - Message length **vs.** transmission time

From Unit 4 to Unit 5

Unit 4 One Continuous Random Variable	Unit 5 Two Random Variables Together
Study one variable (X)	Study two variables (X) and (Y)
Example: battery usage in one hour	Example: opened apps and battery usage
Uses PMF $p(x)$ for discrete variables or PDF $f(x)$ for continuous variables	Uses joint PMF $p(x, y)$ for discrete variables or joint PDF $f(x, y)$ for continuous variables
Probability is **sum** for discrete variables or **area** for continuous variables	Probability is **double sum** for discrete variables or **volume** over a region for continuous variables
Compute $P(X = x)$ $P(a < X < b)$	Compute $P(X = x \text{ and } Y = y)$ $P(a < X < b \text{ and } c < Y < d)$
Mean and **variance** describe one variable	**Covariance** and **correlation** describe the relationship between two variables
Python Applications (in application sessions): plot one distribution	Python Applications: scatter plots, heatmaps, and 3D surfaces

Joint PMF for Two Discrete Random Variables

A **joint probability mass function**, or **joint PMF**, is used when both random variables are discrete. Suppose:

$$X = \textit{Wi-Fi signal level}$$

$$Y = \textit{video call quality}$$

where $Y = 1, 2, \& \ 3$ represents poor, acceptable, and clear video call quality, respectively

The **joint PMF** gives the **joint probability** that two variable values occur together, and it is written as:

$$p(x, y) = P(X = x \text{ and } Y = y)$$

Example Joint PMF Table:

	$X = 1$ *weak signal*	$X = 2$ *medium signal*	$X = 3$ *strong signal*
$Y = 1$ *poor call*	0.18	0.07	0.02
$Y = 2$ *acceptable call*	0.1	0.2	0.08
$Y = 3$ *clear call*	0.03	0.12	0.2

Each cell gives a joint probability for one combination of X and Y. For example:

$$P(X = 1 \textit{ and } Y = 3) = 0.03$$

The joint probability that the Wi-Fi signal is weak and the video call quality is clear is 0.03 (3%).

Joint PMF for Two Discrete Random Variables

The joint probability connects two pieces of information:

- the Wi-Fi signal is weak
- the video call quality is clear

This is useful because one variable alone does not give the full situation.

For example:

$$P(X = 1)$$

only tells us how often the Wi-Fi signal is weak.

$$P(Y = 3)$$

only tells us how often the call quality is clear.

But:

$$P(X = 1 \textit{ and } Y = 3)$$

tells us how often weak signal and clear video call quality happen together.

Conditions of a Joint PMF

A joint PMF must satisfy two conditions:

1. $p(x, y) \geq 0$, for all possible values of x and y
2. and

$$\sum_x \sum_y p(x, y) = 1$$

That means:

- No probability can be negative.
- The total of all individual joint probabilities must equal 1.

	$X = 1$	$X = 2$	$X = 3$	***Total***
$Y = 1$	0.18	0.07	0.02	**0.27**
$Y = 2$	0.1	0.2	0.08	**0.38**
$Y = 3$	0.03	0.12	0.2	**0.35**
Total	**0.31**	**0.39**	**0.3**	**1**

- Row totals and column totals should also lead to the same overall total.
- Therefore, this table is a valid joint PMF.

Discrete Marginal Distributions

- A joint PMF gives joint probabilities for combinations of X and Y:

$$p(x, y) = P(X = x \text{ and } Y = y)$$

- However, sometimes we want to study one variable only.
- From the previous example, we may want to know $P(X = 1)$, the probability that the Wi-Fi signal is weak, regardless of call quality.
- Or $P(Y = 1)$, the probability that the video call quality is poor, regardless of Wi-Fi signal level.
- A **marginal distribution** is the probability distribution of one random variable obtained by summing the joint probabilities over the other variable.
- Basically, a marginal distribution **reduces** the joint distribution into a single-variable distribution.
- To find the marginal distribution of X, add the joint probabilities over all values of Y:

$$p_X(x) = \sum_y p\,(x, y)$$

- To find the marginal distribution of Y, add the joint probabilities over all values of X:

$$p_Y(y) = \sum_x p\,(x, y)$$

- It is called marginal because these probabilities appear in the **margins** of the joint PMF table, as row totals and column totals.

