BSTA 2104 PROBABILITY AND STATISTICS II

UNIT LECTURER: MARY KAMINA

2024/2025 SEP-DEC SEMESTER

GROUPS:

BCS 2.1, BDAT 2.1, BIT 3.1, BSEN 2.1

Pre-requisite: BSTA 1203 Probability and Statistics I

Course Purpose
The course explores probability distributions of functions of random variables, an essential skill in proving standard statistical results.

Expected Learning Outcomes
At the end of this course, students should be able to:

1. Explain the concept of both univariate/one-dimensional and bivariate/two-dimensional random variables.
2. Evaluate the distribution of functions of random variables and calculate expectations.
3. Calculate conditional means and variances based on bivariate/two-dimensional distributions.
4. Establish the relationship (link) between various probability distributions.

   Course Content

   (a) Review of Random variables, probability distributions and mathematical expectation.
   (b) Law of large numbers.
   (c) Conditional probability distributions.
   (d) Marginal and conditional probabilities of bi-variate discrete distributions.
   (e) Marginal and conditional probabilities of continuous distributions.
   (f) Covariance and correlation coefficients.
   (g) Conditional expectation and variance.
   (h) Moments and probability generating function.
   (i) Probability distributions; Binomial, Poisson, hypergeometric, exponential, normal, beta and gamma and their links.
   (j) Bivariate probability distributions; bivariate normal distribution and transformation of variables.

For this unit, other than the notes provided here, use this website in tandem https://www.probabilitycourse.com/preface.php.

1 Review of Random Variables, Probability Distributions and Mathematical Expectation

Random Variables

A random variable is a numerical outcome of a random process. It can be classified into three main types:

1. Discrete Random Variables
   Take on a finite or countably infinite set of values. Examples include the number of heads in a series of coin tosses or the outcome of rolling a die.
2. Continuous Random Variables
   Can take on any value within a range. Examples include the exact height of students in a class or the time taken to run a race.
3. Mixed Random Variables
   These are random variables that are neither discrete nor continuous, but a mixture of both.
   To learn more about Mixed random variables click on this link https://www.probabilitycourse.com/chapter4/4_3_1_mixed.php

Probability Distributions

Probability distributions describe how the values of a random variable are distributed. The main types are:

1. Discrete Distributions
   For example, the binomial distribution describes the number of successes in a fixed number of independent Bernoulli trials.
2. Continuous Distributions
   For example, the normal distribution describes data that clusters around a mean.

Some Distributions

1. Binomial Distribution
   - PMF: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, \ldots, n$.
   - Mean: $E[X] = np$
   - Variance: $\text{Var}(X) = np(1-p)$
2. Normal Distribution
   - PDF: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
   - Mean: μ

- Variance: σ^2

3. Exponential Distribution

   - PDF: $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.
   - Mean: $E[X] = \frac{1}{\lambda}$
   - Variance: $\text{Var}(X) = \frac{1}{\lambda^2}$

There are several important concepts that need to be understood about Probability distributions.

1. Discrete Probability Distributions
   Probability Mass Function (PMF): The PMF of a discrete random variable X, denoted $P(X = x)$, gives the probability that X takes the value x. The PMF satisfies the properties:

$$\sum_x P(X = x) = 1, \quad 0 \leq P(X = x) \leq 1 \; \forall \; x \tag{1}$$

   Example
   For a fair six-sided die, the PMF is $P(X = x) = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5$ or 6

2. Probability Density Function (PDF) for Probability Density Function (PDF): The PDF $f_X(x)$ of a continuous random variable X describes the relative likelihood for X to take on a given value. The probability that X lies within an interval $[a, b]$ is given by:

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx \tag{2}$$

   Example
   The standard normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3}$$

3. Cumulative Distribution Function (CDF) The CDF gives the probability that X takes a value less than or equal to x.

$$F_X(x) = P(X \leq x) \tag{4}$$

   (a) For a discrete random variable

$$F_X(x) = \sum_{t \leq x} P(X = t) \tag{5}$$

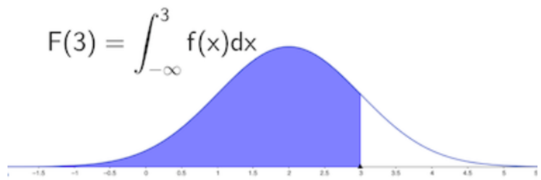

Figure 1: $F_X(3) = P(X \leq 3)$

(b) For a continuous random variable

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx \tag{6}$$

Properties of the CDF

1. $F_X(x)$ is non decreasing: $F_X(x_1) \leq F_X(x_2)$ if x_1¡ x_2
2. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$
3. $F_X(x)$ is right-continuous: $\lim_{x\to 0^+} F_X(x+h) = F_X(x)$

Theorems related to Probability Distributions

1. The Law of Total Probability
2. Bayes Theorem

Mathematical Expectation

Mathematical expectation, also known as the expected value (EV) or mean, is a fundamental concept in probability and statistics that provides a measure of the central tendency of a random variable.

Notation

For a discrete random variable X, the expected value is denoted as $E[X]$.
For a continuous random variable, it's often represented as μ or $E[X]$.

Discrete Random Variables

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$ and corresponding probabilities $P(X = x_i) = p_i$:

$$E[X] = \sum_{i=1}^{n} x_i \cdot p_i$$

Continuous Random Variables

For a continuous random variable with probability density function $f(x)$:

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$

Properties of Expectation

1. **Linearity of Expectation**: If X and Y are random variables, then:

   $$E[aX + bY] = aE[X] + bE[Y]$$

   This holds true regardless of whether X and Y are independent.

2. **Expectation of Constants**: If c is a constant, then:

   $$E[c] = c$$

3. **Non-negativity**: If $X \geq 0$ almost surely, then $E[X] \geq 0$.

4. **Law of Total Expectation**: If Y is another random variable, then:

   $$E[X] = E[E[X|Y]]$$

This link gives a general overview of probability distributions, showing what we are going to cover in this unit.
https://www.statlect.com/probability-distributions/

Now that we have reviewed probability distributions, lets consider their random variables. A probability distribution can either have one, two or more random variables. For this level, let us look at one and two random variables for Probability distributions.

If a distribution has a single random variable, it is called a univariate/one-dimensional distribution. This could either be a Univariate Discrete Probability distribution or a Univariate Continuous distribution. If it has two random variables, it is called a Bivariate/Two-Dimensional distribution.

This could either be a Bivariate Discrete distribution or a Bivariate Continuous distribution. A Bivariate Continuous distribution describes the joint behavior of two continuous random variables, X and Y. It is characterized by a joint probability density function (PDF).

Univariate/One-Dimensional Random Variables

A univariate/one-dimensional random variable is a variable that can take on different values based on the outcomes of a single random phenomenon. It is characterized by its probability distribution, which describes the likelihood of each possible value. We have two types of univariate/one-dimensional random variables can either be from a discrete or continuous distribution.

Examples for Univariate Random Variables

1. The height of students in a class, which can be any value within a range (e.g., 150 cm $\leq X \leq$ 200 cm).

2. Let X be a continuous random variable representing the height of adults in a certain population, measured in centimeters. The heights are normally distributed with a mean (μ) of 170 cm and a standard deviation (σ) of 10 cm.

   The probability density function for a normal distribution is given by:

   $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

   For our example, the PDF becomes:

   $$f(x) = \frac{1}{10\sqrt{2\pi}} e^{-\frac{(x-170)^2}{200}}$$

   To find the probability that a randomly selected adult has a height between 160 cm and 180 cm, we calculate:

   $$P(160 \leq X \leq 180) = \int_{160}^{180} f(x)\, dx$$

   This integral can be computed using standard normal distribution tables or computational tools.

   To find probabilities using the standard normal distribution, we convert the heights to Z-scores:

   $$Z = \frac{X - \mu}{\sigma}$$

   Calculating for our bounds:

   - For $X = 160$:
   $$Z = \frac{160 - 170}{10} = -1$$
   - For $X = 180$:
   $$Z = \frac{180 - 170}{10} = 1$$

   Now, we can use Z-tables:

   $$P(160 \leq X \leq 180) = P(-1 \leq Z \leq 1) \approx 0.6827$$

   Thus, there is approximately a 68.27% chance that a randomly selected adult from this population will have a height between 160 cm and 180 cm.

3. Let X be the outcome of rolling a fair six-sided die. The possible values are $1, 2, 3, 4, 5, 6$ with:

   $$P(X = x) = \frac{1}{6}, \quad \text{for } x = 1, 2, 3, 4, 5, 6.$$

4. Let X be a discrete random variable representing the outcome of rolling a fair six-sided die. The possible values for X are the integers $1, 2, 3, 4, 5, 6$.

   Since the die is fair, each outcome has an equal probability. The probability mass function $P(X = x)$ can be defined as:

$$P(X = x) = \frac{1}{6} \quad \text{for } x = 1, 2, 3, 4, 5, 6$$

   1. Probability of a Specific Outcome
   The probability of rolling a 3:

$$P(X = 3) = \frac{1}{6}$$

   2. Probability of Rolling an Even Number
   The even outcomes are $2, 4, 6$. Thus, the probability of rolling an even number is:

$$P(X \text{ is even}) = P(X = 2) + P(X = 4) + P(X = 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

   3. Cumulative Distribution Function (CDF)
   The cumulative distribution function $F(x)$ gives the probability that X is less than or equal to x:

$$F(x) = P(X \leq x)$$

   For example: - $F(3) = P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$

Bivariate/Two-Dimensional Random Variables

Bivariate random variables involve two random variables, typically denoted as X and Y. They describe the outcomes of two related random phenomena, allowing for the analysis of the relationship between them.

Joint Probability Distribution

1. **Discrete Case**: For discrete random variables, the joint probability distribution is given by the joint probability mass function $P(X = x, Y = y)$.

2. **Continuous Case**: For continuous random variables, the joint distribution is defined by a joint probability density function $f(x, y)$, where:

$$P(A) = \iint_A f(x, y)\, dx\, dy$$

   for any region A.

Independence

Two random variables X and Y are independent if:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

for all x and y.

Examples for Bivariate Random Variables

1. Let X be the height (in cm) and Y be the weight (in kg) of individuals in a sample. The joint distribution can help analyze how height influences weight. A possible joint PMF could be:

   $$P(X = 170, Y = 65) = 0.1$$

   This indicates a 10% chance of selecting an individual who is 170 cm tall and weighs 65 kg.

2. Consider X as the score in Mathematics and Y as the score in English for a group of students. The joint distribution may reveal trends such as higher math scores correlating with higher English scores, which could be visualized using a scatter plot.

Covariance and Correlation

The relationship between two bivariate random variables can also be measured using:

- **Covariance**:

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

- **Correlation Coefficient**:

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

  which standardizes the covariance to a range between -1 and 1, indicating the strength and direction of the linear relationship.

Examples: Discrete and Continuous Distributions

Discrete Case

1. Consider a fair six-sided die. The expected value is:

   $$E[X] = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5$$

2. Let X be the number of heads in two tosses of a fair coin. The values and probabilities are:

$$P(X=0) = \frac{1}{4},$$
$$P(X=1) = \frac{1}{2},$$
$$P(X=2) = \frac{1}{4}.$$

The expected value is:

$$E[X] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 0 + \frac{1}{2} + \frac{2}{4} = 1.$$

3. Suppose you win KES 10 with probability $\frac{1}{10}$, KES 20 with probability $\frac{1}{5}$, and KES 0 with probability $\frac{7}{10}$. Then:

$$E[X] = 10 \cdot \frac{1}{10} + 20 \cdot \frac{1}{5} + 0 \cdot \frac{7}{10} = 1 + 4 = KES5.$$

Continuous Case

1. For a uniform distribution on the interval $[0, 1]$:

$$E[X] = \int_0^1 x \cdot 1 \, dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}.$$

2. For a normal distribution $X \sim N(\mu, \sigma^2)$, the expected value is:

$$E[X] = \mu.$$

For example, if $\mu = 5$ and $\sigma^2 = 2$, then $E[X] = 5$.

Examples: Probability Distributions and Mathematical Expectation

Theoretical Examples

1. Prove that if X is a random variable that only takes the values 0 and 1, then $\mathrm{Var}(X) = E[X](1 - E[X])$.
Solution

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

Since $X^2 = X$ when X is 0 or 1:

$$E[X^2] = E[X]$$

Thus:

$$\mathrm{Var}(X) = E[X] - (E[X])^2 = E[X](1 - E[X])$$

2. Show that if X is a constant random variable, then $\text{Var}(X) = 0$.
Solution
If $X = c$, a constant:

$$\text{Var}(X) = E[(X - E[X])^2] = E[(c - c)^2] = E[0] = 0$$

3. Prove that the expected value of the sum of independent random variables is the sum of their expected values.
Solution
Let $X_1, X_2, \ldots, X_n$ be independent random variables:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Using the linearity of expectation:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

The result holds as the sum is simply the sum of expectations. What if you have random variables X and Y? See the link https://www.youtube.com/watch?v=7KeV3wLw0_o

Practical Examples

4. A fair six-sided die is rolled. Define a random variable X as the outcome of the roll. What is $E[X]$ and $\text{Var}(X)$?
Solution

$$P(X = x) = x, x = 1, 2, 3, 4, 5, 6$$
$$E[X] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5$$
$$\text{Var}(X) = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} - (3.5)^2 = \frac{91}{6} - 12.25 \approx 2.92$$

5. If X is a random variable representing the number of heads in three coin flips, find $E[X]$ and $\text{Var}(X)$.
Solution
- X can take values 0, 1, 2, 3 with probabilities $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$:

$$P(X = 0) = \frac{1}{8}, P(X = 1) = \frac{3}{8}, P(X = 2) = \frac{3}{8}, P(X = 3) = \frac{1}{8}$$

$$E[X] = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = 1.5$$
$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{10.5}{8} - (1.5)^2 = 0.75$$

6. A random variable X represents the number of students who pass an exam. If the probability that any student passes is 0.8, and there are 10 students, what are $E[X]$ and $\mathrm{Var}(X)$?
Solution
X follows a binomial distribution $X \sim \mathrm{Binomial}(10, 0.8)$:

$$E[X] = np = 10 \times 0.8 = 8$$

How would you interpret the $E[X] = 8$?

$$\mathrm{Var}(X) = np(1-p) = 10 \times 0.8 \times 0.2 = 1.6$$

Theoretical Examples: Discrete and Continuous Distributions

1. Proof that the sum of the probability mass function (PMF) over all possible values equals 1.

Problem breakdown
Prove that for a discrete random variable X with probability mass function $p_X(x)$, the sum of $p_X(x)$ over all possible values of x is equal to 1.

