# MATHEMATICAL STATISTICS
## Lecture Note

**Akram Al-sabbagh**
Al-Nahrain University

# CONTENTS

# CHAPTER 1

# INTRODUCTION TO STATISTICS

## 1.1 Basic Probability

This chapter is a reminder of some basics in probability and statistics, it contains some definitions and examples that we will be using in the rest of this notes. Probability or chance can be measured on a scale which runs from zero to one, where zero represents impossibility and one represents certainty.

### 1.1.1 Sample Space

A sample space, $\Omega$, is the set of all possible outcomes of an experiment. The sample space can be classified into two main categories: discrete, where the space contains a finite or countable number of distinct point, and continuous when the space contains an uncountable distinct sample points. An event $E$ is defined to be a subset of the sample space, $E \subseteq \Omega$.

**EXAMPLE 1.1**

A manufacturer has five seemingly identical computer terminals available for shipping. Unknown to her, two of the five are defective. A particular order calls for two of the terminals and is filled by randomly selecting two of the five that are available.

**a** List the sample space for this experiment.

**b** Let $A$ denote the event that the order is filled with two non defective terminals. List the sample points in $A$.

**c** List the possible outcome for event $B$ where both terminals are defective.

**d** Let $C$ represent the case where at least one of the terminals is defective.

**Solution.**

**a** Let the two defective terminals be labelled $D_1$ and $D_2$ and let the three good terminals be labelled $G_1, G_2$, and $G_3$. Any single sample point will consist of a list of the two terminals selected for shipment. The simple events may be denoted by

$$\begin{array}{llll} E_1 = \{D_1, D_2\}, & E_5 = \{D_2, G_1\}, & E_8 = \{G_1, G_2\}, & E_{10} = \{G_2, G_3\}. \\ E_2 = \{D_1, G_1\}, & E_6 = \{D_2, G_2\}, & E_9 = \{G_1, G_3\}, & \\ E_3 = \{D_1, G_2\}, & E_7 = \{D_2, G_3\}, & & \\ E_4 = \{D_1, G_3\}, & & & \end{array}$$

Thus, The sample space $\Omega$ contains 10 sample points $\Omega = \{E_1, E_2, \dots, E_{10}\}$.

**b** Event $A = \{E_8, E_9, E_{10}\}$.

**c** $B = \{E_1\}$.

**d** $C = \{E_1, E_2, E_3, E_4, E_5, E_6, E_7\}$

◀

### 1.1.2 Probability Axioms

Suppose $\Omega$ is a sample space associated with an experiment. To every event $A$ in $\Omega$ ($A$ is a subset of $\Omega$), we assign a number, $P(A)$, called the probability of $A$, so that the following axioms hold:

1. $P(a) \geq 0$.
2. $P(\Omega) = 1$.
3. If $A_1, A_2, A_3, \dots$ form a sequence of pairwise mutually exclusive events in $\Omega$ ($A_i \cap A_j = \emptyset$ for $i \neq j$), then:

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum_{i=1}^{\infty} P(A_i).$$

***Other Consequences:***

**(i)** $P(\bar{A}) = 1 - P(A)$, therefore $P(\phi) = 0$.

**(ii)** For any two events $A_1$ and $A_2$ we have the addition rule:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

.

**EXAMPLE 1.2**

Following example (**??**), evaluate:

**a** Assign probabilities to the simple events in such a way that the information about the experiment is use.

**b** Find the probability of event $A, B, C$.

**c** Find the probability of $A \cap B$, $A \cup C$ and $B \cap C$.

**Solution. a** Because the terminals are selected at random, any pair of terminals is as likely to be selected as any other pair. Thus, $P(E_i) = 1/10$, for $i = 1, 2, \ldots, 10$, is a reasonable assignment of probabilities.

**d** Since $A = E_8 \cup E9 \cup E10$, then $P(A) = P(E_8) + P(E_9) + P(E_{10}) = 3/10$.

Also, $P(B) = 1/10$ and $P(C) = 7/10$.

**c** Since $A \cap B = \emptyset$, then $P(A \cap B) = 0$.

$A \cup C = \Omega$, then $P(A \cup C) = 1$.

$B \cap C = E_1$, then $P(B \cap C) = 1/10$.

◀

### 1.1.3 Conditional Probability

Suppose $P(A_2) \neq 0$. The conditional probability of the event $A_1$ given that the probability of the event $E_2$ is known, is defined as:

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

The conditional probability is undefined if $P(A_2) = 0$. The conditional probability formula above yields the multiplication rule:

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1)P(A_2|A_1) \\ &= P(A_2)P(A_1|A_2) \end{aligned} \tag{1.1}$$

### 1.1.4 Independence

Suppose that events $A_1$ and $A_2$ are in sample space $\Omega$, $A_1$ and $A_2$ are said to be independent if

$$P(A_1 \cap A_2) = P(A_1)P(A_2).$$

In the case of conditional probability, this implies to $P(A_1|A_2) = P(A_1)$ and $P(A_2|A_1) = P(A_2)$. That means that the knowledge of the occurrence of one of the events does not affect the likelihood of occurrence of the other. For more general case, $A_1, A_2, \ldots$ are pairwise independent if $P(A_i \cap A_j) = P(A_i)P(A_j)$, for all $i \neq j$. They are mutually independent if for all subsets $P(\cap_j A_j) = \prod_j P(A_j)$.

**EXAMPLE 1.3**

Back again to example (**??**), evaluate the probability of the event $A$ given $B$ and $B$ given $C$.

**Solution.**

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0.$$

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{1/10}{7/10} = 1/7.$$

◀

***Partition Law:*** Suppose $B_1, B_2, \ldots, B_k$ are mutually exclusive and exhaustive events, (i.e. $B_i \cap B_j = \emptyset$, for all $i \neq j$ and $\cup_i B_i = \Omega$ ). Let $A$ be any event, then

$$P(A) = \sum_{j=1}^{k} P(A|B_j)P(B_j).$$

. ***Bayes' Law:*** Suppose $B_1, B_2, \ldots, B_k$ are mutually exclusive and exhaustive events and $A$ is any event, then

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}.$$

**EXAMPLE 1.4**

(Cancer diagnosis) A screening programme for a certain type of cancer has reliabilities $P(A|D) = 0 : 98$, $P(A|\bar{D}) = 0 : 05$, where $D$ is the event "disease is present" and $A$ is the event "test gives a positive result". It is known that 1 in 10,000 of the population has the disease. Suppose that an individual's test result is positive. What is the probability that the person has the disease?