Example: Finding Discrete Marginal Distributions

Using the same previous Wi-Fi Signal and Video Call Quality Example. The joint PMF table is:

	$X = 1$	$X = 2$	$X = 3$	$\boldsymbol{Total}$
$Y = 1$	0.18	0.07	0.02	$\mathbf{0.27}$
$Y = 2$	0.1	0.2	0.08	$\mathbf{0.38}$
$Y = 3$	0.03	0.12	0.2	$\mathbf{0.35}$
$\boldsymbol{Total}$	$\mathbf{0.31}$	$\mathbf{0.39}$	$\mathbf{0.3}$	**1**

Marginal Distribution of X:

$$p_X(1) = 0.18 + 0.10 + 0.03 = 0.31$$
$$p_X(2) = 0.07 + 0.20 + 0.12 = 0.39$$
$$p_X(3) = 0.02 + 0.08 + 0.20 = 0.30$$

Marginal Distribution of Y:

$$p_Y(1) = 0.18 + 0.07 + 0.02 = 0.27$$
$$p_Y(2) = 0.10 + 0.20 + 0.08 = 0.38$$
$$p_Y(3) = 0.03 + 0.12 + 0.20 = 0.35$$

1. What is the probability that the video call quality level is less than or equal to 2?

$$P(Y \leq 2) = p_Y(1) + p_Y(2) = 0.27 + 0.38 = 0.65$$

2. What is the probability that the Wi-Fi signal level is equal to 3?

$$P(X = 3) = p_X(3) = 0.30$$

Conditional Probability from a Joint PMF

- A conditional probability from a joint PMF tells us how likely one variable value is when another variable value is already known.

$$P(Y = y \mid X = x)$$

- In Unit 2, conditional probability was given as:

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

Note that:

- We first know that event A happened.
- Then we ask how likely event B is.
- $\cap$ = and

- For two discrete random variables X and Y, we use the same idea:

$$\boldsymbol{P(Y = y \mid X = x) = \frac{P(X = x \text{ and } Y = y)}{P(X = x)}}$$

or:

$$\boldsymbol{p_{Y|X}(y \mid x) = \frac{p(x, y)}{p_X(x)}}$$

where:

$$p(x, y) = P(X = x \text{ and } Y = y)$$
$$p_X(x) = P(X = x)$$

Example: Conditional Probability from a Joint PMF

Using also the same previous Wi-Fi Signal and Video Call Quality Example.

	$X = 1$	$X = 2$	$X = 3$	***Total***
$Y = 1$	0.18	0.07	0.02	**0.27**
$Y = 2$	0.1	0.2	0.08	**0.38**
$Y = 3$	0.03	0.12	0.2	**0.35**
Total	**0.31**	**0.39**	**0.3**	**1**

- What is the probability that the video call quality is poor, given that the Wi-Fi signal is weak?

$$P(Y = 1 \mid X = 1)$$

Use:

$$P(Y = 1 \mid X = 1) = \frac{P(X = 1 \text{ and } Y = 1)}{P(X = 1)}$$

From the table:

$$P(X = 1 \text{ and } Y = 1) = 0.18$$
$$P(X = 1) = 0.31$$

Therefore:

$$P(Y = 1 \mid X = 1) = \frac{0.18}{0.31} = 0.581$$

Among weak-signal cases, about **58.1%** result in poor video call quality.

Independence of Two Discrete Random Variables

- Two discrete random variables X and Y are **independent** if knowing the value of one variable does not change the probabilities of the other variable.
- From Unit 2, two events A and B were independent if:

$$P(A \cap B) = P(A)P(B)$$

- For two discrete random variables X and Y, we use the check:

$$P(X = x \text{ and } Y = y) = P(X = x)\, P(Y = y)$$

or,

$$p(x, y) = p_X(x)\, p_Y(y)$$

for **all** possible values of x and y.

Example: Using the Wi-Fi Signal and Video Call Quality Example, are Wi-Fi signal level X and video call quality Y independent?

Check one combination ($X = 1$ and $Y = 1$):

$$p(1,1) = 0.18, \qquad p_X(1) = 0.31, \qquad p_Y(1) = 0.27$$
$$p_X(1)\, p_Y(1) = 0.31 \times 0.27 = 0.0837 \neq p(1,1)$$

Therefore, X and Y are **NOT independent**. Wi-Fi signal level and video call quality are related.