Solution
The probability mass function (PMF) $p_X(x)$ of a discrete random variable X is defined as:

$$p_X(x) = P(X = x)$$

This represents the probability that X takes a specific value x.

Since X can take one of a countable set of values, say $x_1, x_2, \ldots$, the total probability for all possible values of X must sum to 1:

$$\sum_{x_i} p_X(x_i) = 1$$

This equation is a consequence of the axioms of probability, particularly the law of total probability, which states that the sum of probabilities of all mutually exclusive and exhaustive outcomes must equal 1.

Consider a simple case where X can take only three values: x_1, x_2, x_3. The total probability is:

$$p_X(x_1) + p_X(x_2) + p_X(x_3) = 1$$

For a general discrete random variable X, which can take infinitely many values $x_1, x_2, x_3, \ldots$, the sum extends over all these values:

$$\sum_{i=1}^{\infty} p_X(x_i) = 1$$

Therefore, by the definition of a probability mass function, the sum of $p_X(x)$ over all possible values x must equal 1, completing the proof.

2. Proof of the limits of the Cumulative Distribution Function (CDF).

   Problem breakdown
   Show that if X is a random variable with a cumulative distribution function (CDF) $F_X(x)$, then $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.

   Solution
   The cumulative distribution function (CDF) $F_X(x)$ of a random variable X is defined as:
   $$F_X(x) = P(X \leq x)$$
   This represents the probability that the random variable X takes a value less than or equal to x.

   To prove the first part, $\lim_{x\to-\infty} F_X(x) = 0$: As x approaches $-\infty$, the probability $P(X \leq x)$ becomes smaller and smaller, because X has a lower probability of taking extremely negative values. In the limit, as x approaches $-\infty$, the CDF approaches 0 because:
   $$\lim_{x\to-\infty} P(X \leq x) = 0$$
   To prove the second part, $\lim_{x\to\infty} F_X(x) = 1$: As x approaches ∞, the probability $P(X \leq x)$ increases because X has a higher probability of taking values less than or equal to x. In the limit, as x approaches ∞, the CDF approaches 1 because:
   $$\lim_{x\to\infty} P(X \leq x) = 1$$
   This result holds because the total probability of all possible values of X must sum to 1.

   Thus, the limits of the cumulative distribution function are:
   $$\lim_{x\to-\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F_X(x) = 1$$

3. Proof that the expectation of a constant random variable is the constant itself.

   Problem breakdown
   Prove that the expectation of a constant random variable c is c.

   Solution
   Let X be a constant random variable, meaning X takes the value c with probability 1:
   $$P(X = c) = 1$$

- For all other values, X has probability 0:

$$P(X = x) = 0 \quad \text{for} \quad x \neq c$$

The expectation (or mean) of X is defined as:

$$E[X] = \sum_x x \cdot p_X(x)$$

Since $X = c$ with probability 1, the expectation simplifies to:

$$E[X] = c \cdot P(X = c) = c \cdot 1 = c$$

Therefore, the expectation of a constant random variable $X = c$ is simply the constant c, completing the proof.

4. Proof that the mean of a Poisson distribution is λ.

   Problem breakdown
   Prove that the mean of a Poisson distribution with parameter λ is λ.

   Solution

   Let X be a Poisson random variable with parameter λ, i.e., $X \sim \text{Poisson}(\lambda)$. The probability mass function is given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for} \quad k = 0, 1, 2, \ldots$$

   The expectation (mean) of X is defined as:

$$E[X] = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!}$$

   We can manipulate the summation to express it in a simpler form. Notice that:

$$k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \cdot \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!}$$

   This allows us to rewrite the expectation as:

$$E[X] = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!}$$

   Changing the index of summation by setting $j = k - 1$, we obtain:

$$E[X] = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!}$$

The series:

$$\sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!}$$

is recognized as the sum of the Taylor series expansion of e^{λ}, which equals 1. Therefore:

$$E[X] = \lambda \times 1 = \lambda$$

Thus, the mean of a Poisson distribution with parameter λ is λ, completing the proof.

5. Proof that the CDF of a normal distribution is given by the error function.

Problem breakdown
Show that if X follows a normal distribution $N(\mu, \sigma^2)$, then the CDF of X is given by:

$$F_X(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

where erf is the error function.

Solution
The cumulative distribution function (CDF) $F_X(x)$ of a normally distributed random variable X with mean μ and variance σ^2 is defined as:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)d(t)$$

Let X be a random variable with a normal distribution $X \sim N(\mu, \sigma^2)$. The cumulative distribution function (CDF) $F_X(x)$ is defined as:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$$

where $f_X(t)$ is the probability density function (PDF) of X:

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

To express this in terms of the error function, we perform a substitution:

$$u = \frac{t-\mu}{\sigma\sqrt{2}}$$

which implies:

$$du = \frac{dt}{\sigma\sqrt{2}}$$

Thus, the integral becomes:

$$F_X(x) = \int_{-\infty}^{\frac{x-\mu}{\sigma\sqrt{2}}} \frac{1}{\sqrt{\pi}} e^{-u^2}\, du$$

Recognizing this integral as the definition of the error function $\text{erf}(u)$, we get:

$$F_X(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Thus, the CDF of a normal distribution is given by the error function.

6. Proof that the Variance of a Binomial Distribution is $np(1-p)$

   Problem breakdown

   Prove that the variance of a binomial distribution with parameters n and p is $np(1-p)$.

   Solution

   Let X be a random variable following a binomial distribution with parameters n and p, denoted $X \sim \text{Binomial}(n, p)$. The probability mass function (PMF) of X is:

   $$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

   for $k = 0, 1, 2, \ldots, n$.

   The mean (expectation) of X is given by:

   $$E[X] = np$$

   To find the variance, $\text{Var}(X)$, we use:

   $$\text{Var}(X) = E[X^2] - (E[X])^2$$

   First, we find $E[X^2]$. Using the fact that:

   $$E[X^2] = E[X(X-1) + X]$$

   we have:

   $$E[X(X-1)] = n(n-1)p^2$$

   and:

   $$E[X] = np$$

   Thus:

   $$E[X^2] = n(n-1)p^2 + np$$

   Substituting this into the variance formula:

   $$\text{Var}(X) = E[X^2] - (E[X])^2 = n(n-1)p^2 + np - (np)^2$$

   $$\text{Var}(X) = np(1-p)$$

   Therefore, the variance of a binomial distribution with parameters n and p is $np(1-p)$.

2 Law of Large Numbers (LLN)

The Law of Large Numbers (LLN) states that as the sample size n increases, the sample average of independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, X_n$ approaches the expected value $\mu = E[X_i]$. We have two kind os expectations There are two main forms:

1. Weak LLN (WLLN)
   The sample mean converges in probability to the population mean as $n \to \infty$.

2. Strong LLN (SLLN)
   The sample mean converges almost surely to the population mean as $n \to \infty$.

Weak Law of Large Numbers (WLLN)

The Weak Law of Large Numbers (WLLN) ensures that for any $\epsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) = 0.$$

Derivation of WLLN

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2.
Define the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Compute the expectation of the sample mean:

$$E[\overline{X}_n] = \mu.$$

Compute the variance of the sample mean:

$$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Apply Chebyshev's inequality:

$$P\left(\left|\overline{X}_n - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

As $n \to \infty$, the right-hand side goes to 0, proving the WLLN.

Strong Law of Large Numbers (SLLN)

The Strong Law of Large Numbers (SLLN) states that the sample mean converges almost surely to the population mean μ:

$$P\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1.$$

Derivation of SLLN

Define the partial sums $S_n = \sum_{i=1}^{n} X_i$.
Relate the sample mean to the partial sums: $\overline{X}_n = \frac{S_n}{n}$.
Use the Borel-Cantelli Lemma and advanced techniques to show that $\overline{X}_n$ converges almost surely to μ.

Central Limit Theorem (CLT)

The Central Limit Theorem (CLT) states that the sum of i.i.d. random variables, when normalized, converges in distribution to a normal distribution as $n \to \infty$:

$$Z_n = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

Derivation of CLT

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathrm{Var}(X_i)$. Define the normalized sum Z_n as:

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right).$$

As $n \to \infty$, the distribution of Z_n converges in distribution to a standard normal distribution:

$$Z_n \xrightarrow{d} N(0,1).$$

This means that the standardized sum of the random variables approaches a normal distribution with mean 0 and variance 1, regardless of the distribution of X_i, provided X_i has finite mean and variance.
To derive the CLT, we use characteristic functions, which are a powerful tool in probability theory for studying sums of random variables.
Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. The sum of these random variables is:

$$S_n = X_1 + X_2 + \cdots + X_n.$$

The sample mean is:

$$\overline{X}_n = \frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

We are interested in the behavior of S_n as $n \to \infty$. To standardize S_n, we subtract the expected value $n\mu$ and divide by the standard deviation $\sigma\sqrt{n}$, forming a normalized sum Z_n:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right).$$

Thus, Z_n is a standardized sum of the random variables. The goal is to show that Z_n converges in distribution to $N(0,1)$.
The characteristic function of a random variable X is defined as the expected value of e^{itX}:

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right].$$

The characteristic function uniquely determines the distribution of a random variable and helps in analyzing sums of independent random variables.
Let $Y_i = \frac{X_i - \mu}{\sigma}$. These are i.i.d. random variables with mean 0 and variance 1. The characteristic function of Y_i is:

$$\varphi_{Y_i}(t) = \mathbb{E}\left[e^{itY_i}\right].$$

Since the Y_i's are independent, the characteristic function of the normalized sum $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$ is:

$$\varphi_{Z_n}(t) = \left(\varphi_{Y_1}\left(\frac{t}{\sqrt{n}} \right) \right)^n.$$

We now expand $\varphi_{Y_i}(t)$ in a Taylor series around $t = 0$. Since $\mathbb{E}[Y_i] = 0$ and $\mathrm{Var}(Y_i) = 1$, we have:

$$\varphi_{Y_i}(t) = 1 - \frac{t^2}{2} + o(t^2).$$

Substituting this approximation into the characteristic function of Z_n, we get:

$$\varphi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n} \right) \right)^n.$$

For large n, we use the fact that $\left(1 + \frac{x}{n}\right)^n \to e^x$ as $n \to \infty$. Thus:

$$\varphi_{Z_n}(t) \approx \exp\left(-\frac{t^2}{2} \right).$$

This is the characteristic function of the standard normal distribution $N(0,1)$. Since the characteristic function of Z_n converges to the characteristic function of $N(0,1)$, we conclude that Z_n converges in distribution to $N(0,1)$:

$$Z_n \xrightarrow{d} N(0,1).$$

Relationship Between LLN and CLT

1. The LLN ensures that the sample mean converges to the population mean as $n \to \infty$.
2. The CLT describes the distribution of the sample mean for large, but finite n, and shows that it becomes approximately normal.
3. **Key Difference**: The LLN is about convergence to the mean, while the CLT is about the distribution of the sample mean for finite samples.

Theoretical Examples

1. Prove the Weak Law of Large Numbers using Chebyshev's inequality for i.i.d. random variables $X_1, X_2, \ldots, X_n$ with mean μ and variance σ^2.

   Solution
   Compute $E[\overline{X}_n] = \mu$.
   Compute $\text{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$. Apply Chebyshev's inequality:

   $$P\left(|\overline{X}_n - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

   As $n \to \infty$, the probability goes to 0, proving the WLLN.

2. Sketch the proof of the Strong Law of Large Numbers for i.i.d. random variables $X_1, X_2, \ldots, X_n$ with mean μ and variance σ^2.

   Solution
   Define $S_n = \sum_{i=1}^{n} X_i$. Show that $\overline{X}_n = \frac{S_n}{n}$ converges almost surely to μ.
   (Hint: Use the Borel-Cantelli Lemma)

3. Show that the sample proportion in a sequence of Bernoulli trials with probability of success p converges to p using the WLLN.

   Solution
   Let $X_i = 1$ if the i-th trial is a success, and 0 otherwise. Then X_i has mean p and variance $p(1-p)$.
   Use WLLN to show that the sample proportion $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges to p in probability.

4. Derive the Central Limit Theorem for i.i.d. random variables $X_1, X_2, \ldots, X_n$ with mean μ and variance σ^2.

   Solution

Define $Z_n = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}}$.
Show that $Z_n \xrightarrow{d} N(0, 1)$.

5. Prove the Strong Law of Large Numbers using martingale techniques (advanced).

   Solution
   (Hint: use the Martingale Convergence Theorem).

Practical Examples

1. You have collected daily temperatures over 365 days. Use the Weak Law of Large Numbers to estimate the population mean temperature.

   Solution
   Let X_i be the temperature on day i.
   Compute the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
   By the WLLN, $\overline{X}_n$ converges in probability to the true population mean temperature as $n \to \infty$.

2. Completion times of 100 employees are recorded. Estimate the average completion time using the sample mean and explain why this is a good estimate using the Weak LLN.

   Solution
   Let X_i be the completion time for the i-th employee.
   Compute the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
   By the WLLN, the sample mean approximates the population mean completion time as n grows large.

3. Heights of 1000 individuals are measured. Estimate the probability that the sample mean height is within 2 cm of the population mean using the Central Limit Theorem.

   Solution
   Assume the heights are normally distributed with mean μ and variance σ^2.
   Use the CLT to approximate the distribution of the sample mean:

$$\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

   Compute the probability that $|\overline{X}_n - \mu| \leq 2$ cm.

4. Over 200 days, the number of customers visiting a store is recorded daily. Use the Strong Law of Large Numbers to explain why the average number of visits will stabilize as the number of days increases.

   Solution
   Let X_i be the number of customer visits on day i. Compute the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
   By the SLLN, $\overline{X}_n$ will almost surely converge to the true mean number of customer visits as $n \to \infty$.

5. A sample of 1000 products is inspected, and 5% are found to be defective. Estimate the probability that the proportion of defective products in future samples will be within 1% of the true defect rate using the Central Limit Theorem.

   Solution
   Let X_i be 1 if the i-th product is defective, and 0 otherwise. The sample proportion is $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
   Apply the CLT to approximate the distribution of $\overline{X}_n$.
   Use this distribution to compute the probability that the sample proportion is within 1% of the true defect rate.

Here is a link to more examples in this lesson
https://www.probabilitycourse.com/chapter7/7_1_3_solved_probs.php

3 Conditional Probability Distributions

Conditional Probability

Conditional probability quantifies the likelihood of an event A occurring given that another event B has occurred. The formula is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{where} \quad P(B) > 0$$

Examples on Conditional Probability

1. What is the probability that it will rain tomorrow given that it is cloudy today? given that $P(R \cap C) = 0.15$ (Probability of rain and cloudy) and $P(C) = 0.5$ (Probability of cloudy)?