**Solution.** We require $P(D|A)$. First, we need to find $P(A)$:

$$P(A) = P(A|D)P(D) + P(A|\bar{D})P(\bar{D}) = 0.98 \times 0.0001 + 0.05 \times 0.9999 = 0.050093.$$

By the use of Bayes' rule;

$$P(D|A) = \frac{P(A|D)P(D)}{P(D)} = \frac{0.0001 \times 0.98}{0.050093} = 0.002.$$

Therefore, the person is still very unlikely to have the disease even though the test is positive. ◄

**EXAMPLE 1.5**

(Bertrand's Box Paradox) Three indistinguishable boxes contain black and white beads as shown: [ww], [wb], [bb]. A box is chosen at random and a bead chosen at random from the selected box. What is the probability of that the [wb] box was chosen given that selected bead was white?

**Solution.** Let $E$ represent the event of choosing the [wb] box, $W$ is the event of that the selected bead is white. By partition law: $P(W) = 1 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3}$. Then, using Bayes' rule gives:

$$P(E|W) = \frac{P(E)P(W|E)}{P(W)} = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3}.$$

This means, even though a bead from the selected box has been seen, the probability that the box is [wb] is still $1/3$. ◄

## 1.2 Random Variables

A ***Random variable*** $X$ is a real-valued function for which the domain is a sample space. Given a random experiment with sample space $\Omega$, then $X : \Omega \to \mathbb{R}$. The space of the r.v $X$ is the set of real numbers $\mathcal{A} = \{x : x = X(\omega), \omega \in \Omega\}$. Furthermore, for any event $A \subset \mathcal{A}$, then there is an event $\Psi \subset \Omega$, such that $P(A) = \Pr\{X \in A\} = P(\Psi)$, where $\Psi = \{\omega : \omega \in \Omega, X(\omega) \in A\}$ and $A = \{x : x = X(\omega), \omega \in \Psi\}$, knowing that $P(A)$ satisfy the probability axiom **??**.
***Note:*** A r.v $X$ is called discrete if it defined on a discrete sample space (countable or finite), and it is called a continuous r.v otherwise.

**EXAMPLE 1.6**

Toss a coin twice, the the sample space is: $\Omega = \{HH, HT, TH, TT\}$, suppose a r.v $X$ represent the number of heads. Then:

$$X(\omega) = \begin{cases} 0, & \text{if } \omega = TT \\ 1, & \text{if } \omega = TH, HT \\ 2, & \text{if } \omega = HH \end{cases} \tag{1.2}$$

Therefore, the space of $X$ is $\mathcal{A} = \{x : x = 0, 1, 2\}$, and the probability of $x = 0, 1, 2$: $\Pr X = 0 = 1/4$, $\Pr X = 1 = 1/2$ and $\Pr X = 2 = 1/4$.

Assume the event $A = \{x : x = 0, 1\} \subset \mathcal{A}$, then $P(A) = \Pr(X \in A) = \Pr(X = 0, 1) = \Pr(X = 0) + \Pr(X = 1) = 3/4$.

**EXAMPLE 1.7**

Let $\mathcal{A} = \{x : 0 < x < 2\}$ be the sample space of a r.v $X$. For each event $A \subset \mathcal{A}$, we define the probability set function $P(A)$ as

$$\begin{aligned} P(A) &= \int_{x \in A} \frac{3}{8} x^2 dx, && x \in \mathcal{A} \\ &= 0, && e.w \end{aligned} \tag{1.3}$$

If $A_1 = \{x : 0 < x < 1/2\}$ and $A_2 = \{x : 1/2 < x < 1\}$. Find the $P(A_1), P(A_1^c), P(A_2), P(A_2^c), P(A_1 \cap A_2), P(A_1 \cup A_2)$

**Solution.**

$$P(A_1) = \int_{x \in A_1} \frac{3}{8} x^2 dx = \frac{3}{8} \int_0^{1/2} x^2 dx = \frac{1}{64}.$$

$$P(A_1^c) = 1 - P(A_1) = 1 - \frac{1}{64} = \frac{63}{64}.$$

$$P(A_2) = \int_{x \in A_2} \frac{3}{8} x^2 dx = \frac{3}{8} \int_{1/2}^{1} x^2 dx = \frac{7}{8}.$$

$$P(A_2^c) = 1 - P(A_2) = 1 - \frac{7}{8} = \frac{1}{8}.$$

Since $A_1 \cap A_2 = \emptyset$, Then $P(A_1 \cap A_2) = 0$, and then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) = \frac{57}{64}.$$

◀

**EXAMPLE 1.8**

Let $\mathcal{A} = \{x : x = 1, 2, \ldots\}$ be the sample space of a r.v $X$. For each event $A \subset \mathcal{A}$, we define the probability set function $P(A)$ as

$$P(A) = \Pr(X \in A) = \sum_{x \in \mathring{A}} \left(\frac{1}{2}\right)^x, \quad x \in \mathcal{A}$$

If $A = \{x : x = 1, 2\}, B = \{x : x = 2, 3\}, C = \{x : x = 1, 3, 5, \dots\}$. Find $P(A),\ P(B),\ P(C),\ P(A^c),\ P(B^c),$ $P(C^c),\ P(A \cap B)$, and $P(A \cup B)$.

**Solution.**

$$P(A) = \sum_{x \in A} \left(\frac{1}{2}\right)^x = \sum_{x=1}^{2} \left(\frac{1}{2}\right)^x = \frac{1}{2} + \left(\frac{1}{2}\right)^2 = \frac{3}{4} \Rightarrow P(A^c) = \frac{1}{4}$$

$$P(B) = \sum_{x \in B} \left(\frac{1}{2}\right)^x = \sum_{x=2}^{3} \left(\frac{1}{2}\right)^x = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 = \frac{3}{8} \Rightarrow P(B^c) = \frac{5}{8}$$

$$P(C) = \sum_{x \in C} \left(\frac{1}{2}\right)^x = \sum_{x=1, step 2}^{\infty} \left(\frac{1}{2}\right)^x = \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \cdots = \frac{1/2}{1 - (1/2)^2} = \frac{2}{3} \Rightarrow P(C^c) = \frac{1}{3}$$

$$A \cap B = \{x : x = 2\} \Rightarrow P(A \cap B) = \sum_{x \in A \cap B} \left(\frac{1}{2}\right)^x = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) == \frac{3}{4} + \frac{7}{8} - \frac{1}{4} = \frac{7}{8}$$

◀

## CHAPTER 2

# DISTRIBUTION OF RANDOM VARIABLES

### 2.1 The Probability Density Function (PDF)

Notationally, we will use an upper case letter, such as $X$ or $Y$ , to denote a random variable and a lower case letter, such as $x$ or $y$, to denote a particular value that a random variable may assume. For example, let $X$ denote any one of the six possible values that could be observed on the upper face when a die is tossed. After the die is tossed, the number actually observed will be denoted by the symbol $x$. Note that $X$ is a random variable, but the specific observed value, $x$, is not random.