Covariance Between Two Discrete Random Variables

- After studying whether two variables are independent, we may want to know how they move together.
- **Covariance** helps us answer questions such as:
  - When Wi-Fi signal level increases, does video call quality usually improve?
  - When the number of running apps increases, does battery usage usually increase?
  - When the number of active users increases, does response time usually increase?
- Covariance $\boldsymbol{Cov(X,Y)}$ gives us the **direction** of the relationship between variables X and Y:
  - If $\boldsymbol{Cov(X,Y) > 0}$: they increase together
  - If $\boldsymbol{Cov(X,Y) < 0}$: they move in opposite directions
  - If $\boldsymbol{Cov(X,Y) \approx 0}$: there is little or no linear relationship
- For two discrete random variables X and Y:

$$\mathbf{Cov}(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{E(XY) - E(X)E(Y)}$$

where:

$$E(XY) = \sum_x \sum_y xy\, p(x,y)\,, \qquad E(X) = \sum_x x\; p_X(x) \qquad , \qquad E(Y) = \sum_y y\; p_Y(y)$$

Correlation Between Two Discrete Random Variables

- Covariance is useful, but its value depends on the scale of the variables.
- For example, measuring time in seconds or milliseconds can change the covariance value.
- **Correlation** gives a clearer and easier-to-compare measure of relationship. It tells us:
  - the **direction** of the relationship
  - the **strength** of the relationship
- Unlike covariance, correlation is always between -1 *and* 1. So, it is easier to interpret.
- Formula:

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

where:

$$\sigma_X = \sqrt{\mathrm{Var}(X)} \quad \text{and} \quad \sigma_Y = \sqrt{\mathrm{Var}(Y)}$$

- If $\rho_{X,Y} > 0$, then X and Y increase together.
- If $\rho_{X,Y} < 0$, they move in opposite directions.
- If $\rho_{X,Y} = 0$, there is no linear relationship.

Example: Covariance and Correlation from a Joint PMF

Using the Wi-Fi Signal and Video Call Quality Example, find the covariance and correlation between X and Y.

	$X = 1$	$X = 2$	$X = 3$	***Total***
$Y = 1$	0.18	0.07	0.02	**0.27**
$Y = 2$	0.1	0.2	0.08	**0.38**
$Y = 3$	0.03	0.12	0.2	**0.35**
Total	**0.31**	**0.39**	**0.3**	**1**

Step 1: Find $E(X)$, $E(Y)$, $E(X^2)$, $E(Y^2)$, and $E(XY)$

$$E(X) = \sum x\ p(x) = 1.99\ , \qquad E(X^2) = \sum x^2\ p(x) = 4.57$$

$$E(Y) = \sum y\ p(y) = 2.08\ , \qquad E(Y^2) = \sum y^2\ p(y) = 4.94$$

$$E(XY) = \sum_x \sum_y xy\ p(x,y) = 4.47$$

Step 2: Find Covariance

$$\boldsymbol{Cov(X,Y)} = E(XY) - E(X)E(Y) = 4.47 - (1.99)(2.08) = 0.331$$

Step 3: Find Correlation

$$\sigma_X = \sqrt{E(X^2) - E(X)^2} \approx 0.781 \quad , \qquad \sigma_Y = \sqrt{E(Y^2) - E(Y)^2} \approx 0.784$$

$$\boldsymbol{\rho_{X,Y}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{0.331}{(0.781)(0.784)} \approx 0.541$$

This indicates a moderate positive relationship between X and Y.

Joint PDF for Two Continuous Random Variables

- So far, we studied two discrete random variables using a joint PMF. Now we move to two **continuous** random variables.
- A joint probability density function, or joint PDF, is used when both random variables are continuous.
- Example: $X = \textit{network delay in milliseconds}$
  $Y = \textit{server processing time in milliseconds}$
  Both X and Y can take any value in an interval, so they are continuous random variables.