   Solution
   Identify events: R (Rain), C (Cloudy).
   Use the formula: $P(R|C) = \frac{P(R \cap C)}{P(C)}$.

Substitute values: $P(R|C) = \frac{0.15}{0.5}$.
Calculate: $P(R|C) = 0.3$.
Conclusion: The probability of rain given it is cloudy is 0.3.

2. What is the probability of passing an exam given that a student studied given $P(P \cap S) = 0.72$ (Probability of passing and studying) and $P(S) = 0.8$ (Probability of studying)

   Solution
   Identify events: P (Pass), S (Studied).
   Use the formula: $P(P|S) = \frac{P(P \cap S)}{P(S)}$.
   Substitute values: $P(P|S) = \frac{0.72}{0.8}$.
   Calculate: $P(P|S) = 0.9$.
   Conclusion: The probability of passing given studying is 0.9.

Conditional Probability Distributions

For two random variables X and Y, the conditional probability distribution of X given $Y = y$ is defined as:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad P(Y = y) > 0$$

Examples on Conditional Probability Distributions

1. What is the probability that a student received an 'A' given that they studied for 5 hours given $P(X = A, Y = 5) = 0.1$ and $P(Y = 5) = 0.25$?

   Solution
   Identify events: X (Grade), Y (Hours studied).
   Use the formula: $P(X = A|Y = 5) = \frac{P(X=A,Y=5)}{P(Y=5)}$.
   Substitute values: $P(X = A|Y = 5) = \frac{0.1}{0.25}$.
   Calculate: $P(X = A|Y = 5) = 0.4$.
   Conclusion: The probability of receiving an 'A' given 5 hours of study is 0.4.

2. What is the probability that food is served in under 10 minutes on Monday given $P(X = \text{Under 10 minutes}, Y = \text{Monday}) = 0.2$ and $P(Y = \text{Monday}) = 0.5$?

   Solution
   Identify events: X (Wait time), Y (Day).
   Use the formula: $P(X = \text{Under 10 minutes}|Y = \text{Monday}) = \frac{P(X=\text{Under 10 minutes},Y=\text{Monday})}{P(Y=\text{Monday})}$.

Substitute values: $P(X = \text{Under 10 minutes}|Y = \text{Monday}) = \frac{0.2}{0.5}$.
Calculate: $P(X = \text{Under 10 minutes}|Y = \text{Monday}) = 0.4$. Conclusion: The probability of food being served in under 10 minutes on Monday is 0.4.

More Examples for you to practise

1. A software system is tested by two teams. Team A detects bugs 70% of the time, and Team B detects bugs 80% of the time. Team A performs 40% of the tests, and Team B performs 60%. Given that a bug was detected, what is the probability that Team A conducted the test?

   Solution: Define A as Team A testing and D as a bug being detected. We want to find $P(A|D)$.
   Using Bayes' Theorem:

   $$P(A|D) = \frac{P(D|A)P(A)}{P(D)}$$

   First, calculate $P(D)$:

   $$P(D) = P(D|A)P(A) + P(D|B)P(B) = (0.7 \times 0.4) + (0.8 \times 0.6) = 0.74$$

   Now calculate $P(A|D)$:

   $$P(A|D) = \frac{0.7 \times 0.4}{0.74} = 0.378$$

2. A data center has two types of servers. Type A servers fail with a probability of 0.1 when CPU usage is high, while Type B servers fail with a probability of 0.2. 30% of the servers are Type A, and 70% are Type B. If a server fails, what is the probability that it is a Type A server?

   Solution
   Let A be Type A server, F be a server failure.
   We are asked to find $P(A|F)$.
   Using Bayes' Theorem:

   $$P(A|F) = \frac{P(F|A)P(A)}{P(F)}$$

   Compute $P(F)$:

   $$P(F) = P(F|A)P(A) + P(F|B)P(B) = (0.1 \times 0.3) + (0.2 \times 0.7) = 0.17$$

Now, calculate $P(A|F)$:

$$P(A|F) = \frac{0.1 \times 0.3}{0.17} = 0.176$$

3. A cloud service provider has two types of servers: high performance (HP) and low performance (LP). High-performance servers experience overload with a probability of 0.05, while low-performance servers experience overload with a probability of 0.2. 20% of the servers are high-performance. If an overload occurs, what is the probability it was a high-performance server?
   Solution
   Let H be the event of using a high-performance server, and O be an overload.
   We want to find $P(H|O)$.
   Use Bayes' Theorem:

   $$P(H|O) = \frac{P(O|H)P(H)}{P(O)}$$

   Calculate $P(O)$:

   $$P(O) = P(O|H)P(H) + P(O|L)P(L) = (0.05 \times 0.2) + (0.2 \times 0.8) = 0.17$$

   Now, calculate $P(H|O)$:

   $$P(H|O) = \frac{0.05 \times 0.2}{0.17} = 0.0588$$

4. In a SaaS (Software as a Service, is a cloud computing model that delivers software applications over the internet on a subscription basis) company, 60% of users who haven't logged in for over a month churn, while 20% of regular users churn. 25% of users haven't logged in for over a month. If a user churns, what is the probability that they hadn't logged in for over a month?

   Solution
   Define L as not logged in for over a month and C as churn.
   We want to calculate $P(L|C)$.
   Using Bayes' Theorem:

   $$P(L|C) = \frac{P(C|L)P(L)}{P(C)}$$

Calculate $P(C)$:

$$P(C) = P(C|L)P(L) + P(C|R)P(R) = (0.6 \times 0.25) + (0.2 \times 0.75) = 0.3$$

Now, calculate $P(L|C)$:

$$P(L|C) = \frac{0.6 \times 0.25}{0.3} = 0.5$$

4 Conditional Probability Distributions

A conditional probability distribution describes the probability of one random variable taking specific values given that another random variable has already taken a certain value. This applies to both discrete and continuous random variables.

Given two random variables X and Y, the conditional probability distribution of X given $Y = y$ is:

- For discrete random variables:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad P(Y = y) > 0$$

- For continuous random variables:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad f_Y(y) > 0$$

  where $f_{X|Y}(x|y)$ is the conditional probability density function (PDF) of X given $Y = y$.

Examples on Discrete Conditional Probability Distributions

1. Consider two systems, A and B, where system A fails with probability 0.2, and both fail together with probability 0.1. What is the conditional probability that system A fails given that system B fails?

   Solution
   Define the events:

$$A : \text{System A fails}, \quad B : \text{System B fails}$$

   Known values:

$$P(A \cap B) = 0.1, \quad P(B) = 0.2$$

Apply the conditional probability formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.2} = 0.5$$

Conclusion: The probability that system A fails given that system B fails is 0.5.

2. Suppose a bug detection system identifies bugs in two categories: X and Y. If the probability of finding a bug in category X is 0.6, and the joint probability of finding a bug in both categories is 0.3, what is the conditional probability of finding a bug in X given that there's a bug in Y?

   Solution
   Define the events:

   $$X : \text{Bug found in category X}, \quad Y : \text{Bug found in category Y}$$

   Known values:
   $$P(X \cap Y) = 0.3, \quad P(Y) = 0.5$$

   Apply the conditional probability formula:

   $$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{0.3}{0.5} = 0.6$$

   Conclusion: The conditional probability of finding a bug in X given that there is a bug in Y is 0.6.

3. A data science team tracks whether a customer clicks on a product link C given that they visited the product page V. If 30% of all users click on the link and 40% visit the page, with 20% clicking after visiting the page, find the conditional probability of a click given a visit.

   Solution
   Define the events:

   $$C : \text{Customer clicks}, \quad V : \text{Customer visits the page}$$

   Known values:
   $$P(C \cap V) = 0.2, \quad P(V) = 0.4$$

Apply the conditional probability formula:

$$P(C|V) = \frac{P(C \cap V)}{P(V)} = \frac{0.2}{0.4} = 0.5$$

Conclusion: The probability of a click given that the customer visited the page is 0.5.

4. A company manufactures two types of products, and it is known that 15% of all products are defective. Of the defective products, 5% are of type A, and 10% are of type B. If a product is defective, what is the probability it is of type A?

   Solution
   Define the events:

   $$A : \text{Product is of type A}, \quad D : \text{Product is defective}$$

   Known values:

   $$P(A \cap D) = 0.05, \quad P(D) = 0.15$$

   Apply the conditional probability formula:

   $$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{0.05}{0.15} = 0.333$$

   Conclusion: The probability that a defective product is of type A is approximately 0.333.

Examples on Continuous Conditional Probability Distributions

1. In a network, the time T to transfer data is modeled as a continuous random variable. The conditional distribution of T, given that network latency L is fixed at $l = 2$, is represented by:

   $$f_{T|L}(t|l) = \frac{f_{T,L}(t,l)}{f_L(l)}$$

   If $f_{T,L}(t,l) = e^{-(t+l)}$ and $f_L(l) = e^{-l}$, find $f_{T|L}(t|2)$.

Solution
Joint PDF:

$$f_{T,L}(t,l) = e^{-(t+l)}$$

Marginal PDF for L:

$$f_L(l) = e^{-l}$$

Apply the conditional PDF formula:

$$f_{T|L}(t|l) = \frac{e^{-(t+l)}}{e^{-l}} = e^{-t}$$

Conclusion: Therefore, $f_{T|L}(t|2) = e^{-t}$.

2. The CPU usage X at a given time is modeled as a continuous variable. Given that at a certain time the CPU usage is known to be 60%, find the conditional distribution of future CPU usage assuming the joint PDF is:

$$f_{X,Y}(x,y) = \lambda e^{-\lambda(x+y)}$$

and the marginal distribution $f_Y(y)$ is $\lambda e^{-\lambda y}$, where $\lambda = 1$.

Solution
Joint PDF:

$$f_{X,Y}(x,y) = e^{-(x+y)}$$

Marginal PDF for Y:

$$f_Y(y) = e^{-y}$$

Apply the conditional PDF formula:

$$f_{X|Y}(x|y) = \frac{e^{-(x+y)}}{e^{-y}} = e^{-x}$$

Conclusion: Therefore, $f_{X|Y}(x|0.6) = e^{-x}$.

3. A data scientist tracks customer spending X based on the time Y spent on the website. The joint distribution of X and Y is given by $f_{X,Y}(x,y) = 2e^{-2(x+y)}$. Find the conditional distribution $f_{X|Y}(x|y)$.

Solution
Joint PDF:

$$f_{X,Y}(x,y) = 2e^{-2(x+y)}$$

Marginal PDF for Y:

$$f_Y(y) = \int_0^\infty 2e^{-2(x+y)} dx = e^{-2y}$$

Apply the conditional PDF formula:

$$f_{X|Y}(x|y) = \frac{2e^{-2(x+y)}}{e^{-2y}} = 2e^{-2x}$$

Conclusion: The conditional distribution is $2e^{-2x}$.

The Law of Total Probability

The Law of Total Probability states that the probability of an event can be computed by considering all possible ways the event can occur based on a partition of the sample space:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Derivation of The Law of Total Probability

1. Express $P(A)$ as a sum over disjoint events B_i:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$$

2. Use the definition of conditional probability:

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

3. Substitute back into the sum:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Examples of the Law of Total Probability

1. What is the probability of server failure given hardware and software failures given $P(A|B_1) = 0.6$ (hardware), $P(B_1) = 0.3$ and $P(A|B_2) = 0.4$ (software), $P(B_2) = 0.7$?

   Solution
   Identify events: A (Server failure), B_1 (Hardware failure), B_2 (Software

failure).
Apply the Law of Total Probability:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$

Substitute values:

$$P(A) = (0.6)(0.3) + (0.4)(0.7)$$

Calculate:

$$P(A) = 0.18 + 0.28 = 0.46$$

Conclusion: The probability of server failure is 0.46.

2. What is the probability of a successful software deployment given $P(A|B_1) = 0.7$ (manual testing), $P(B_1) = 0.4$ and $P(A|B_2) = 0.9$ (automated testing), $P(B_2) = 0.6$?

   Solution
   Identify events: A (Success), B_1 (Manual testing), B_2 (Automated testing).
   Apply the Law of Total Probability:

   $$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$

   Substitute values:

   $$P(A) = (0.7)(0.4) + (0.9)(0.6)$$

   Calculate:

   $$P(A) = 0.28 + 0.54 = 0.82$$

   Conclusion: The probability of a successful software deployment is 0.82.

Bayes Theorem

Bayes' Theorem relates the conditional and marginal probabilities of random events:

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

Bayes Theorem Proof

Start with the definition of conditional probability:

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

Rewrite $P(A \cap B)$ using conditional probability:

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

Substitute into the first equation:

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

Click this link to see the differences between Bayes Theorem and Conditional Probability https://www.cuemath.com/data/bayes-theorem/

Examples on Bayes Theorem

1. What is the probability that a server is faulty given that it has crashed given $P(F) = 0.02$ (Faulty), $P(C|F) = 0.95$ (Crash given Faulty) and $P(C) = 0.1$?

   Solution
   Identify events: F (Faulty), C (Crash).
   Apply Bayes' Theorem:

   $$P(F|C) = \frac{P(C|F)P(F)}{P(C)}$$

   Substitute values:

   $$P(F|C) = \frac{(0.95)(0.02)}{0.1}$$

   Calculate:

   $$P(F|C) = \frac{0.019}{0.1} = 0.19$$

   Conclusion: The probability that the server is faulty given that it crashed is 0.19.

2. What is the probability that a file is malicious given it was flagged given $P(M) = 0.01$ (Malicious), $P(F|M) = 0.9$ (Flagged given Malicious) and $P(F) = 0.05$?

   Solution
   Identify events: M (Malicious), F (Flagged).
   Apply Bayes' Theorem:

   $$P(M|F) = \frac{P(F|M)P(M)}{P(F)}$$

   Substitute values:

   $$P(M|F) = \frac{(0.9)(0.01)}{0.05}$$

   Calculate:

   $$P(M|F) = \frac{0.009}{0.05} = 0.18$$

   Conclusion: The probability that the file is malicious given it was flagged is 0.18.

Conditional Means and Variances

Conditional Mean
It is the expected value of X given $Y = y$:

$$E[X|Y = y] = \sum_x xP(X = x|Y = y)$$

Conditional Variance
It is the variance of X given $Y = y$:

$$Var(X|Y = y) = E[X^2|Y = y] - (E[X|Y = y])^2$$

Examples on Conditional Means and Variances

1. What is the expected salary of an employee given 5 years of experience given salary distributions: $S = \{50k, 60k, 70k\}$ with probabilities based on experience?