It is now meaningful to talk about the probability that $X$ takes on the value $x$, denoted by $\Pr(X = x)$.

***Definition***: The probability that $X$ takes on the value $x$, $\Pr(X = x)$, is defined as the sum of the probabilities of all sample points in $\Omega$ that are assigned the value $x$. We will sometimes denote $\Pr(X = x)$ by $p(x)$ or $f(x)$. Because $p(x)$ or $f(x)$ is a function that assigns probabilities to each value $x$ of the random variable $X$.

***Definition***: The probability distribution for a discrete variable $X$ can be represented by a formula, a table, or a graph that provides $p(x) = \Pr(X = x)$ for all $x$.

**EXAMPLE 2.1**

A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. Not wishing to show any biases in his selection, he decides to select the two workers at random. Let $X$ denote the number of women in his selection. Find the probability distribution for $X$.

**Solution.** The supervisor can select two workers from six in $\binom{6}{2} = 15$ ways. Hence $\Omega$ contains 15 sample points, which we assume to be equally likely because random sampling was employed. Thus, $\Pr(E_i) = 1/15, i = 1, 2, \ldots, 15$.

The values for $X$ that have nonzero probability are $0, 1$, and $2$. The number of ways of selecting $X = 0$ women is $\binom{3}{0}\binom{3}{2} = 1 \times 3 = 3$ sample points in the event $X = 0$, and

$$p(0) = \Pr(X = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{3}{15} = \frac{1}{5}.$$

Similarly,

$$p(1) = \Pr(X = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{9}{15} = \frac{3}{5}.$$

$$p(2) = \Pr(X = 2) = \frac{\binom{3}{2}\binom{3}{0}}{15} = \frac{3}{15} = \frac{1}{5}.$$

Notice that $(X = 1)$ is by far the most likely outcome. This should seem reasonable since the number of women equals the number of men in the original group. Therefore, we can write the probability function in the formula:

$$p(x) = \frac{\binom{3}{x}\binom{3}{2-x}}{\binom{6}{2}}, \quad x = 0, 1, 2.$$

Notice that, since $p(x) = \Pr(X = x)$ is a probability function, this means that the sum of $p(x)$ over the space is equal to one. ◀

***Theorem***: If $f(x)$ is a probability density function (pdf) for a discrete or continuous random variable $X$, then the following properties should be satisfied:

1. $f(x) \geq 0$, for all $x \in \mathcal{A}$.
2. 
   - discrete: $\sum_{x\in\mathcal{A}} f(x) = 1$.
   - continuous: $\int_{x\in\mathcal{A}} f(x) = 1$.

For any subset of the sample of the sample points $(A \subset \mathcal{A})$, a probability set function $p(x)$ can be expressed in term of the pdf $f(x)$ as:

$$p(A) = \Pr(x \in A) = \begin{cases} \sum_{x\in A} f(x), & discrete \\ \int_{x\in A} f(x), & continuous \end{cases}$$

**EXAMPLE 2.2**

Let $X$ be a discrete r.v defined on a sample set $\mathcal{A} = \{x : x = 0, 1, 2, 3\}$, and let $f(x)$ be a function defined on $\mathcal{A}$ as: $f(x) = \frac{1}{8}\binom{3}{x}, x \in \mathcal{A}$. Examine whether $f(x)$ is a pdf of $X$ or not. If so, find: $p(A)$ and $p(A^c)$, where $A = \{x : x = 1, 2\}$.

**Solution.**

- For the fist condition:

$$f(x) > 0, \forall x \in \mathcal{A}.$$

- For the second condition:

$$\sum_{x\in\mathcal{A}} f(x) = \frac{1}{8}\sum_{x=0}^{3}\binom{3}{x} = \frac{1}{8}\left[\binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3}\right] = \frac{1}{8}(1 + 3 + 3 + 1) = 1.$$

This proves that $f(x)$ is a pdf of $X$. Hence,

$$p(A) = \sum_{x \in A} f(x) = \frac{1}{8}\sum_{x=1}^{2}\binom{3}{x} = \frac{1}{8}\left[\binom{3}{1} + \binom{3}{2}\right] = \frac{3}{4},$$

and $p(A^c) = 1 - p(A) = 1 - \frac{3}{4} = \frac{1}{4}$. ◀

**EXAMPLE 2.3**

Let $X$ be a r.v defined on a sample set $\mathcal{A} = \{x : x = 1, 2, 3, \dots\}$ and let $f(x)$ be a function defined on $\mathcal{A}$ as: $f(x) = \left(\frac{1}{2}\right)^x,\ x \in \mathcal{A}$. Is $f(x)$ a pdf of $X$? If so, evaluate the following probabilities: $p(A), p(B), p(A \cap B), p(A \cup B)$ and $p(A|B)$, knowing that $A = \{x : x = 1, 2, 3\}$ and $B = \{x : x = 1, 3, 5, \dots\}$

**Solution.**

- $f(x) > 0, \quad \forall x \in \mathcal{A} = \{x : x = 1, 2, 3, \dots\}$.
- $$\sum_{x \in \mathcal{A}} f(x) = \sum_{x=1}^{\infty}\left(\frac{1}{2}\right)^x = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \dots = \frac{1/2}{1 - 1/2} = 1.$$

  Hence, $f(x)$ is a pdf of $X$.

In order to evaluate the probabilities:

- $$p(A) = \sum_{x \in A} f(x) = \sum_{x=1}^{3}\left(\frac{1}{2}\right)^x = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 = \frac{7}{8}$$
- $$p(B) = \sum_{x \in B} f(x) = \sum_{x=1, step2}^{\infty}\left(\frac{1}{2}\right)^x = \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \dots = \frac{1/2}{1 - 1/4} = \frac{2}{3}.$$
- $$A \cap B = \{x : x = 1, 3\} \Rightarrow p(A \cap B) = \sum_{x \in A \cap B} f(x) = \sum_{x=1,3}\left(\frac{1}{2}\right)^x = \frac{1}{2} + \left(\frac{1}{2}\right)^3 = \frac{5}{8}.$$
- $$p(A \cup B) = p(A) + p(B) - p(A \cap B) = \frac{7}{8} + \frac{2}{3} - \frac{5}{8} = \frac{11}{12}.$$
- $$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{5/8}{2/3} = \frac{15}{16}.$$

◀

**EXAMPLE 2.4**

Let $X$ be a r.v defined on a sample set $\mathcal{A} = \{x : 2 \le x \le 4\}$ and let $f(x)$ be a function defined on $\mathcal{A}$ as: $f(x) = \frac{1}{8}(x+1),\ x \in \mathcal{A}$. Examine whether $f(x)$ is a pdf of $X$? If it is, find $p(A)$, where $A = \{x : 1.5 \le x \le 2.5\}$.