- **Joint PDF** describes how probability is distributed over pairs of values (x, y). It is written as:

$$f(x,y) \quad \textit{or} \quad f_{X,Y}(x,y)$$

- $f(x,y) \neq P(X = x \textit{ and } Y = y)$ *because* $P(X = x \textit{ and } Y = y) = 0$
- $f(x,y)$ is a **density value** that helps us calculate probability over a region.

$$\boldsymbol{P(a < X < b \textit{ and } c < Y < d) = \int_a^b \int_c^d f(x,y)\, dy\, dx}$$

- ➢ Joint PMF: probability is found by adding table values.
- ➢ Joint PDF: probability is found by area/volume over a region.

Joint PDF for Two Continuous Random Variables

- A joint PDF $f(x, y)$ must satisfy **two** conditions.
  1. $f(x, y) \geq 0$ for all possible values of $x$and y.
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f\,(x, y) dy\, dx = 1$

Example: Probability from a Joint PDF

Let: $X =$ *processing time of a computing task* in seconds
$Y =$ *energy used to complete the same task* in joules

Suppose X and Y have the following joint PDF, find $P(0 < X < 1 \text{ and } 0 < Y < 2)$.

$$f(x, y) = \begin{cases} \frac{x + y}{24}, & 0 < x < 2,\ 0 < y < 4 \\ 0, & \text{otherwise} \end{cases}$$

$$P(0 < X < 1 \text{ and } 0 < Y < 2) = \int_0^1 \int_0^2 \frac{x + y}{24}\, dy\, dx = \frac{1}{24}\left[x^2 + 2x\right]_0^1 = \frac{3}{24} = 0.125$$

There is a 12.5% probability that the task takes between 0 and 1 seconds and uses between 0 and 2 joules of energy.

Marginal PDFs for Continuous Random Variables

- For a joint PDF $f(x, y)$, we may want to study one continuous random variable alone.
- This is similar to marginal distributions from a joint PMF, but now we use **integration** instead of summation.

- **Marginal PDF of X :**
  To get the PDF of X alone, integrate the joint PDF over all possible values of Y:

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

- **Marginal PDF of Y:**
  To get the PDF of Y alone, integrate the joint PDF over all possible values of X:

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

Example: Marginal PDFs from a Joint PDF

Using the same previous example:

$X =$ *processing time of a computing task in seconds*

$Y =$ *energy used to complete the same task in joules*

$$f(x,y) = \begin{cases} \dfrac{x+y}{24}, & 0 < x < 2,\ 0 < y < 4 \\ 0, & \textit{otherwise} \end{cases}$$

Marginal PDF of X:

$$f_X(x) = \int_0^4 \frac{x+y}{24}\, dy = \frac{4x+8}{24} = \frac{x+2}{6}, \qquad 0 < x < 2$$

Marginal PDF of Y:

$$f_Y(y) = \int_0^2 \frac{x+y}{24}\, dx = \frac{2+2y}{24} = \frac{1+y}{12}, \qquad 0 < y < 4$$

Independence of Two Continuous Random Variables

- Two continuous random variables X and Y are **independent** if knowing one variable does not change the distribution of the other.
- For continuous random variables, independence means:

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

for **all** possible values of x and y.

Using the Processing Time and Energy Usage Example:

From the previous slide:

$$f(x,y) = \frac{x+y}{24}, \qquad f_X(x) = \frac{x+2}{6}, \qquad f_Y(y) = \frac{1+y}{12}$$

At $x = 0.5,\ y = 1$:

$$f(0.5,1) = 1.5/24 = 0.0625$$
$$f_X(0.5) = 2.5/6 = 0.4167$$
$$f_Y(1) = 2/12 = 0.1667$$

$$f_X(0.5)f_Y(1) \approx 0.0695 \neq f(0.5,1)$$

Therefore, X and Y are **not independent**.

Covariance and Correlation for Continuous Random Variables

Similar to the joint PMF case, covariance and correlation can also be computed for two continuous random variables. The main difference is that we use integration instead of summation.

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \qquad \textit{and} \qquad \sigma_X = \sqrt{E(X^2) - [E(X)]^2}$$

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy \qquad \textit{and} \qquad \sigma_Y = \sqrt{E(Y^2) - [E(Y)]^2}$$

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, y f(x,y) dy\, dx$$

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

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

Multinomial Distribution

- The **multinomial distribution** is an extension of the binomial distribution.
- In the binomial distribution, each trial has only two possible outcomes (success *or failure*).
- In the multinomial distribution, each trial can have more than two possible outcomes.