   Solution
   Define possible salaries and probabilities for 5 years of experience.
   Calculate the conditional mean:

   $$E[S|Y = 5] = (50k)(0.3) + (60k)(0.4) + (70k)(0.3)$$

Calculate:

$$E[S|Y=5] = 15k + 24k + 21k = 60k$$

Conclusion: The expected salary given 5 years of experience is KES 60,000.

2. What is the expected bug fix time given that there are 3 developers given bug fix times: $T = \{5, 7, 10\}$ hours with probabilities $P(T=5) = 0.2$, $P(T=7) = 0.5$, $P(T=10) = 0.3$?

   Solution
   Calculate the expected time:

   $$E[T|D=3] = (5)(0.2) + (7)(0.5) + (10)(0.3)$$

   Calculate:

   $$E[T|D=3] = 1 + 3.5 + 3 = 7.5 \text{ hours}$$

   Conclusion: The expected bug fix time is 7.5 hours when 3 developers are working.

Is Bayes Theorem and Conditional Probability the same?

The **conditional probability** of an event A given another event B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{where } P(B) > 0$$

This formula tells us how to calculate the probability of A occurring given that B has already occurred.

Bayes' Theorem is a specific application of the conditional probability formula. It allows us to reverse conditional probabilities, i.e., to compute $P(A|B)$ when we know $P(B|A)$, $P(A)$, and $P(B)$. Bayes' Theorem is derived from the definition of conditional probability and is written as:

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

where $P(B)$ can be expanded using the **Law of Total Probability** if necessary:

$$P(B) = P(B|A)P(A) + P(B|\neg A)P(\neg A)$$

Key Differences

1. Conditional Probability Formula: Directly calculates the probability of one event given another.

2. Bayes' Theorem: Reverses the known conditional probability to find another conditional probability in cases where direct computation is not possible. It is derived from the conditional probability formula, but it is more specific as it gives a way to reverse or "update" probabilities based on new information.

5 Marginal and Conditional Probabilities of Bivariate/Two-Dimensional Discrete Distributions

Bivariate Discrete Random Variables

In many real-world problems, we are often interested in two or more variables simultaneously. These variables might depend on each other, and analyzing their behavior jointly gives us a better understanding of the system under study. For example, we might want to study the relationship between the number of defective products and the number of complaints received in a factory.

A bivariate discrete random variable consists of two discrete random variables, X and Y, which are defined on the same sample space Ω. The pair (X, Y) takes values in a finite or countably infinite set, and each value corresponds to an ordered pair (x, y) in the xy-plane.

Bivariate Discrete Probability Distributions

In probability theory, a bivariate discrete distribution refers to the probability distribution of two discrete random variables defined on the same sample space. These variables can be dependent or independent, and the relationship between them can be captured through joint, marginal, and conditional distributions. Understanding these distributions is crucial in fields such as data science, software engineering, computer science, IT and statistics, where two or more related phenomena are often studied together. Let's denote the two discrete random variables by X and Y

Joint Probability Mass Function (Joint PMF)

The joint probability mass function (PMF) of two discrete random variables X and Y is denoted by $P(X = x, Y = y)$ and gives the probability that X takes a specific value x, and Y takes a specific value y. Formally, the joint PMF satisfies two conditions:

1. $P(X = x, Y = y) \geq 0$
2. $\sum_x \sum_y P(X = x, Y = y) = 1$

The joint PMF describes the complete probabilistic behavior of X and Y together.

Examples on Joint Distribution of Two Random Variables

1. Consider the example of placing three balls into three cells. Let X represent the number of balls in Cell 1, and Y represent the number of cells occupied. The joint probability distribution for the random variables X and Y is shown in the Total column.

$X\backslash Y$	1	2	3	Total
0	$\frac{2}{27}$	$\frac{6}{27}$	0	$\frac{8}{27}$
1	0	$\frac{6}{27}$	$\frac{6}{27}$	$\frac{12}{27}$
2	0	$\frac{6}{27}$	0	$\frac{6}{27}$
3	$\frac{1}{27}$	0	0	$\frac{1}{27}$

2. In a computer network, packets are sent to two servers, and each packet either reaches its destination or is lost. Let X represent the number of packets reaching Server A and Y represent the number of packets reaching Server B. If the joint PMF is given by:

$X\backslash Y$	0	1	2	Total
0	0.1	0.1	0.05	0.25
1	0.1	0.2	0.15	0.45
2	0.05	0.1	0.15	0.30

How can we verify that this is a valid joint PMF?

Solution
We need to verify that the total probability sums to 1. Summing the table:

$$0.1 + 0.1 + 0.05 + 0.1 + 0.2 + 0.15 + 0.05 + 0.1 + 0.15 = 1$$

Thus, this is a valid joint PMF.

3. In an e-commerce platform, let X represent the number of items a customer buys in a single session, and Y represent the number of items returned. The joint distribution is given by:

$X\backslash Y$	0	1	2	Total
0	0.2	0.05	0.02	0.27
1	0.15	0.2	0.05	0.4
2	0.08	0.15	0.1	0.33

How can we compute the total probability?

Solution
We need to sum all joint probabilities to check that they equal 1:

$$0.2 + 0.05 + 0.02 + 0.15 + 0.2 + 0.05 + 0.08 + 0.15 + 0.1 = 1$$

Thus, the distribution is valid.

4. In a computer system, tasks are scheduled across two processors. Let X represent the number of tasks on Processor A, and Y represent the number of tasks on Processor B. The joint PMF is given by:

$X \backslash Y$	0	1	2	Total
0	0.15	0.1	0.05	0.3
1	0.2	0.2	0.1	0.5
2	0.05	0.1	0.05	0.2

What is the probability that Processor A handles exactly 1 task?

Solution
To find $P(X = 1)$, sum the probabilities for $X = 1$:

$$P(X = 1) = 0.2 + 0.2 + 0.1 = 0.5$$

5. In a software testing process, two test suites are run on the same code. Let X represent the number of failed test cases in Suite A, and Y represent the number of failed test cases in Suite B. The joint distribution is given as:

$X \backslash Y$	0	1	2	Total
0	0.3	0.15	0.05	0.5
1	0.1	0.15	0.05	0.3
2	0.05	0.1	0.05	0.2

What is the probability that neither test suite has any failed test cases?

Solution
The probability that neither test suite has any failed test cases is $P(X = 0, Y = 0) = 0.3$.

Marginal Probability

Marginal Probability from Set Theory

Marginal probability can be understood from set theory by considering it as the probability of a single event occurring without regard to other related events. It is derived from joint probabilities by "summing out" or "marginalizing" over the outcomes of the other events.

Definitions

- **Sample Space** (Ω): The set of all possible outcomes.
- **Event** (A): A subset of the sample space (i.e., $A \subseteq \Omega$) representing a specific outcome or a set of outcomes.
- **Joint Probability**: The probability of two events A and B happening simultaneously is denoted as $P(A \cap B)$, where $\cap$ is the intersection of the sets.

Marginal Probability from Set Theory

Given two events A and B, the marginal probability of event A, denoted $P(A)$, is the probability of event A happening, regardless of whether event B occurs. In terms of joint probabilities, the marginal probability of A can be expressed as:

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

Here, we are "summing over" all possible occurrences of B, considering every way that event A can occur in conjunction with different outcomes for event B.

Set Theory Interpretation

- Joint Probability $P(A \cap B)$
  This is the probability of both A and B occurring together, which is represented as the intersection of the sets A and B.
- Marginal Probability $P(A)$
  Marginal probability is the probability of being in the set A, regardless of the status of set B. This means we are not focusing on the intersection with B anymore but are looking at the overall likelihood of A happening, which includes every scenario where B might occur or not.

Example from Set Theory

Consider two events A is the event that it rains and B is the event that a football game is played.

The joint probability $P(A \cap B)$ is the probability that it rains and the football game is played.
The marginal probability $P(A)$ would be the total probability that it rains, regardless of whether the game happens or not. This would include both:

- The probability that it rains and the game is played.
- The probability that it rains and the game is not played.

In set notation, marginal probability can be seen as focusing on the event A, whether $A \cap B$ or $A \cap B^c$ (where B^c is the complement of B) occurs.

Marginal Probability Mass Function (Marginal PMF)

The marginal probability mass function of a discrete random variable is obtained by summing the joint PMF over all possible values of the other variable. The marginal PMF gives the probability distribution of one variable, disregarding the other. The marginal PMF of X, denoted as $P(X = x)$, is given by:

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

The marginal PMF of Y, denoted as $P(Y = y)$, is given by:

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

Examples on Marginal Distributions from Joint PMF

1. Using the joint distribution from Example 1 in the Joint Distribution of Two Random Variables section, we can compute the marginal distributions as follows:

   Solution

   $$P(X = 0) = \frac{8}{27}, \quad P(X = 1) = \frac{12}{27}, \quad P(X = 2) = \frac{6}{27}, \quad P(X = 3) = \frac{1}{27}$$
   $$P(Y = 1) = \frac{3}{27}, \quad P(Y = 2) = \frac{18}{27}, \quad P(Y = 3) = \frac{6}{27}$$

2. In the example of network packet loss, the second example in the Joint Distribution of Two Random Variables section, the joint distribution of packets reaching Servers A and B is:

$X \backslash Y$	0	1	2	Total
0	0.1	0.1	0.05	0.25
1	0.1	0.2	0.15	0.45
2	0.05	0.1	0.15	0.30

   What is the marginal distribution of packets reaching Server A?

   Solution
   The marginal distribution of packets reaching Server A, $P(X)$, is:

   $$P(X = 0) = 0.25, \quad P(X = 1) = 0.45, \quad P(X = 2) = 0.30$$

3. In the example of network packet loss, the joint distribution of packets reaching Servers A and B is:

$X \backslash Y$	0	1	2	Total
0	0.1	0.1	0.05	0.25
1	0.1	0.2	0.15	0.45
2	0.05	0.1	0.15	0.30

What is the marginal distribution of packets reaching Server A?

Solution
The marginal distribution of packets reaching Server A, $P(X)$, is:

$$P(X = 0) = 0.25, \quad P(X = 1) = 0.45, \quad P(X = 2) = 0.30$$

4. In the earlier example of customer purchases and returns, the joint PMF is given by:

$X \backslash Y$	0	1	2	Total
0	0.2	0.05	0.02	0.27
1	0.15	0.2	0.05	0.4
2	0.08	0.15	0.1	0.33

What is the marginal distribution of the number of items returned, Y?

Solution
We sum the joint probabilities over all values of X:

$$P(Y = 0) = 0.2 + 0.15 + 0.08 = 0.43$$
$$P(Y = 1) = 0.05 + 0.2 + 0.15 = 0.4$$
$$P(Y = 2) = 0.02 + 0.05 + 0.1 = 0.17$$

Thus, the marginal distribution of Y is $P(Y = 0) = 0.43$, $P(Y = 1) = 0.4$, and $P(Y = 2) = 0.17$.

5. Given the joint distribution of tasks handled by two processors:

$X \backslash Y$	0	1	2	Total
0	0.15	0.1	0.05	0.3
1	0.2	0.2	0.1	0.5
2	0.05	0.1	0.05	0.2

What is the marginal distribution of tasks handled by Processor B?

Solution
To find the marginal distribution of Y (tasks on Processor B):

$$P(Y = 0) = 0.15 + 0.2 + 0.05 = 0.4$$
$$P(Y = 1) = 0.1 + 0.2 + 0.1 = 0.4$$
$$P(Y = 2) = 0.05 + 0.1 + 0.05 = 0.2$$

Thus, the marginal distribution of Y is $P(Y = 0) = 0.4$, $P(Y = 1) = 0.4$, and $P(Y = 2) = 0.2$.

6. In a software testing process, the joint distribution of the number of bugs found by two teams, X and Y, is given as:

$X \backslash Y$	0	1	2	Total
0	0.3	0.15	0.05	0.5
1	0.1	0.15	0.05	0.3
2	0.05	0.1	0.05	0.2

What is the marginal distribution of the number of bugs found by Team A?

Solution
We sum the joint probabilities for each value of X:

$$P(X = 0) = 0.3 + 0.15 + 0.05 = 0.5$$
$$P(X = 1) = 0.1 + 0.15 + 0.05 = 0.3$$
$$P(X = 2) = 0.05 + 0.1 + 0.05 = 0.2$$

Thus, the marginal distribution of X is $P(X = 0) = 0.5$, $P(X = 1) = 0.3$, and $P(X = 2) = 0.2$.

Marginal Probability in Discrete Random Variables

When working with discrete random variables, computing marginal and conditional probabilities is essential. Here's how they are defined:

- **Marginal Probability**: The probability of an event occurring without considering other variables. For instance, $P(X = x)$ is the marginal probability of $X = x$, computed by summing over all possible values of the other variable Y.
- **Joint Probability**: The probability that two random variables X and Y take on specific values simultaneously, denoted $P(X = x, Y = y)$.
- **Conditional Probability**: The probability of one event occurring given that another has occurred, denoted $P(Y = y \mid X = x)$.

In the context of bivariate discrete distributions, these probabilities are key to understanding how two random variables interact.

Joint Probability Mass Function (PMF)

For two discrete random variables X and Y, the joint probability mass function (PMF) is defined as:

$$P_{XY}(x, y) = P(X = x, Y = y)$$

This represents the probability that X takes value x and Y takes value y simultaneously. The joint PMF contains all the information about the distribution of X and Y.

Joint Range R_{XY} is the set of all pairs (x, y) where $P_{XY}(x, y) > 0$:

$$R_{XY} = \{(x, y) \mid P_{XY}(x, y) > 0\}$$

The joint PMF can be written as:

$$P(X = x, Y = y) = P((X = x) \cap (Y = y))$$

Lemma 1: The sum of all probabilities in the joint PMF must equal 1:

$$\sum_{x \in R_X} \sum_{y \in R_Y} P_{XY}(x, y) = 1$$

Proof

Let (X, Y) be a pair of discrete random variables with a joint PMF $P_{XY}(x, y)$ defined over their respective ranges R_X and R_Y. The joint PMF $P_{XY}(x, y)$ assigns probabilities to every pair (x, y), ensuring that all possible outcomes are accounted for. By the definition of a probability measure, the total probability across the sample space must equal 1:

$$\sum_{x \in R_X} \sum_{y \in R_Y} P_{XY}(x, y) = 1$$

This is a consequence of the axioms of probability, specifically the normalization condition, which states that the probability of the entire sample space must sum to 1. For a finite number of outcomes, we can express:

$$n = |R_X| \text{ and } m = |R_Y| \text{ (the number of elements in } R_X \text{ and } R_Y).$$

Therefore, we can write:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P_{XY}(x_i, y_j) = 1$$

Lemma 2: The marginal PMF for X can be derived by summing over the joint PMF with respect to Y:

$$P_X(x) = \sum_{y \in R_Y} P_{XY}(x, y)$$

Proof

To derive the marginal PMF $P_X(x)$, we sum the joint PMF $P_{XY}(x, y)$ over all possible values of Y:

$$P_X(x) = \sum_{y \in R_Y} P_{XY}(x, y)$$

This signifies that the probability of X being equal to x is the aggregation of all probabilities corresponding to each possible y given x.