**Solution.**

- $f(x) > 0, \quad \forall x \in \mathcal{A} = \{x : 2 \le x \le 4\}$.
- $$\int_{x\in\mathcal{A}} f(x)dx = \int_2^4 \frac{1}{8}(x+1)dx = \frac{1}{16}(x+1)^2\Big|_2^4 = \frac{1}{16}(25-9) = 1.$$
  Hence, $f(x)$ is a pdf of $X$.
- $$p(A) = \int_{x\in A} f(x)dx = \int_{1.5}^{2} f(x)dx + \int_2^{2.5} f(x)dx = 0 + \int_2^{2.5} \frac{1}{8}(x+1)dx = \frac{1}{16}(x+1)^2\Big|_2^2.5 = \frac{13}{320}.$$

◀

**EXAMPLE 2.5**

Suppose that the function $f(x) = e^{-x},\ x \in \mathcal{A}$ is defined os a sample set $\mathcal{A} = \{x : 0 < x < \infty\}$ and that $X$ is a r.v. Show that $f(x)$ is a pdf of $X$ and evaluate $p(A), p(B), p(A \cap B)$ and $p(A \cup B)$, if $A = \{x : 0 < x < 3\}$ and $B = \{x : 1 < x < \infty\}$.

**Solution.**

- $f(x) > 0, \quad \forall x \in \mathcal{A} = \{x : 2 \le x \le 4\}$.
- $$\int_{x\in\mathcal{A}} f(x)dx = \int_0^\infty e^{-x}dx = -e^{-x}\Big|_0^\infty = -(e^{-\infty} - e^0) = 1.$$
  Hence, $f(x)$ is a pdf of $X$.

For the probability evaluations:

- $$p(A) = \int_{x\in A} f(x)dx = \int_0^3 e^{-x}dx = 1 - e^{-3}.$$
- $$p(B) = \int_{x\in B} f(x)dx = \int_1^\infty e^{-x}dx = e^{-1}.$$
- $$A \cap B = \{x : 1 < x < 3\} \Rightarrow p(A \cap B) = \int_1^3 e^{-x}dx = e^{-1} - e^{-3}.$$
- $$p(A \cup B) = p(A) + p(B) - p(A \cap B) = 1 - e^{-3} + e^{-1} - e^{-1} + e^{-3} = 1.$$

◀

**EXAMPLE 2.6**

Verify that the following functions are pdf's of a r.v $X$ that defined as:

1. $f(x) = x^{-2}, \quad \mathcal{A} = \{x : 1 < x < \infty\}$.
2. $f(x) = \frac{4}{9}\binom{2}{x}\left(\frac{1}{2}\right)^x, \quad \mathcal{A} = \{x : x = 0, 1, 2\}$.
3. $f(x) = 1 - |1 - x|, \quad \mathcal{A} = \{x : 0 < x < 2\}$.
4. $f(x) = \begin{cases} 1 + x, & -1 < x < 0 \\ 1 - x, & 0 \le x < 1 \end{cases}$

**Solution.**
3. $f(x) \ge 0, \forall x \in \mathcal{A} = \{x : 0 < x < 2\}$.

$$f(x) = 1 - |1 - x| = \begin{cases} 1 - (1 - x), & 1 - x \ge 0 \to x \le 1 \\ 1 + (1 - x), & 1 - x < 0 \to x > 1 \end{cases}$$

$$f(x) = \begin{cases} x, & 0 < x \le 1 \\ 2 - x, & 1 < x < 2 \end{cases}$$

then,

$$\int_{x \in \mathcal{A}} f(x)dx = \int_0^1 x dx + \int_1^2 (2 - x)dx = \frac{1}{2}x^2\Big|_0^1 + \frac{1}{2}(2 - x)^2\Big|_1^2 = \frac{1}{2} + \frac{1}{2} = 1$$

◀

**EXAMPLE 2.7**

Find the constant $c$ that makes each of the following function pdf of a r.v $X$:

1. $f(x) = c(x + 1), \quad x = 0, 1, 2, 3$.
2. $f(x) = c(x^{\alpha-1} - x^{\beta-1}), \quad 0 < x < 1, \alpha > 1, \beta > 0$.
3. $f(x) = c(1 + x^2)^{-1}, \quad -\infty < x < \infty$.

**Solution.** Since $f(x)$ is a pdf, then it should satisfy the properties of the pdf, hence:

1. $$1 = \sum_{x \in \mathcal{A}} f(x) = \sum_{x=0}^{3} c(x + 1) = c(1 + 2 + 3 + 4) = 10c \Rightarrow c = 1/10.$$

2. $$1 = \int_{x \in \mathcal{A}} f(x)dx = \int_0^1 c(x^{\alpha-1} - x^{\beta-1})dx = c\left[\frac{x^\alpha}{\alpha} - \frac{x^\beta}{\beta}\right]_0^1 = c\left(\frac{1}{\alpha} - \frac{1}{\beta}\right) = c\frac{\beta - \alpha}{\alpha\beta} \Rightarrow c = \frac{\alpha\beta}{\beta - \alpha}, \beta \ne \alpha$$

3. $$1 = \int_{-\infty}^{\infty} c(1 + x^2)^{-1}dx = c\tan^{-1} x\Big|_{-\infty}^{\infty} = c\left[\tan^{-1}(\infty) - \tan^{-1}(-\infty)\right] = c\left[\tan^{-1}(\infty) + \tan^{-1}(\infty)\right]$$

$$= 2c\tan^{-1}(\infty) = 2c\frac{\pi}{2} \Rightarrow c = \frac{1}{\pi}.$$

◀

***Definition***: The **Mode** of the distribution is the value of $x$ that maximises the pdf $f(x)$ of a r.v $X$. Note that the mode of a continuous r.v is the solution of $f'(x) = 0$ and $f(x) < 0$. Also, the mode may not be existed or a distribution may have more than one mode.