Example:

Suppose a cloud storage system receives uploaded files. Each uploaded file is classified into one of three types:

$1 = \textit{image file}$ $2 = \textit{document file}$ $3 = \textit{video file}$

If we observe n uploaded files, we can define:

$$X_1 = \textit{number of image files},$$
$$X_2 = \textit{number of document files}$$
$$X_3 = \textit{number of video files}$$

Then $X_1 + X_2 + X_3 = n$

Assume $p1, p2, and\ p3$ are the probabilities of each category of uploaded file and are given as:

$$p_1 = 0.50, \quad p_2 = 0.30, \quad p_3 = 0.20$$

If $n = 10$ files are uploaded, we can for example find $P(X_1 = 5, X_2 = 3, X_3 = 2)$

Multinomial Distribution Formula

- Suppose there are k possible categories. Let $p_1, p_2, \ldots, p_k$ be the probabilities of the categories, where:

$$p_1 + p_2 + \cdots + p_k = 1$$

Let:

$$X_1, X_2, \ldots, X_k$$

be the number of outcomes in each category, where:

$$x_1 + x_2 + \cdots + x_k = n$$

- Formula:

$$\boldsymbol{P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!\, x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}}$$

- If $k = 2$, the multinomial distribution becomes the binomial distribution covered in unit 3.

$$P(X_1 = x_1, X_2 = x_2) = \frac{n!}{x_1! x_2!} p_1^{x_1} p_2^{x_2}$$

Since $p_1 = p, p_2 = 1 - p$, and $x_1 + x_2 = n$, this probability becomes:

$$P(X = x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

Example: Multinomial Distribution

Using the cloud storage file type example given earlier:

$$p_1 = 0.50,\ \ p_2 = 0.30\ ,\ \ p_3 = 0.20\ ,\quad n = 10$$

$$P(X_1 = 5, X_2 = 3, X_3 = 2) = \frac{10!}{5!\,3!\,2!}(0.50)^5(0.30)^3(0.20)^2$$

$$= 2520(0.50)^5(0.30)^3(0.20)^2$$

$$= 2520(0.03125)(0.027)(0.04)$$

$$\approx 0.0851$$

There is approximately an 8.51% probability that among 10 uploaded files:

- 5 are images
- 3 are documents
- 2 are videos

Bivariate Normal Distribution

- In Unit 4, we studied the normal distribution for one continuous random variable.
- In Unit 5, we study two continuous random variables together through the joint PDF.
- A **bivariate normal distribution** is a **joint PDF** for two normally distributed continuous random variables.

  X is normally distributed *variable* with $\mu_X,\ \sigma_X$

  Y is normally distributed *variable* with $\mu_Y,\ \sigma_Y$

- Together, they are described by one joint density $f(x, y)$
- The two variables may be related or not related.
- The relationship between X and Y is described by the correlation ρ :
  - If $\rho > 0$, they tend to increase together (positive relationship).
  - If $\rho < 0$, they move in opposite directions (negative relationship).
  - If $\rho = 0$, there is no linear relationship.
- A bivariate normal distribution is like a 3D bell shape for two variables and is determined by the five parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$.

Unit 5 Summary

- In this unit, we studied how to describe and analyze two random variables together.
- A joint distribution describes X and Y together.
- For discrete variables, we use a joint PMF: $p(x, y) = P(X = x \text{ and } Y = y)$
- For continuous variables, we use a joint PDF $f(x, y)$ and probabilities are found over a region using integration.
- A marginal distribution studies one variable alone by summing or integrating over the other variable:
  - $p_X(x) = \sum_y p\,(x, y)$
  - $f_X(x) = \int_{-\infty}^{\infty} f\,(x, y)dy$
- Independence means the joint distribution equals the product of the marginal distributions:
  - $p(x, y) = p_X(x)p_Y(y)$
  - $f(x, y) = f_X(x)f_Y(y)$
- Covariance tells the direction of the relationship.
- Correlation tells the direction and strength of the relationship.
- Multinomial distribution extends binomial distribution to more than two categories.
- Bivariate normal distribution is a joint PDF for two normally distributed continuous variables.