Mathematically, this is equivalent to the law of total probability:

$$P_X(x) = \sum_{y \in R_Y} P(Y \mid X = x) P_X(x) = P(X = x)$$

The total probability law states that we can calculate the probability of an event by conditioning on a partition of the sample space.

The marginalization process ensures that $P_X(x)$ captures all possible interactions with Y while isolating the effects of Y.

Thus, we express $P_X(x)$ as:

$$P_X(x) = \sum_{y \in R_Y} P_{XY}(x, y)$$

Therefore, Lemma 2 is validated, demonstrating how to derive marginal distributions from joint distributions.

Application of Lemmas in Theorems

Theorem: Marginal PMF for Y

Statement: The marginal PMF for Y can also be derived using a similar approach:

$$P_Y(y) = \sum_{x \in R_X} P_{XY}(x, y)$$

Proof

Using the same reasoning as in Lemma 2, we can derive the marginal PMF for Y by summing the joint PMF with respect to X:

$$P_Y(y) = \sum_{x \in R_X} P_{XY}(x, y)$$

This reflects the total probability of observing Y taking the value y, accounting for all possible values of X.
We can express it as:

$$P_Y(y) = \sum_{x \in R_X} P(X \mid Y = y)P_Y(y) = P(Y = y)$$

This application of the law of total probability allows us to conclude:

$$P_Y(y) = \sum_{x \in R_X} P_{XY}(x, y)$$

This relationship demonstrates the interdependence of joint and marginal distributions, reaffirming the validity of Lemmas 1 and 2.

Theorem: Marginal PMF for X

Statement: The marginal PMF for X can also be derived as follows:

$$P_X(x) = \sum_{y \in R_Y} P_{XY}(x, y)$$

Proof
By applying Lemma 2, we can assert:

$$P_X(x) = \sum_{y \in R_Y} P_{XY}(x, y)$$

This is equivalent to the aggregation of probabilities of X conditioned on all possible outcomes of Y.
Thus, it can also be expressed as:

$$P_X(x) = \sum_{y \in R_Y} P(Y \mid X = x)P_X(x)$$

This shows how marginal PMFs capture the essential probabilities of one variable, eliminating the influence of the other.The equation confirms that $P_X(x)$ is derived from summing the contributions from Y:

$$P_X(x) = \sum_{y \in R_Y} P_{XY}(x, y)$$

Thus, the relationship highlights the importance of joint distributions in understanding marginal behaviors.

Marginal Probability from Joint Probability

From the joint PMF, we can derive the marginal probability mass function for each variable.

- **Marginal PMF of X**:

$$P_X(x) = \sum_{y \in R_Y} P_{XY}(x, y)$$

This gives the probability that $X = x$, regardless of the value of Y.

- **Marginal PMF of Y**:

$$P_Y(y) = \sum_{x \in R_X} P_{XY}(x, y)$$

This gives the probability that $Y = y$, regardless of the value of X.

Example

1. Given the joint PMF of X and Y, find

$X \backslash Y$	0	1	2
0	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$
1	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$

(a) Find the marginal PMFs of X and Y.

Solution
The marginal PMF of X For $X = 0$

$$P_X(0) = \sum_y P_{XY}(0, y) = \frac{1}{6} + \frac{1}{4} + \frac{1}{8} = \frac{9}{24} = 0.375$$

For $X = 1$

$$P_X(1) = \sum_y P_{XY}(1, y) = \frac{1}{8} + \frac{1}{6} + \frac{1}{6} = \frac{15}{24} = 0.625$$

The marginal PMF of Y
For $Y = 0$

$$P_Y(0) = P_{XY}(0, 0) + P_{XY}(1, 0) = \frac{1}{6} + \frac{1}{8} = \frac{7}{24}$$

For $Y = 1$

$$P_Y(1) = P_{XY}(0, 1) + P_{XY}(1, 1) = \frac{1}{4} + \frac{1}{6} = \frac{10}{24}$$

For $Y = 2$

$$P_Y(2) = P_{XY}(0,2) + P_{XY}(1,2) = \frac{1}{8} + \frac{1}{6} = \frac{7}{24}$$

(b) Find the conditional probability $P(Y = 1 \mid X = 0)$

$$P(Y = 1 \mid X = 0) = \frac{P(X = 0, Y = 1)}{P_X(0)} = \frac{1/4}{0.375} = \frac{2}{3}$$

Conditional Probability from Joint PMF

The conditional probability of $Y = y$ given that $X = x$ can be computed from the joint PMF:

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P_X(x)}$$

This gives the probability that $Y = y$, provided that $X = x$. Similarly, the conditional probability $P(X = x \mid Y = y)$ can be computed using the same principles.

Examples on Conditional Probability from Joint PMF

1. In a machine learning model, two discrete random variables are observed where X represents the user age group in years, and Y represents the time spent on the website in hours. The joint PMF of the user age group X and the time spent on the website Y is given by the following table:

$X \backslash Y$	10	20	30
2 (Age Group: 20-29)	0.2	0.1	0.05
3 (Age Group: 30-39)	0.1	0.15	0.1

(a) Find the marginal PMF of X.

Solution
The marginal PMF of X is found by summing the joint PMF over all possible values of Y for each X.

For $X = 2$ (Age Group: 20-29)

$$P_X(2) = \sum_{y \in \{10,20,30\}} P_{XY}(2, y) = 0.2 + 0.1 + 0.05 = 0.35$$

For $X = 3$ (Age Group: 30-39)

$$P_X(3) = \sum_{y \in \{10,20,30\}} P_{XY}(3, y) = 0.1 + 0.15 + 0.1 = 0.35$$

Thus, the marginal PMF of X is:

$$P_X(x) = \begin{cases} 0.35, & \text{if } X = 2 \\ 0.35, & \text{if } X = 3 \end{cases}$$

(b) Find the marginal PMF of Y

Solution

To find the marginal PMF of Y from the joint PMF, we sum the joint probabilities over all possible values of X for each value of Y. This shows the probabilities associated with each value of Y
For $Y = 10$

$$P_Y(10) = \sum_{x\in\{2,3\}} P_{XY}(x, 10) = P_{XY}(2, 10)+P_{XY}(3, 10) = 0.2+0.1 = 0.3$$

For $Y = 20$

$$P_Y(20) = \sum_{x\in\{2,3\}} P_{XY}(x, 20) = P_{XY}(2, 20)+P_{XY}(3, 20) = 0.1+0.15 = 0.25$$

For $Y = 30$

$$P_Y(30) = \sum_{x\in\{2,3\}} P_{XY}(x, 30) = P_{XY}(2, 30)+P_{XY}(3, 30) = 0.05+0.1 = 0.15$$

Thus, the marginal PMF of Y is:

$$P_Y(y) = \begin{cases} 0.3, & \text{if } Y = 10 \\ 0.25, & \text{if } Y = 20 \\ 0.15, & \text{if } Y = 30 \end{cases}$$

(c) Find the conditional probability $P(Y = 20 \mid X = 3)$.
The conditional probability is given by:

$$P(Y = 20 \mid X = 3) = \frac{P(X = 3, Y = 20)}{P_X(3)}$$

From the joint PMF table, we know:

$$P(X = 3, Y = 20) = 0.15$$

and from the marginal PMF of X:

$$P_X(3) = 0.35$$

Thus, the conditional probability is:

$$P(Y = 20 \mid X = 3) = \frac{0.15}{0.35} \approx 0.4286$$

2. Let X represent the number of defects detected in a code module and Y the number of test cases executed. The joint PMF is given as:

$X\backslash Y$	10	20	30
0	0.1	0.2	0.15
1	0.05	0.25	0.25

Solution
To find the marginal PMF of X, sum across the columns for each value of X:

For $X = 0$

$$P_X(0) = 0.1 + 0.2 + 0.15 = 0.45$$

For $X = 1$

$$P_X(1) = 0.05 + 0.25 + 0.25 = 0.55$$

6 Marginal and Conditional Probabilities of Continuous Distributions

Bivariate/Two-Dimensional Continuous Random Variables

If X and Y are continuous random variables defined on the sample space Ω of a random experiment, then the pair (X, Y) is called a bivariate continuous random variable. This means that (X, Y) assigns a point in the xy-plane corresponding to each outcome in the sample space Ω.

Joint and Marginal Distribution and Density Functions

Two-Dimensional Continuous Distribution Function

The distribution function of a two-dimensional continuous random variable (X, Y) is a real-valued function and is defined as:

$$F(x, y) = P(X \leq x, Y \leq y) \quad \text{for all real } x \text{ and } y.$$

Remark: $F(x, y)$ can also be written as $F_{X,Y}(x, y)$.

Joint Probability Density Function

Let (X, Y) be a continuous random variable assuming all values in some region R of the xy-plane. Then, a function $f(x, y)$ such that:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x', y')\, dx'\, dy'$$

is defined to be a joint probability density function.

As in the one-dimensional case, a joint probability density function has the following properties:

- $f(x, y) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1$

Examples

1. Given the joint PDF $f_{X,Y}(x, y) = 6xy$ for $0 < x < 1$ and $0 < y < 1$, how do we find the value of the joint distribution function $F_{X,Y}(0.5, 0.5)$?

   Solution
   To compute $F_{X,Y}(0.5, 0.5)$, we integrate the joint PDF:
   $$F_{X,Y}(0.5, 0.5) = \int_0^{0.5} \int_0^{0.5} 6xy\, dx\, dy$$
   First, we integrate with respect to x:
   $$\int_0^{0.5} 6xy\, dx = 6y \left[\frac{x^2}{2}\right]_0^{0.5} = 6y \times \frac{0.25}{2} = 1.5y$$
   Now, we integrate with respect to y:
   $$\int_0^{0.5} 1.5y\, dy = 1.5 \left[\frac{y^2}{2}\right]_0^{0.5} = 1.5 \times \frac{0.25}{2} = 0.1875$$
   Thus, $F_{X,Y}(0.5, 0.5) = 0.1875$.

2. Suppose the response times X and Y of two servers are modeled with the joint PDF $f_{X,Y}(x, y) = 8xy$ for $0 < x < 1$ and $0 < y < 1$. How do we calculate the probability that both servers respond within 0.3 seconds?

   Solution
   We need to find:
   $$P(0 < X < 0.3, 0 < Y < 0.3) = \int_0^{0.3} \int_0^{0.3} 8xy\, dx\, dy$$
   First, we integrate with respect to x:
   $$\int_0^{0.3} 8xy\, dx = 8y \left[\frac{x^2}{2}\right]_0^{0.3} = 8y \times \frac{0.09}{2} = 0.36y$$
   Next, we integrate with respect to y:
   $$\int_0^{0.3} 0.36y\, dy = 0.36 \left[\frac{y^2}{2}\right]_0^{0.3} = 0.36 \times \frac{0.09}{2} = 0.0162$$
   Thus, the probability that both servers respond within 0.3 seconds is 0.0162.

Marginal Continuous Distribution Function

Let (X, Y) be a two-dimensional continuous random variable having $f(x, y)$ as its joint probability density function. Now, the marginal distribution function of the continuous random variable X is defined as:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

The marginal PDF of X is obtained by integrating the joint PDF over all values of Y.
The marginal distribution function of the continuous random variable Y is defined as:

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

The marginal PDF of Y is obtained by integrating the joint PDF over all values of X.

Marginal Probability Density Function

Let (X, Y) be a two-dimensional continuous random variable having $F(X, Y)$ and $f(X, Y)$ as its distribution function and joint probability density function, respectively. Let $F(x)$ and $F(y)$ be the marginal distribution functions of X and Y, respectively. Then, the marginal probability density function of X is given by:

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

or it may also be obtained as:

$$f_X(x) = \frac{d}{dx} F_X(x).$$

The marginal probability density function of Y is given by:

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

or

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

Examples

1. Given the joint PDF $f_{X,Y}(x, y) = 6xy$ for $0 < x < 1$ and $0 < y < 1$, how can we determine the marginal PDF $f_X(x)$?

Solution
We integrate the joint PDF over all values of y:

$$f_X(x) = \int_0^1 6xy\,dy = 6x\int_0^1 y\,dy = 6x\left[\frac{y^2}{2}\right]_0^1 = 6x \times \frac{1}{2} = 3x$$

Thus, the marginal PDF of X is $f_X(x) = 3x$ for $0 < x < 1$.

2. With the joint PDF $f_{X,Y}(x, y) = 6xy$, how do we calculate the marginal probability that X is less than 0.4?

Solution
We find $P(X < 0.4)$ by integrating the marginal PDF:

$$P(X < 0.4) = \int_0^{0.4} 3x\,dx = 3\left[\frac{x^2}{2}\right]_0^{0.4} = 3 \times \frac{0.16}{2} = 0.24$$

Thus, the probability that X is less than 0.4 is 0.24.

Conditional Probability Density Function

Let (X, Y) be a two-dimensional continuous random variable having the joint probability density function $f(x, y)$. The conditional probability density function of Y given $X = x$ is defined as:

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

where $f_X(x) > 0$ is the marginal density of X.

Similarly, the conditional probability density function of X given $Y = y$ is defined as:

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

where $f_Y(y) > 0$ is the marginal density of Y.

Conditional Continuous Distribution Function

For a two-dimensional continuous random variable (X, Y), the conditional distribution function of Y given $X = x$ is defined as:

$$F_{Y|X}(y|x) = P(Y \le y|X = x) = \int_{-\infty}^{y} f_{Y|X}(y'|x)\,dy'.$$

Similarly, the conditional distribution function of X given $Y = y$ is defined as:

$$F_{X|Y}(x|y) = P(X \le x|Y = y) = \int_{-\infty}^{x} f_{X|Y}(x'|y)\,dx'.$$

Joint PDF of Continuous Random Variables

For two continuous random variables X and Y, the **joint probability density function (PDF)** describes the likelihood that X takes a value in a small interval around x and Y takes a value in a small interval around y.