**EXAMPLE 2.8**

Find the mode for the following pdf's:

1. $f(x) = \left(\frac{1}{2}\right)^x, \quad x = 1, 2, 3, \ldots.$
2. $f(x) = 12x^2(1-x), \quad 0 < x < 1.$

**Solution.**

1. $f(x) = \left(\frac{1}{2}\right)^x, \quad x = 1, 2, 3, \ldots$, then $x = 1$ is the mode.
2. $f(x) = 12x^2(1-x), \quad 0 < x < 1 \Rightarrow f'(x) = 12(2x - 3x^2)$, set $f'(x) = 0$:

$$12(2x - 3x^2) = 0 \Rightarrow x = 0, 2/3.$$

then, $f''(x) = 12(2 - 6x) = 24(1 - 3x)$.

$$f''(0) = 24(1 - 0) = 24 > 0$$

$$f''(2/3) = 24(1 - 3(2/3)) = -24 < 0$$

hence, $x = 2/3$ is the mode.

◀

### 2.1.1 The Probability Density Function in $n-$Dimensional Space

Let $X_1, X_2, \ldots, X_n$ be an $n$ r.v's (discrete or continuous) defined on $n-$D sample space $\mathcal{A}$, and let $\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(x_1, x_2, \ldots, x_n)$ be a function defined on $\mathcal{A}$, such that:

1. $f(x_1, x_2, \ldots, x_n) \geq 0, \forall (x_1, x_2, \ldots, x_n) \in \mathcal{A}$.
2. $$1 = \begin{cases} \sum\sum\cdots\sum_{(x_1,x_2,\ldots,x_n)\in\mathcal{A}} f(x_1, x_2, \ldots, x_n) \\ \int\int\cdots\int_{(x_1,x_2,\ldots,x_n)\in\mathcal{A}} f(x_1, x_2, \ldots, x_n)dx_1, dx_2, \ldots, dx_n \end{cases}$$

Then the function $f(x_1, x_2, \ldots, x_n)$ is called pdf of r.v's $X_1, X_2, \ldots, X_n$. Furthermore, for all event $A \subset \mathcal{A}$, the probability of $A$, $p(A)$, can be expressed in terms of the pdf by:

$$p(A) = \Pr\{(X_1, X_2, \ldots, X_n) \in A\} = \begin{cases} \sum\sum\cdots\sum_{(x_1,x_2,\ldots,x_n)\in A} f(x_1, x_2, \ldots, x_n). \\ \int\int\cdots\int_{(x_1,x_2,\ldots,x_n)\in A} f(x_1, x_2, \ldots, x_n)dx_1, dx_2, \ldots, dx_n. \end{cases}$$

**EXAMPLE 2.9**

Let $X$ and $Y$ be discrete r.v's defined on a sample space $\mathcal{A} = \{(x, y) : x = 1, 2, 3; y = 1, 2\}$, and let $f(x, y)$ be a function defined on $\mathcal{A}$ by $f(x, y) = \frac{1}{21}(x + y), \quad (x, y) \in \mathcal{A}$.

1. Is $f(x, y)$ a pdf of $X$ and $Y$?
2. If so, find $p(A)$ and $p(A^c)$, where $A = \{(x, y) : x = 1, 2; y = 1\}$.

**Solution.**

1. - $f(x, y) > 0, \quad \forall (x, y) \in \mathcal{A}$.
   -
$$\sum\sum\nolimits_{(x,y)\in\mathcal{A}} f(x, y) = \sum_{x=1}^{3}\sum_{y=1}^{2} \frac{1}{21}(x + y) = \frac{1}{21}\sum_{x=1}^{3}[(x + 1) + (x + 2)]$$
$$= \frac{1}{21}\sum_{x=1}^{3}(2x + 3) = \frac{1}{21}[5 + 7 + 9] = 1.$$

   Then $f(x, y)$ is a pdf of $X$ and $Y$.

2.
$$p(A) = \sum\sum_{(x,y)\in A} f(x, y) = \sum_{x=1}^{2}\sum_{y=1}^{1} \frac{1}{21}(x + y) = \frac{1}{21}\sum_{x=1}^{2}(x + 1) = \frac{1}{21}(2 + 3) = \frac{5}{21}$$
$$\Rightarrow p(A^c) = 1 - \frac{5}{21} = \frac{16}{21}.$$

◀

**EXAMPLE 2.10**

Let $X$ and $Y$ be two r.v's defined on a sample space $\mathcal{A} = \{(x, y) : x = 1, 2, \ldots; y = 0, 1, 2\}$, and let $f(x, y)$ be a function defined on $\mathcal{A}$ by
$$f(x, y) = \binom{2}{y}\left(\frac{1}{2}\right)^{x+2}, \quad (x, y) \in \mathcal{A}$$

1. Is $f(x, y)$ a pdf of $X$ and $Y$?
2. Find $p(A)$, where $A = \{(x, y) : x = 1, 3, 5, \ldots; y = 1\}$.

**Solution.**

1. - $f(x, y) > 0, \quad \forall (x, y) \in \mathcal{A}$.
   -
$$\sum\sum\nolimits_{(x,y)\in\mathcal{A}} f(x, y) = \sum_{x=1}^{\infty}\sum_{y=0}^{2}\binom{2}{y}\left(\frac{1}{2}\right)^{x+2} = \sum_{x=1}^{\infty}\left(\frac{1}{2}\right)^{x+2}\left[\binom{2}{0} + \binom{2}{1} + \binom{2}{2}\right]$$
$$= \sum_{x=1}^{\infty}\left(\frac{1}{2}\right)^{x+2}(1 + 2 + 1) = 4\left[\left(\frac{1}{2}\right)^{3} + \left(\frac{1}{2}\right)^{4} + \left(\frac{1}{2}\right)^{5} + \ldots\right] = 4\frac{(1/2)^3}{1 - (1/2)} = 1$$

Then $f(x,y)$ is a pdf of $X$ and $Y$.

2.

$$p(A) = \sum\sum_{(x,y)\in A} f(x,y) = \sum_{x=1,step2}^{\infty}\sum_{y=1}^{1}\binom{2}{y}\left(\frac{1}{2}\right)^{x+2} = \sum_{x=1,step2}^{\infty}\left(\frac{1}{2}\right)^{x+2}\binom{2}{1} = 2\sum_{x=1,step2}^{\infty}\left(\frac{1}{2}\right)^{x+2}$$

$$= 2\left[\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^7 + \dots\right] = 2\frac{(1/2)^3}{1-(1/2)^2} = \frac{1}{3}.$$

◀

**EXAMPLE 2.11**

Let $X$ and $Y$ be continuous r.v's defined on a sample space $\mathcal{A} = \{(x,y) : 0 < x < 2; 2 < y < 4\}$, and let $f(x,y)$ be a function defined on $\mathcal{A}$ as $f(x,y) = \frac{1}{8}(6-x-y), \quad (x,y)\in\mathcal{A}$

1. Is $f(x,y)$ a pdf of $X$ and $Y$?
2. Find $p(A)$, where $A = \{(x,y) : 0 < x < 1; 2 < y < 3\}$.

**Solution.**

1. - $f(x,y) \geq 0, \quad \forall(x,y)\in\mathcal{A}$.
   -

$$\int\int_{(x,y)\in\mathcal{A}} f(x,y) = \int_{x=0}^{2}\int_{y=2}^{4}\frac{1}{8}(6-x-y)dydx = \frac{1}{8}\int_0^2\left[6y - xy - \frac{y^2}{2}\right]_2^4 dx$$

$$= \frac{1}{8}\int_0^2\left[(24-4x-8)-(12-2x-2)\right]dx = \frac{1}{8}\int_0^2(6-2x)dx = \frac{1}{8}\left[6x-x^2\right]_0^2 = \frac{12-4}{8} = 1$$

Then $f(x,y)$ is a pdf of $X$ and $Y$.