The joint PDF, denoted by $f_{X,Y}(x, y)$, is such that the probability that X is between a and b, and Y is between c and d is given by:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x, y)\, dx\, dy$$

The joint PDF must satisfy the following conditions:

1. $f_{X,Y}(x, y) \geq 0 \quad \text{for all } x \text{ and } y$ 2. The total probability must equal 1:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$$

Independence of Continuous Random Variables

Two continuous random variables X and Y are independent if the joint PDF can be factored as the product of their marginal PDFs:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \quad \text{for all } x \text{ and } y$$

Examples on Independence of Continuous Random Variables

1. Consider a system where the response times (in seconds) of two servers, X and Y, are jointly distributed with the PDF:

   $$f_{X,Y}(x, y) = \begin{cases} 6xy & \text{if } 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

   Find the marginal PDF of X, the response time of Server A.

   Solution
   To find the marginal PDF of X, integrate the joint PDF over all values of y:

   $$f_X(x) = \int_0^1 6xy\, dy = 6x \left[\frac{y^2}{2}\right]_0^1 = 3x, \quad 0 < x < 1$$

2. In a data processing pipeline, the time taken for processing X (in seconds) and the number of errors Y are jointly distributed according to the PDF:

   $$f_{X,Y}(x, y) = 2e^{-x}(1 - e^{-y}) \quad x > 0, y > 0$$

   Find the marginal PDF of the processing time X.

Solution
The marginal PDF of X is found by integrating the joint PDF over all values of y:

$$f_X(x) = \int_0^\infty 2e^{-x}(1 - e^{-y})\,dy = 2e^{-x}\int_0^\infty (1 - e^{-y})\,dy$$

$$= 2e^{-x}\left[y + \frac{1}{e^y}\right]_0^\infty = 2e^{-x}$$

3. Two processors in a distributed computing system are allocated tasks. Let X be the time (in seconds) it takes for Processor A to finish its tasks, and Y be the time for Processor B. The joint PDF is given by:

$$f_{X,Y}(x, y) = e^{-(x+y)}, \quad x > 0, y > 0$$

Check if the task completion times X and Y are independent.

Solution
First, find the marginal PDFs:

$$f_X(x) = \int_0^\infty e^{-(x+y)}\,dy = e^{-x}$$

$$f_Y(y) = \int_0^\infty e^{-(x+y)}\,dx = e^{-y}$$

Since $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$, the variables X and Y are independent.

4. In a software testing scenario, let X represent the number of bugs found in module A, and Y represent the number of bugs found in module B. Suppose the joint PDF of X and Y is:

$$f_{X,Y}(x, y) = \begin{cases} 10xy & \text{if } 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the probability that both modules have bug counts less than 0.5?

Solution
We need to compute:

$$P(X < 0.5, Y < 0.5) = \int_0^{0.5}\int_0^{0.5} 10xy\,dx\,dy$$

First, integrate with respect to x:

$$\int_0^{0.5} 10xy\,dx = 10y\left[\frac{x^2}{2}\right]_0^{0.5} = 10y \times 0.125 = 1.25y$$

Now, integrate with respect to y:

$$\int_0^{0.5} 1.25y\,dy = 1.25\left[\frac{y^2}{2}\right]_0^{0.5} = 1.25 \times 0.125 = 0.15625$$

Thus, the probability that both modules have bug counts less than 0.5 is 0.15625.

More Examples

1. Consider two continuous random variables, X and Y, which have a joint PDF defined as $f_{X,Y}(x, y) = 6xy$ for values $0 < x < 1$ and $0 < y < 1$. How can we verify that this joint PDF is valid by ensuring that the total probability integrates to 1 over the specified range?

   Solution
   To confirm that this is a valid joint PDF, we need to compute the total probability:
   $$\int_0^1 \int_0^1 6xy\,dx\,dy = 6\int_0^1 \int_0^1 xy\,dx\,dy$$
   First, we perform the integration with respect to x:
   $$\int_0^1 xy\,dx = \frac{y}{2}\bigg|_0^1 = \frac{1}{2}y$$
   Next, we integrate with respect to y:
   $$\int_0^1 \frac{1}{2}y\,dy = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$
   Thus, the total integral is:
   $$6 \times \frac{1}{4} = 1$$
   This confirms that $f_{X,Y}(x, y)$ is indeed a valid joint PDF.

2. Given the joint PDF $f_{X,Y}(x, y) = 6xy$ for $0 < x < 1$ and $0 < y < 1$, how can we calculate the probability that both random variables X and Y fall within the range of 0.2 to 0.5?

   Solution
   We need to find:
   $$P(0.2 < X < 0.5, 0.2 < Y < 0.5) = \int_{0.2}^{0.5} \int_{0.2}^{0.5} 6xy\,dx\,dy$$
   First, we integrate with respect to x:
   $$\int_{0.2}^{0.5} 6xy\,dx = 6y\left[\frac{x^2}{2}\right]_{0.2}^{0.5} = 6y\left(\frac{0.25}{2} - \frac{0.04}{2}\right) = 3y \times 0.21 = 0.63y$$

Next, we integrate with respect to y:

$$\int_{0.2}^{0.5} 0.63y\, dy = 0.63 \left[\frac{y^2}{2}\right]_{0.2}^{0.5} = 0.63 \times \frac{0.25 - 0.04}{2} = 0.63 \times 0.105 = 0.06615$$

Thus, the probability that both X and Y are in the specified range is 0.06615.

3. Given the joint PDF $f_{X,Y}(x, y) = 6xy$ for $0 < x < 1$ and $0 < y < 1$, how do we calculate the conditional PDF $f_{X|Y}(x|0.5)$?

   Solution
   First, find $f_Y(0.5)$:

   $$f_Y(y) = \int_0^1 f_{X,Y}(x, y)\, dx = \int_0^1 6xy\, dx = 3y \quad \Rightarrow \quad f_Y(0.5) = 3 \times 0.5 = 1.5$$

   Now, calculate $f_{X|Y}(x|0.5)$:

   $$f_{X|Y}(x|0.5) = \frac{f_{X,Y}(x, 0.5)}{f_Y(0.5)} = \frac{6x(0.5)}{1.5} = 2x \quad \text{for } 0 < x < 1$$

4. If the joint PDF for two software system response times X and Y is given by $f_{X,Y}(x, y) = 8xy$ for $0 < x < 1$ and $0 < y < 1$, how can we calculate the conditional PDF $f_{X|Y}(x|0.2)$?

   Solution
   1. Find $f_Y(0.2)$:

   $$f_Y(y) = \int_0^1 8xy\, dx = 4y \quad \Rightarrow \quad f_Y(0.2) = 4 \times 0.2 = 0.8$$

   2. Now compute $f_{X|Y}(x|0.2)$:

   $$f_{X|Y}(x|0.2) = \frac{f_{X,Y}(x, 0.2)}{f_Y(0.2)} = \frac{8x(0.2)}{0.8} = 2x \quad \text{for } 0 < x < 1$$

5. For the joint PDF $f_{X,Y}(x, y) = 10xy$ where $0 < x < 1$ and $0 < y < 1$, how do we find $P(X < 0.4|Y = 0.6)$?

   Solution
   Calculate the conditional PDF $f_{X|Y}(x|0.6)$: First, find $f_Y(0.6)$:

   $$f_Y(y) = \int_0^1 10xy\, dx = 5y \quad \Rightarrow \quad f_Y(0.6) = 5 \times 0.6 = 3$$

   Now find $f_{X|Y}(x|0.6)$:

   $$f_{X|Y}(x|0.6) = \frac{f_{X,Y}(x, 0.6)}{f_Y(0.6)} = \frac{10x(0.6)}{3} = \frac{20x}{3} \quad \text{for } 0 < x < 1$$

Compute $P(X < 0.4|Y = 0.6)$:

$$P(X < 0.4|Y = 0.6) = \int_0^{0.4} \frac{20x}{3}\,dx = \frac{20}{3}\left[\frac{x^2}{2}\right]_0^{0.4} = \frac{20}{3}\times\frac{0.16}{2} = \frac{20 \times 0.08}{3} = \frac{1.6}{3} \approx 0.5333$$

6. If the joint PDF $f_{X,Y}(x, y) = 4xy$ for $0 < x < 1$ and $0 < y < 1$ is given, how do we find $f_{X|Y}(x|0.8)$?

   Solution
   First, find $f_Y(0.8)$:

$$f_Y(y) = \int_0^1 4xy\,dx = 2y \quad \Rightarrow \quad f_Y(0.8) = 2 \times 0.8 = 1.6$$

   Now calculate $f_{X|Y}(x|0.8)$:

$$f_{X|Y}(x|0.8) = \frac{f_{X,Y}(x, 0.8)}{f_Y(0.8)} = \frac{4x(0.8)}{1.6} = 2x \quad \text{for } 0 < x < 1$$

7. In a data analysis project, the joint PDF of the heights X and weights Y of individuals is given by $f_{X,Y}(x, y) = 12xy$ for $0 < x < 1$ and $0 < y < 1$. How do we find the probability that both height and weight are less than 0.5?

   Solution
   We need to calculate:

$$P(X < 0.5, Y < 0.5) = \int_0^{0.5}\int_0^{0.5} 12xy\,dx\,dy$$

   First, integrate with respect to x:

$$\int_0^{0.5} 12xy\,dx = 12y\left[\frac{x^2}{2}\right]_0^{0.5} = 12y \times \frac{0.25}{2} = 1.5y$$

   Next, integrate with respect to y:

$$\int_0^{0.5} 1.5y\,dy = 1.5\left[\frac{y^2}{2}\right]_0^{0.5} = 1.5 \times \frac{0.25}{2} = 0.1875$$

   Thus, the probability is 0.1875.

8. In a software application, the joint PDF of response time X and memory usage Y is defined as $f_{X,Y}(x, y) = 16xy$ for $0 < x < 1$ and $0 < y < 1$. How do we find the probability that the response time is less than 0.4 and memory usage is less than 0.3?

Solution
Calculate:

$$P(X < 0.4, Y < 0.3) = \int_0^{0.3} \int_0^{0.4} 16xy \, dx \, dy$$

First, integrate with respect to x:

$$\int_0^{0.4} 16xy \, dx = 16y \left[\frac{x^2}{2}\right]_0^{0.4} = 16y \times \frac{0.16}{2} = 1.28y$$

Next, integrate with respect to y:

$$\int_0^{0.3} 1.28y \, dy = 1.28 \left[\frac{y^2}{2}\right]_0^{0.3} = 1.28 \times \frac{0.09}{2} = 0.0576$$

Thus, the probability is 0.0576.

9. In a machine learning model, suppose the joint PDF for features X and Y is given by $f_{X,Y}(x,y) = 20xy$ for $0 < x < 1$ and $0 < y < 1$. How do we determine the probability that both features are below 0.2?

   Solution
   We calculate:

   $$P(X < 0.2, Y < 0.2) = \int_0^{0.2} \int_0^{0.2} 20xy \, dx \, dy$$

   Integrate with respect to x:

   $$\int_0^{0.2} 20xy \, dx = 20y \left[\frac{x^2}{2}\right]_0^{0.2} = 20y \times \frac{0.04}{2} = 0.4y$$

   Next, integrate with respect to y:

   $$\int_0^{0.2} 0.4y \, dy = 0.4 \left[\frac{y^2}{2}\right]_0^{0.2} = 0.4 \times \frac{0.04}{2} = 0.008$$

   Thus, the probability is 0.008.

10. In an experimental study, the joint PDF of two variables X (temperature) and Y (pressure) is defined as $f_{X,Y}(x,y) = 14xy$ for $0 < x < 1$ and $0 < y < 1$. How do we compute the probability that temperature is less than 0.6 and pressure is less than 0.4?

    Solution
    We find:

    $$P(X < 0.6, Y < 0.4) = \int_0^{0.4} \int_0^{0.6} 14xy \, dx \, dy$$

First, integrate with respect to x:

$$\int_0^{0.6} 14xy\,dx = 14y\left[\frac{x^2}{2}\right]_0^{0.6} = 14y \times \frac{0.36}{2} = 2.52y$$

Next, integrate with respect to y:

$$\int_0^{0.4} 2.52y\,dy = 2.52\left[\frac{y^2}{2}\right]_0^{0.4} = 2.52 \times \frac{0.16}{2} = 0.2016$$

Thus, the probability is 0.2016.

11. In a software performance test, the joint PDF of CPU usage X and memory usage Y is given as $f_{X,Y}(x,y) = 18xy$ for $0 < x < 1$ and $0 < y < 1$. How do we find the probability that CPU usage is less than 0.5 and memory usage is less than 0.7?

    Solution
    We need to compute:

$$P(X < 0.5, Y < 0.7) = \int_0^{0.7}\int_0^{0.5} 18xy\,dx\,dy$$

First, integrate with respect to x:

$$\int_0^{0.5} 18xy\,dx = 18y\left[\frac{x^2}{2}\right]_0^{0.5} = 18y \times \frac{0.25}{2} = 2.25y$$

Next, integrate with respect to y:

$$\int_0^{0.7} 2.25y\,dy = 2.25\left[\frac{y^2}{2}\right]_0^{0.7} = 2.25 \times \frac{0.49}{2} = 0.55125$$

Thus, the probability is 0.55125.

Exercise

1. Let X and Y be two random variables. Then for

$$f_{X,Y}(x,y) = \begin{cases} kxy & \text{for } 0 < x < 4 \text{ and } 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

to be a joint density function, what must be the value of k? As $f_{X,Y}(x,y)$ is the joint probability density function,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx = 1$$

$$\int_0^4 \int_1^5 kxy\,dy\,dx = 1$$
$$\int_0^4 \int_1^5 kxy\,dy\,dx = 1$$
$$\int_1^5 y\,dy \int_0^4 kx\,dx = 1$$
$$\int_1^5 y\,dy = \frac{5^2}{2} - \frac{1^2}{2} = \frac{25-1}{2} = \frac{24}{2} = 12$$
$$k\int_0^4 x\,dx = \frac{4^2}{2} = 8$$

So,

$$12k \cdot 8 = 1 \Rightarrow 96k = 1 \Rightarrow k = \frac{1}{96}.$$

2. If the joint p.d.f. of a two-dimensional random variable (X, Y) is given by

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{for } 0 < x < 1 \text{ and } 0 < y < x \\ 0 & \text{otherwise} \end{cases}$$

Then,

(a) Find the marginal density functions of X and Y.