2.

$$p(A) = \int\int_{(x,y)\in A} f(x,y) = \int_{x=0}^{1}\int_{y=2}^{3}\frac{1}{8}(6-x-y)dydx = \cdots = \frac{3}{8}$$

◀

**EXAMPLE 2.12**

$X$ and $Y$ are two r.v's defined on a sample space $\mathcal{A} = \{(x,y) : 0 < x < y < 1\}$, and let $f(x,y) = 2 \quad (x,y)\in\mathcal{A}$ is a function defined on $\mathcal{A}$. is $f(x,y)$ a pdf of $X$ and $Y$?

**Solution.**

- $f(x,y) = 2 > 0, \quad \forall(x,y)\in\mathcal{A}$.
-

$$\int\int_{(x,y)\in\mathcal{A}} f(x,y) = 2\int_{x=0}^{1}\int_{y=x}^{1} dydx = 2\int_0^1 y\Big|_x^1 dx = 2\int_0^1(1-x)dx = 2\left[x - \frac{x^2}{2}\right]_0^1 = 2(1-\frac{1}{2}) = 1$$

Then $f(x,y)$ is a pdf of $X$ and $Y$. ◀

***Note***: If $f(x_1,x_2,\ldots,x_n)$ is a pdf of r.v's $X_1,X_2,\ldots,X_n$ defined on a sample space $\mathcal{A}=\{(x_1,x_2,\ldots,x_n):-\infty<x_i<\infty; i=1,2,\ldots,n\}$, and if event $A\subset\mathcal{A}$, where $A=\{(x_1,x_2,\ldots,x_n):a_i<x_i<b_i; i=1,2,\ldots,n\}$. Then the probability of $A$ is:

$$p(A)=\Pr\{a_i<x_i<b_i; i=1,2,\ldots,n\}=\begin{cases}\sum\limits_{x_1=a_1}^{b_1}\sum\limits_{x_2=a_2}^{b_2}\cdots\sum\limits_{x_n=a_n}^{b_n} f(x_1,x_2,\ldots,x_n)\\ \int\limits_{x_1=a_1}^{b_1}\int\limits_{x_2=a_2}^{b_2}\cdots\int\limits_{x_n=a_n}^{b_n} f(x_1,x_2,\ldots,x_n)dx_1,dx_2,\ldots,dx_n\end{cases}$$

**EXAMPLE 2.13**

Find the constant $c$ that makes the function $f(x,y)=ce^{-x-y},\quad 0<x<y<\infty$ a pdf of r.v's $X$ and $Y$.

**Solution.** Since $f(x,y)$ is a pdf of $X$ and $Y$, then by definition $\int\int_{(x,y)\in\mathcal{A}} f(x,y)dxdy=1$:

$$1=c\int_{x=0}^{\infty}\int_{y=x}^{\infty}e^{-x-y}dydx=c\int_0^{\infty}e^{-x}\left[\int_x^{\infty}e^{-y}dy\right]dx=c\int_0^{\infty}e^{-x}\left[-e^{-y}\right]_x^{\infty}dx$$

$$=c\int_0^{\infty}e^{-x}(0+e^{-x})dx=c\int_0^{\infty}e^{-2x}dx=c\left[\frac{1}{2}e^{-2x}\right]_0^{\infty}=c(0+\frac{1}{2})=\frac{c}{2}.$$

Therefore, $c=2$. ◀

## 2.2 Cumulative Distribution Function (CDF)

***Definition***: The cumulative distribution function or the probability distribution of a r.v $X$, denoted by $F(x)=\Pr(X\le x), -\infty<x<\infty$. Let $X$ be a r.v has a pdf $f(x)$ defined on a sample space $\mathcal{A}$, we define:

$$F(x)=\Pr(X\le x)=\Pr(-\infty<x<\infty)=\begin{cases}\sum\limits_{t=-\infty}^{x} f(t), & \text{discrete.}\\ \int\limits_{t=-\infty}^{x} f(t)dt, & \text{continuous.}\end{cases}$$

**EXAMPLE 2.14**

Suppose that $X$ has a pdf $p(x)$, defined on $x=0,1,2$ as:

$$p(x)=\binom{2}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{2-x},\quad x=0,1,2.$$

Find $F(x)$ for all $x$.

**Solution.** From the pdf $p(x)$, we have: $p(0)=1/4, p(1)=1/2, p(2)=1/4$. In order to find the distribution function $F(x)$, from definition of $F(x)=\Pr(X\le x)$, we should evaluate the probability for four regions of $x$:

$$-\infty<x<0,\ 0\le x<1,\ 1\le x<2 \text{ and } 2\le x<\infty.$$

- For $x < 0$: Because the only value of $X$ that are assigned positive probabilities are $0, 1$ and 2; and that non of these values are less than 0, then $f(x) = 0, \ \forall - \infty < x < 0$.
- For $0 \leq x < 1$:
$$F(x) = \Pr(0 \leq x < 1) = \Pr(X < 0) + \Pr(X = 0) = 0 + \frac{1}{4} = \frac{1}{4}.$$
- For $1 \leq x < 2$:
$$F(x) = \Pr(1 \leq x < 2) = \Pr(X = 0) + \Pr(X = 1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}.$$
- For $x \geq 2$:
$$F(x) = \Pr(X \geq 2) = \Pr(X < 0) + \Pr(0 \leq X \leq 1) + \Pr(1 \leq X < 2) = 0 + \frac{1}{4} + \frac{3}{4} = 1.$$

Therefore, the distribution probability is:

$$F(x) = \Pr(X \leq x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

◀

***Notes***:

1. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$.
2. $F(\infty) = \lim_{x \to \infty} F(x) = 1$.
3. $F(x)$ is a non-decreasing function of $x$, if $x_1 < x_2$ then $F(x_1) \leq F(x_2)$.
4. $0 \leq F(x) \leq 1$, because $0 \leq \Pr(X \leq x) \leq 1$.
5. For a continuous r.v $X$, the pdf could be evaluated in terms of the cdf as:
$$f(x) = \frac{d}{dx} F(x).$$
6. For a discrete r.v $X$, the pdf could be evaluated in terms of the cdf as:
$$f(x) = \Pr(X = x) = \Pr(X \leq x) - \Pr(X \leq x - 1) = F(x) - F(x - 1).$$

**EXAMPLE 2.15**

Let the r.v has a pdf $f(x) = \frac{x}{6}; \ x = 1, 2, 3$. Find the cdf of $X$

**Solution.**

$$F(x) = \Pr(X \leq x) = \sum_{\tau=-\infty}^{x} f(\tau) = \sum_{\tau=1}^{x} \frac{\tau}{6} = \frac{1}{6}(1 + 2 + \cdots + x) = \frac{x(x+1)}{12}$$

$$\therefore F(x) = \begin{cases} 0, & x < 1 \\ \frac{x(x+1)}{12}, & 1 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

◀

**EXAMPLE 2.16**

Let the r.v has a pdf $f(x) = 2x^{-3};\ 1 < x < \infty$. Find cdf of $X$

**Solution.**

$$F(x) = \Pr(X \leq x) = \int_{\tau=-\infty}^{x} f(\tau)d\tau = \int_{\tau=1}^{x} 2x^{-3}d\tau = -\tau^{-2}\Big|_{1}^{x} 1 - \frac{1}{x^2}$$

$$\therefore F(x) = \begin{cases} 0, & x \leq 1 \\ 1 - \frac{1}{x^2}, & 1 \leq x < \infty \\ 1, & x = \infty \end{cases}$$

◄

***Remark***: In order to evaluate the probability of a r.v $X$ between $a$ and $b$, we have two ways:

1. Using the pdf as:

$$\Pr(a \leq X \leq b) = \begin{cases} \sum_{x=a}^{b} f(x), & discrete \\ \int_{x=a}^{b} f(x), & continuous \end{cases}$$

2. Using the cdf as:

$$\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X \leq a) = F(b) - F(a).$$

**EXAMPLE 2.17**

Let the r.v has a pdf $f(x) = \left(\frac{1}{2}\right)^x;\ x = 1, 2, 3, \ldots.$

1. Find the cdf of $X$.
2. Evaluate $\Pr(2 \leq X \leq 4)$ using the pdf and the cdf.

**Solution.** 1.

$$F(x) = \sum_{\tau=-\infty}^{x} f(\tau) = \sum_{\tau=1}^{x} \left(\frac{1}{2}\right)^{\tau} = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \cdots + \left(\frac{1}{2}\right)^x = \frac{(1/2)[1 - (1/2)^x]}{1 - (1/2)}$$

$$\therefore F(x) = \begin{cases} 0, & x < 1 \\ 1 - \left(\frac{1}{2}\right)^x, & 1 \leq x < \infty \\ 1, & x = \infty \end{cases}$$

2. Using the pdf:

$$\Pr(2 \leq x \leq 4) = \sum_{x=2}^{4} \left(\frac{1}{2}\right)^x = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 = \frac{7}{16}.$$

Using the cdf:

$$\Pr(2 \leq x \leq 4) = F(4) - F(2) = 1 - \left(\frac{1}{2}\right)^4 - 1 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} - \frac{1}{16} = \frac{7}{16}$$

◄

***Definition***: The **Median** of a r.v $X$ is the value of $x$ for which the cdf $F(x) = \frac{1}{2}$.

**EXAMPLE 2.18**

Find the median of the following distributions, where their pdf's are defined as follows:

1. $f(x) = \left(\frac{1}{2}\right)^x, \; x = 1, 2, 3, \ldots.$
2. $f(x) = \frac{x}{6}, \; x = 1, 2, 3.$
3. $f(x) = 3x^2, \; 0 < x < 1.$
4. $f(x) = \frac{1}{\pi}(1 + x^2)^{-1}, \; -\infty < x < \infty.$

**Solution.** 1. From ex (**??**), $F(x) = 1 - \left(\frac{1}{2}\right)^x$. To evaluate the median, set $F(x) = \frac{1}{2}$:

$$1 - \left(\frac{1}{2}\right)^x = \frac{1}{2} \Rightarrow \left(\frac{1}{2}\right)^x = \frac{1}{2} \Rightarrow x = 1.$$

The median $x = 1$.

2. From ex (**??**), $F(x) = \frac{x(x+1)}{12}$. To evaluate the median, set $F(x) = \frac{1}{2}$:

$$\frac{x(x+1)}{12} = \frac{1}{2} \Rightarrow x(x+1) = 6 \Rightarrow x^2 + x - 6 = 0 \Rightarrow x = -3, 2.$$

The median $x = 2$ since $-3 \notin \mathcal{A}$.

3. $F(X) = \int_{\tau=-\infty}^{x} f(\tau)d\tau = \int_{\tau=0}^{x} 3\tau^2 d\tau = \tau^3\Big|_0^x = x^3, \; 0 < x < 1.$

Set $F(x) = \frac{1}{2} \Rightarrow x^3 = \frac{1}{2} \Rightarrow x = \frac{1}{\sqrt[3]{2}}$ is the median.

4. $F(X) = \int_{\tau=-\infty}^{x} f(\tau)d\tau = \int_{\tau=-\infty}^{x} \frac{1}{\pi}(1+\tau)^{-1}d\tau = \frac{1}{\pi}\tan^{-1}\tau|_{-\infty}^{x} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}x, \; -\infty < x < \infty.$

Set $F(x) = \frac{1}{2} \Rightarrow \frac{1}{2} + \frac{1}{\pi}\tan^{-1}x = \frac{1}{2} \Rightarrow x = 0$ is the median.

◀

**EXAMPLE 2.19**

Find the constant $c$ in the following cdf's and find the pdf for each case.