Solution
Marginal density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^2 \,dx$$

As x is involved in both the given ranges, i.e. $0 < x < 1$ and $0 < y < x$; therefore, here we will combine both these intervals and hence have $0 < y < x < 1$. $\Rightarrow x$ takes the values from y to 1]

$$[2x]_y^1$$
$$= 2 - 2y = 2(1-y), \quad 0 < y < 1$$

Marginal density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$
$$\int_0^x 2dy \quad [0 < y < x < 1]$$
$$= 2\,[y]_0^x$$
$$= 2x, 0 < x < 1$$

(b) Find the conditional density functions.

Solution
Conditional density function of Y given X $(0 < X < 1)$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2}{2x} \quad \text{for } 0 < y < x.$$

Conditional density function of X given Y $(0 < Y < 1)$ is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2}{2(1-y)} = \frac{1}{1-y} \quad \text{for } y < x < 1.$$

(c) Check for independence of X and Y.

Solution $f_{X,Y}(x,y) = 2$,

$$f_X(x) f_Y(y) = 2(2x)(1-y) \neq f_{X,Y}(x,y).$$

$\Rightarrow X$ and Y are not independent.

3. If (X, Y) be a two-dimensional random variable having joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{8}(6-x-y) & \text{for } 0 < x < 2,\, 2 < y < 4 \\ 0 & \text{elsewhere} \end{cases}$$

Find

(a) $P(X < 1, Y < 3)$
Solution

$$\begin{aligned}
P(X < 1, Y < 3) &= \int_{-\infty}^{1} \int_{-\infty}^{3} f_{X,Y}(x,y)\, dy\, dx \\
&= \int_0^1 \int_2^3 \frac{1}{8}(6-x-y)\, dy\, dx \\
&= \int_0^1 \left(\int_2^3 \frac{6-x-y}{8}\, dy \right) dx \\
&= \int_0^1 \left(\frac{6y - xy - \frac{y^2}{2}}{8} \Bigg|_2^3 \right) dx \\
&= \int_0^1 \left(\frac{(18 - 3x - \frac{9}{2}) - (12 - 2x - 2}{8} \right) dx \\
&= \int_0^1 \left(\frac{(18 - 3x - \frac{9}{2}) - (10 - 2x)}{8} \right) dx
\end{aligned}$$

$$= \int_0^1 \left(\frac{\frac{7}{2} - x)}{8} \right) dx$$

$$= \left(\left(\frac{\frac{7}{2}x - \frac{x^2}{2}}{8} \right) \Big|_0^1 \right) = \frac{3}{8} = 0.375$$

(b) $P(X < 1|Y < 3)$.

Solution

$$P\left[X < 1|Y < 3\right] = \frac{P[X < 1, Y, 3}{P[Y < 3]}$$

Where $Y < 3$

$$= \int_0^2 \int_2^3 \frac{1}{8}(6 - x - y)\, dy\, dx$$

$$= \int_0^2 \left(\int_2^3 \frac{6 - x - y}{8}\, dy \right) dx$$

$$= \int_0^2 \left(\frac{6y - xy - \frac{y^2}{2}}{8} \Big|_2^3 \right) dx$$

$$= \int_0^2 \left(\frac{(18 - 3x - \frac{9}{2}) - (12 - 2x - 2}{8} \right) dx$$

$$= \int_0^2 \left(\frac{(18 - 3x - \frac{9}{2}) - (10 - 2x)}{8} \right) dx$$

$$= \int_0^2 \left(\frac{\frac{7}{2} - x)}{8} \right) dx$$

$$= \left(\left(\frac{\frac{7}{2}x - \frac{x^2}{2}}{8} \right) \Big|_0^2 \right) = \frac{5}{8}$$

where

$$P(Y < 3) = \int_0^2$$

$$\therefore P\left[X < 1|Y < 3\right] = \frac{\frac{3}{8}}{\frac{5}{8}} = \frac{3}{5} = 0.6$$

7 Covariance and Correlation

For this topic, refer to the link. When you see solution, click on it to display the workings. https://www.probabilitycourse.com/chapter5/5_3_1_covariance_correlation.php

8 Conditional Expectation and Variance.

For this topic, refer to the link. When you see solution, click on it to display the workings. https://www.probabilitycourse.com/chapter5/5_1_5_conditional_expectation.php

9 Moments and moment generating functions.

For this topic, refer to the link.
Click here for moments.https://www.statlect.com/fundamentals-of-probability/moments
Click here for moment generating functions. https://www.statlect.com/fundamentals-of-probability/moment-generating-function
Click here for characteristic functions. https://www.statlect.com/fundamentals-of-probability/characteristic-function

10 Probability Distributions and their linkages

At this point we have learnt about univariate discrete distribution, univariate continuous distributions, bivariate discrete distributions and bivariate continuous distributions. These probability distributions are connected/linked to other probability distributions under certain conditions. https://www.statlect.com/probability-distributions/relationships-among-probability-distributions

Once you have read the above find more worked examples below.

1. A subscription service has 100 customers, and each has a 10% chance of leaving. What is the probability that exactly 5 will leave in a given month?

   Solution
   Using the Binomial PMF with $n = 100$, $p = 0.1$, and $k = 5$:

   $$P(X = 5) = \binom{100}{5}(0.1)^5(0.9)^{95} \approx 0.1871$$

2. A web server receives an average of 3 requests per second. What is the probability of receiving exactly 5 requests in a second?

   Solution
   Using the Poisson PMF with $\lambda = 3$ and $k = 5$:

   $$P(X = 5) = \frac{3^5 e^{-3}}{5!} \approx 0.1008$$

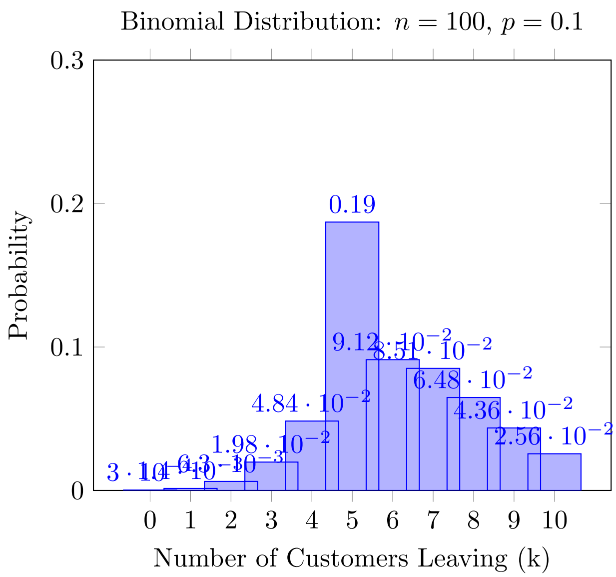

Figure 2: Binomial Distribution PMF for $n = 100$, $p = 0.1$

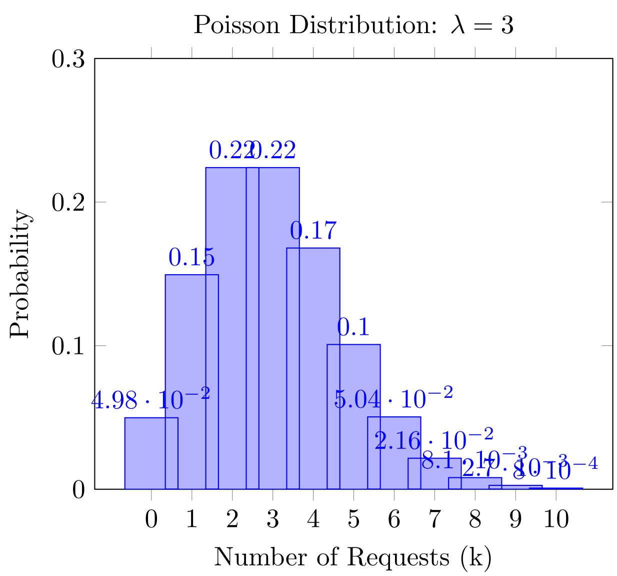

Figure 3: Poisson Distribution PMF for $\lambda = 3$

3. A data scientist is analyzing customer churn in a subscription service. The service has 100 customers, and each customer has a 10% chance of leaving the service in any given month. What is the probability that exactly 5

customers will leave in a given month?

Solution
This is a classic example of a Binomial distribution where $n = 100$, $p = 0.1$, and we are interested in finding the probability of $k = 5$ customers leaving.

Using the Binomial PMF:

$$P(X = 5) = \binom{100}{5} (0.1)^5 (0.9)^{95}$$

Calculating the binomial coefficient:

$$\binom{100}{5} = \frac{100!}{5!(100-5)!} = 75287520$$

Now, substituting the values:

$$P(X = 5) = 75287520 \times (0.1)^5 \times (0.9)^{95} \approx 0.1871$$

Thus, the probability that exactly 5 customers will leave in the given month is approximately 0.1871.

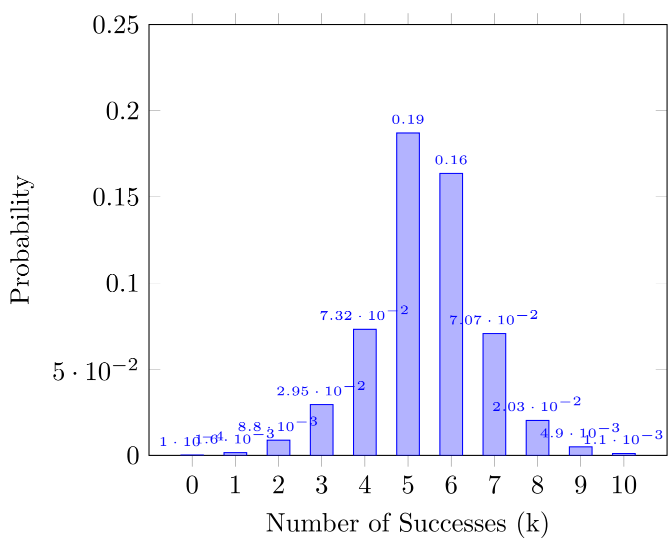

Figure 4: Binomial Distribution for $n = 100$, $p = 0.1$

4. A web server experiences an average of 3 requests per second. What is the probability that the server will receive exactly 5 requests in a second?

   Solution
   Here, $\lambda = 3$ and we are interested in finding $P(X = 5)$.

   Using the Poisson PMF:

$$P(X = 5) = \frac{3^5 e^{-3}}{5!}$$

   Calculating the factorial:

$$5! = 120$$

   Substituting the values:

$$P(X = 5) = \frac{243e^{-3}}{120} \approx 0.1008$$

   Thus, the probability that the server will receive exactly 5 requests in a second is approximately 0.1008.

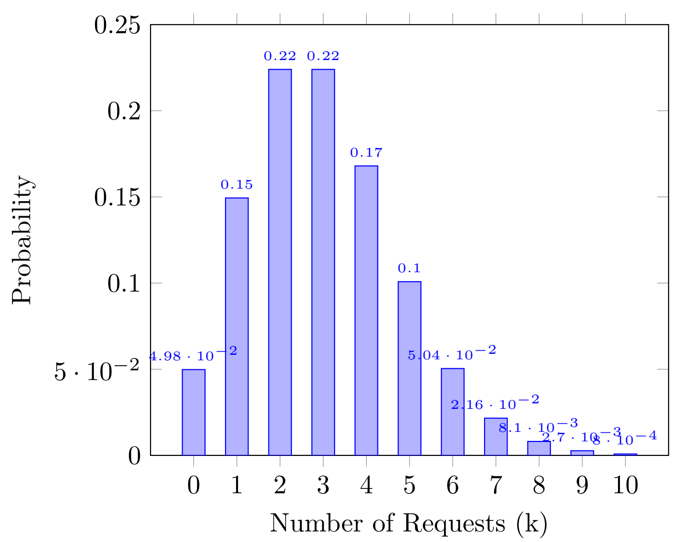

Figure 5: Poisson Distribution for $\lambda = 3$

5. What is the probability that exactly 3 packets out of 5 transmitted are successful if the probability of success in each transmission is 0.7?

   Solution
   We use the Binomial Probability Mass Function (PMF):

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Here:

$$n = 5, \quad k = 3, \quad p = 0.7$$

Substituting values:

$$P(X = 3) = \binom{5}{3} (0.7)^3 (0.3)^2$$

Calculating:

$$P(X = 3) = \frac{5!}{3!(5-3)!} (0.7)^3 (0.3)^2 = 10 \times 0.343 \times 0.09 \approx 0.3087$$

Thus, the probability of exactly 3 successful transmissions is 0.3087.

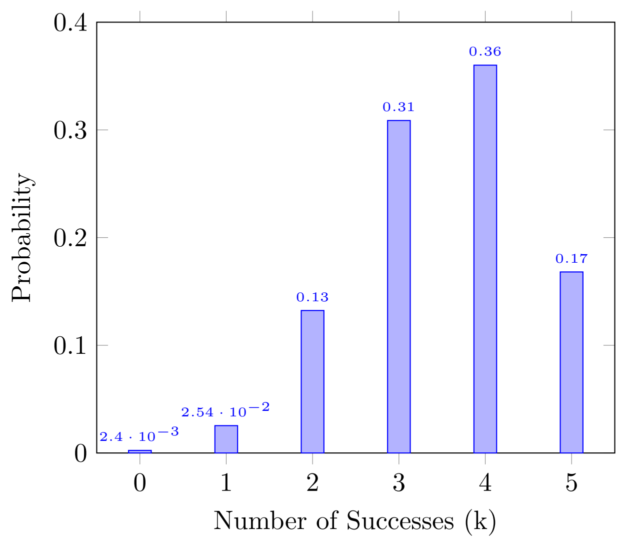

Figure 6: Binomial Distribution for $n = 5$, $p = 0.7$

6. In a software testing scenario, defects occur with an average rate of 2 defects per hour. What is the probability that exactly 5 defects will be found in an hour?

   Solution
   We use the Poisson PMF:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Here:

$$\lambda = 2, \quad k = 5$$

Substituting values:

$$P(X = 5) = \frac{2^5 e^{-2}}{5!} = \frac{32e^{-2}}{120} \approx 0.0361$$

Thus, the probability of finding exactly 5 defects in an hour is 0.0361.

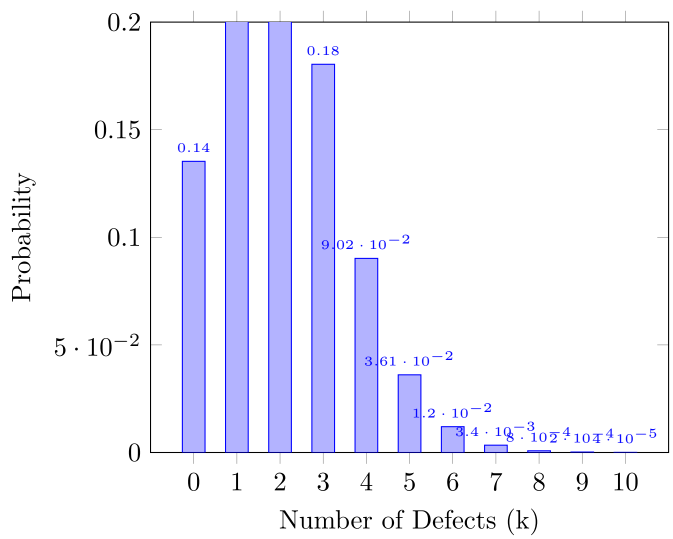

Figure 7: Poisson Distribution for $\lambda = 2$

7. Suppose a batch of 20 electronic components contains 5 defective ones. If 4 components are selected at random, what is the probability that exactly 2 of them are defective?

   Solution
   We use the Hypergeometric PMF:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

   Here:

$$N = 20, \quad K = 5, \quad n = 4, \quad k = 2$$

Substituting values:

$$P(X = 2) = \frac{\binom{5}{2}\binom{15}{2}}{\binom{20}{4}} = \frac{10 \times 105}{4845} \approx 0.2166$$

Thus, the probability of selecting exactly 2 defective components is 0.2166.

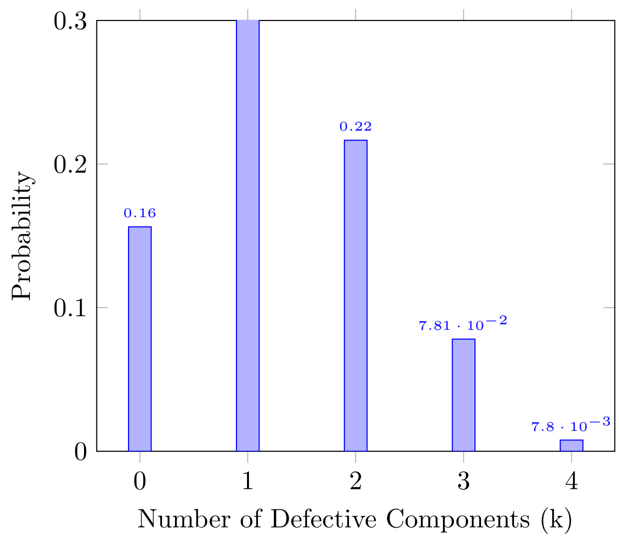

Figure 8: Hypergeometric Distribution for $N = 20$, $K = 5$, $n = 4$

8. In a call center, the average time between customer calls is 10 minutes. What is the probability that the next call will come within 5 minutes?

   Solution
   We use the Exponential PDF:

$$P(T \leq t) = 1 - e^{-\lambda t}$$

   Here:

$$\lambda = \frac{1}{10}, \quad t = 5$$

   Substituting values:

$$P(T \leq 5) = 1 - e^{-\frac{1}{10} \times 5} = 1 - e^{-0.5} \approx 1 - 0.6065 = 0.3935$$

Thus, the probability that the next call will come within 5 minutes is 0.3935.

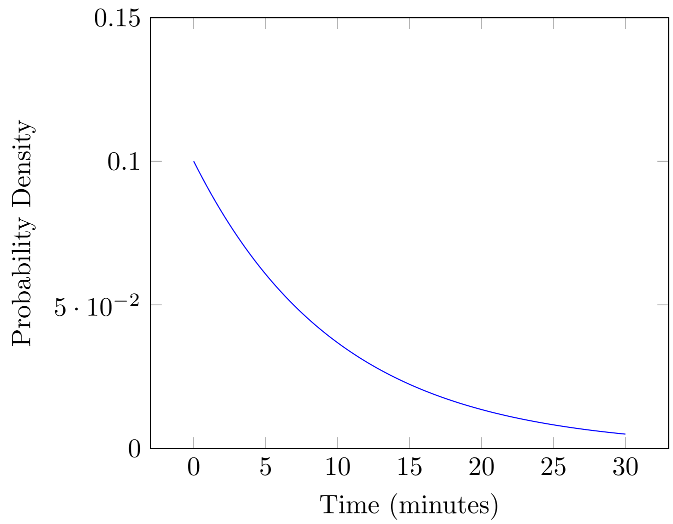

Figure 9: Exponential Distribution for $\lambda = 0.1$

9. If the scores on a standardized test are normally distributed with a mean of 100 and a standard deviation of 15, what is the probability that a randomly selected score is between 85 and 115?

   Solution
   We use the Normal Distribution CDF:

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

   Where Z is the standard normal variable:

$$\mu = 100, \quad \sigma = 15, \quad a = 85, \quad b = 115$$

   First, compute the Z-scores:

$$Z_1 = \frac{85-100}{15} = -1, \quad Z_2 = \frac{115-100}{15} = 1$$

   Using the standard normal table, we find:

$$P(-1 < Z < 1) \approx 0.6826$$

   Thus, the probability that the score is between 85 and 115 is 0.6826.

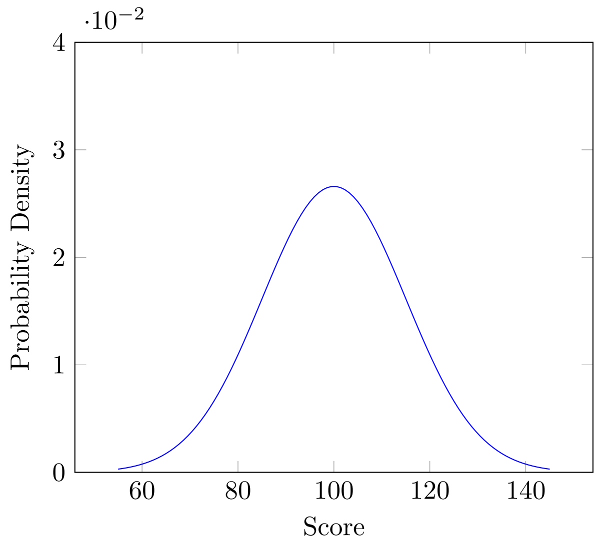

Figure 10: Normal Distribution with $\mu = 100$, $\sigma = 15$

10. In a machine learning model, the prior distribution of a parameter θ is Beta-distributed with $\alpha = 2$ and $\beta = 3$. What is the expected value of θ?

    Solution
    The expected value of a Beta distribution is:

    $$E(\theta) = \frac{\alpha}{\alpha + \beta}$$

    Substituting values:

    $$E(\theta) = \frac{2}{2+3} = \frac{2}{5} = 0.4$$

    Thus, the expected value of θ is 0.4.

11. The time until a server experiences two failures follows a Gamma distribution with shape parameter $\alpha = 2$ and rate parameter $\lambda = \frac{1}{3}$. What is the probability that the server will experience the second failure within 6 hours?

    Solution
    We use the Gamma CDF for $\alpha = 2$:

    $$P(T \leq t) = 1 - e^{-\lambda t}\,(1 + \lambda t)$$

    Here:

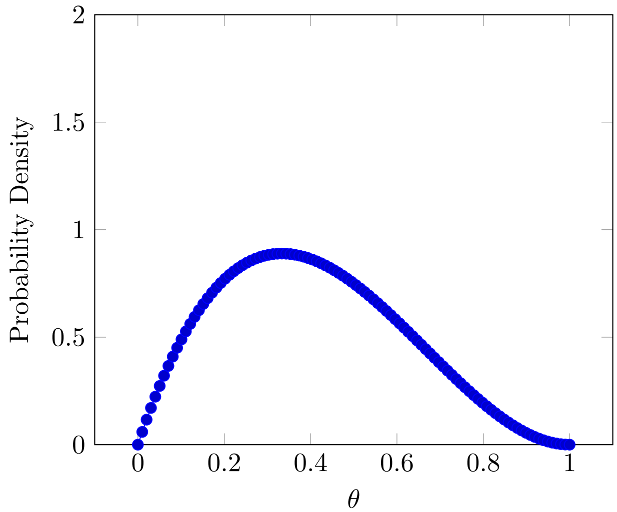

Figure 11: Beta Distribution with $\alpha = 2$, $\beta = 3$

$$\alpha = 2, \quad \lambda = \frac{1}{3}, \quad t = 6$$

Substituting values:

$$P(T \leq 6) = 1 - e^{-\frac{1}{3}\times 6}\left(1 + \frac{1}{3} \times 6\right) = 1 - e^{-2} \times 3 \approx 0.5941$$

Thus, the probability that the server will experience the second failure within 6 hours is 0.5941.

When we have some data and we want to categorize a random variable, we can use the chart (Figure 61.15:Distributional Choices) named in the link as a guide https://tinyheero.github.io/2016/03/17/prob-distr.html

11 Bivariate Distributions

For this topic, refer to the link. When you see solution, click on it to display the workings. https://www.probabilitycourse.com/chapter5/5_3_2_bivariate_normal_dist.php

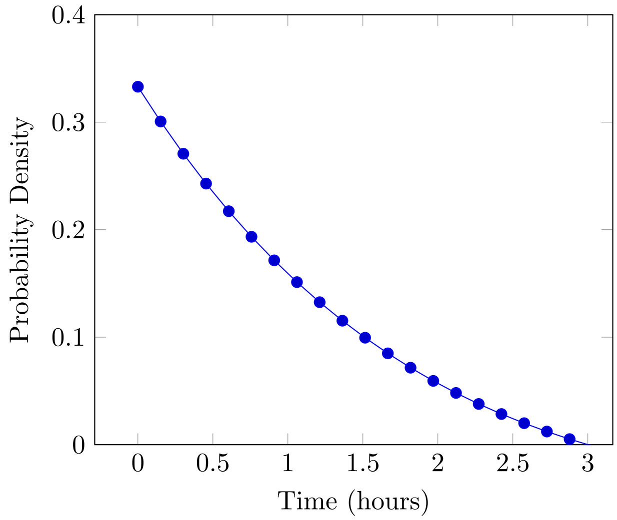

Figure 12: Gamma Distribution with $\alpha = 2$, $\lambda = \frac{1}{3}$

Sample Examination Paper 1
Total Marks: 70

Section A: Short Questions (1 Mark Each)

1. Define a discrete random variable.
2. What is a continuous random variable?
3. State one property of the Probability Mass Function (PMF).
4. What is the Law of Large Numbers (LLN)?
5. Define covariance.

Section B: Medium Questions (2 Marks Each)

1. Differentiate between joint probability and marginal probability.
2. Explain the significance of a joint probability mass function (PMF).
3. If the probability that it will rain tomorrow given that it is cloudy today is 0.3, express this using conditional probability notation.
4. Give an example of a bivariate discrete probability distribution.
5. For a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$, compute the value of $f(0)$.

Section C: Practical Applications (3 Marks Each)

1. In a software testing process, if Suite A fails with a probability of 0.3 and Suite B with 0.4, what is the joint probability of both suites failing if they are independent?
2. For a fair six-sided die, calculate the expected value of the outcome.
3. A data center records 2 failures per hour. What is the probability of recording exactly 3 failures in an hour? Use the Poisson distribution formula.
4. In a bivariate distribution, if the probability of purchasing an item and returning it is 0.1 and the probability of returning the item is 0.2, find the conditional probability that the item was purchased given that it was returned.

Section D: Problem Solving (4 Marks Each)

1. A machine produces 10,000 widgets in a day. The probability that a widget is defective is 0.01. What is the probability that exactly 2 widgets will be defective on a given day? Use the binomial distribution.
2. For a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that a randomly selected value will fall between 40 and 60.
3. In an IT network, the time taken to process a request is exponentially distributed with a mean of 2 seconds. What is the probability that a request will take less than 1 second to process?

Section E: Long Answer (20 Marks)

1. (a) Define and explain the joint probability distribution for bivariate discrete random variables. (5 marks)
   (b) Using an example from Data Science, explain how to compute the marginal and conditional probabilities from a joint probability mass function (PMF). (5 marks)
   (c) Solve the following problem: A web server receives an average of 3 requests per second. What is the probability of receiving exactly 5 requests in a second? (Use the Poisson distribution formula) (5 marks)
   (d) Explain the relationship between the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT), and how they apply to statistics. (5 marks)

Sample Examination Paper 2
Total Marks: 70

Section A: Short Questions (1 Mark Each)

1. Define the Central Limit Theorem (CLT).
2. What is the expected value of a random variable?
3. State one property of the cumulative distribution function (CDF).
4. Define a bivariate random variable.
5. What is a probability mass function (PMF)?

Section B: Medium Questions (2 Marks Each)

1. Differentiate between discrete and continuous random variables.
2. Explain the concept of conditional probability with an example.
3. In a binomial distribution, what happens as the number of trials n increases?
4. Write down the general form of the Poisson distribution formula.
5. What is a covariance matrix?

Section C: Practical Applications (3 Marks Each)

1. In a web server, the time between requests is exponentially distributed with a mean of 3 seconds. What is the probability that the next request will arrive in less than 2 seconds?
2. For a discrete random variable with probabilities $P(X = 1) = 0.2, P(X = 2) = 0.5, P(X = 3) = 0.3$, compute the expected value of X.
3. A server receives an average of 4 requests per second. What is the probability of receiving exactly 6 requests in a second? Use the Poisson distribution formula.
4. In a bivariate probability distribution, the joint probability of two variables X and Y is given. How do you compute the marginal distribution of X?

Section D: Problem Solving (4 Marks Each)

1. For a binomial distribution with $n = 10$ and $p = 0.6$, find the probability of getting exactly 7 successes.

2. In a data science project, the number of bugs found in two different code modules follows a bivariate distribution. How would you compute the conditional probability of bugs in one module given the number of bugs in the other?

3. The time taken to process a task in a system is normally distributed with a mean of 15 seconds and a standard deviation of 5 seconds. What is the probability that a task will take between 10 and 20 seconds to complete?

Section E: Long Answer (20 Marks)

1. (a) Define and explain the concept of conditional probability for bivariate distributions. (5 marks)

   (b) Explain how the Central Limit Theorem applies to sampling distributions, with an example from IT or Data Science. (5 marks)

   (c) Solve the following: A call center receives an average of 5 calls per minute. What is the probability of receiving exactly 8 calls in the next minute? (Use the Poisson distribution formula). (5 marks)

   (d) Explain how the Exponential distribution is related to the Poisson distribution, and give a real-world application of this in computer networks. (5 marks)