1. $X$ is a r.v with a cdf:

$$F(x) = \begin{cases} 0, & x < 1 \\ c[1 - (1/2)^x], & 1 \leq x < \infty \\ 1, & x = \infty \end{cases}$$

2. $X$ is a r.v has a cdf:

$$F(x) = \begin{cases} 0, & x \leq \\ cx(x+1), & 0 < x < 3 \\ 1, & x \geq 3 \end{cases}$$

**Solution.**

1. Since $F(x)$ is a cdf, then: $F(\infty) = 1 \Rightarrow c[1 - (1/2)^{\infty}] = 1 \Rightarrow c(1 - 0) = 1 \Rightarrow c = 1$. then:

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - \left(\frac{1}{2}\right)^x, & 1 \le x < \infty \\ 1, & x = \infty \end{cases}$$

Therefore, the pdf $f(x) = F(x) - F(x-1)$:

$$f(x) = 1 - \left(\frac{1}{2}\right)^x - 1 + \left(\frac{1}{2}\right)^{x-1} = \left(\frac{1}{2}\right)^{x-1} - \left(\frac{1}{2}\right)^x = \left(\frac{1}{2}\right)^{x-1}\left(1 - \frac{1}{2}\right)$$

$$\therefore f(x) = \left(\frac{1}{2}\right)^x, \quad x = 1, 2, 3, \ldots$$

2. Since $F(\infty)$ is a cdf, then: $F(3) = 1 \Rightarrow 3c(3+1) = 1 \Rightarrow 12c = 1 \Rightarrow c = \frac{1}{2}$, then:

$$F(x) = \begin{cases} 0, & x \le 0 \\ \frac{1}{12}x(x+1), & 0 < x < 3 \\ 1, & x \ge 3 \end{cases}$$

Therefore, the pdf $f(x) = F'(x)$:

$$\therefore f(x) = \frac{1}{12}(2x+1), \quad 0 < x < 3$$

◀

**EXAMPLE 2.20**

Find the cdf of the r.v $X$ which has a pdf $f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \le x < 2 \end{cases}$

**Solution.**

$$F(x) = \Pr(X \le x) = \begin{cases} 0, & x \le 0 \\ \int\limits_0^x f(\tau)d\tau = \int\limits_0^x \tau d\tau = \frac{1}{2}\tau^2\Big|_0^x = \frac{1}{2}x^2, & 0 < x < 1 \\ \int\limits_0^1 f(\tau)d\tau + \int\limits_1^x f(\tau)d\tau = 1 - \frac{1}{2}(2-x)^2, & 1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$

◀

### 2.2.1 The Cumulative Distribution Function in $n-$Dimensional Space

Let $X_1, X_2, \ldots, X_n$ be $n$ r.v's defined on an $n-$Dimensional sample space $\mathcal{A}$ with pdf $f(x_1, x_2, \ldots, x_n)$. We define the cdf of $X_1, X_2, \ldots, X_n$ as:

$$\begin{aligned} F(x_1, x_2, \ldots, x_n) &= \Pr(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) \\ &= \begin{cases} \sum\limits_{\tau_1=-\infty}^{x_1} \sum\limits_{\tau_2=-\infty}^{x_2} \cdots \sum\limits_{\tau_n=-\infty}^{x_n} f(\tau_1, \tau_2, \ldots, \tau_n), & \text{discrete.} \\ \int\limits_{\tau_1=-\infty}^{x_1} \int\limits_{\tau_2=-\infty}^{x_2} \cdots \int\limits_{\tau_n=-\infty}^{x_n} f(\tau_1, \tau_2, \ldots, \tau_n)d\tau_1 d\tau_2 \ldots d\tau_n, & \text{continuous.} \end{cases} \end{aligned}$$

**EXAMPLE 2.21**

Let the pdf of r.v's $X$ and $Y$ be: $f(x,y) = \frac{x}{6}\left(\frac{1}{2}\right)^y$; $x = 1,2,3,\ y = 1,2,\ldots$. Find the cdf of $X$ and $Y$.

**Solution.**

$$
\begin{aligned}
F(x,y) = \Pr(X \le x, Y \le y) &= \sum_{t=-\infty}^{x}\sum_{s=-\infty}^{y} f(t,s) = \sum_{t=1}^{x}\sum_{s=1}^{y}\frac{t}{6}\left(\frac{1}{2}\right)^s \\
&= \sum_{t=1}^{3}\frac{t}{6}\left[\frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1}{2}\right)^y\right] = \sum_{t=1}^{3}\frac{t}{6}\frac{(1/2)[1-(1/2)^y]}{1-(1/2)} \\
&= [1-(1/2)^y]\frac{1}{6}(1+2+\cdots+x) = \frac{1}{6}\frac{x(x+1)}{2}[1-(1/2)^y]
\end{aligned}
$$

$$
\therefore F(x,y) = \begin{cases} 0, & x<1, y<1 \\ \frac{x(x+1)}{12}[1-\left(\frac{1}{2}\right)^y], & 1 \le x < 3, 1 \le y < \infty \\ 1, & x \ge 3, y = \infty \end{cases}
$$

◀

**EXAMPLE 2.22**

Let $X$ and $Y$ be two r.v's defined on sample space $\mathcal{A} = \{(x,y) : 0 < x < 2;\ 2 < y < 4\}$, and let $f(x,y) = \frac{1}{8}(6-x-y)$ is a pdf of $X$ and $Y$. Find the cdf.

**Solution.**

$$
\begin{aligned}
F(x,y) = \Pr(X \le x, Y \le y) &= \int_{t=-\infty}^{x}\int_{s=-\infty}^{y} f(t,s)dtds = \int_{0}^{x}\int_{2}^{y}\frac{1}{8}(6-t-s)dsdt \\
&= \frac{1}{8}\int_2^y\left[6t - \frac{1}{2}t^2 - st\right]_0^x ds = \frac{1}{8}\int_2^y\left(6x - \frac{1}{2}x^2 - xs\right)ds = \frac{1}{8}\left[6xs - \frac{1}{2}x^2s - \frac{1}{2}xs^2\right]_2^y \\
&= \frac{1}{8}\left(6xy - \frac{1}{2}x^2y - \frac{1}{2}xy^2 - 10x + x^2\right) = \frac{x}{16}(12y - xy - y^2 - 20 + 2x).
\end{aligned}
$$

$$
\therefore F(x,y) = \begin{cases} 0, & x \le 0,\ y \le 2 \\ \frac{x}{16}(12y - xy - y^2 - 20 + 2x), & 0 < x < 2,\ 2 < y < 4 \\ 1, & x \ge 2,\ y \ge 4 \end{cases}
$$

◀

***Note***:

1. If $X_1, X_2, \ldots, X_n$ are continuous r.v's, then the pdf is:

$$
f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1, \partial x_2, \ldots, \partial x_n}.
$$

2. For $2-$dimension discrete sample space, the pdf could be written as:

$$
f(x,y) = F(x,y) - F(x,y-1) - F(x-1,y) + F(x-1,y-1).
$$

**EXAMPLE 2.23**

Let the r.v's $X$, $Y$ and $Z$ have pdf $f(x,y,z) = 6\, e^{-(x+y+z)},\ 0 < x < y < z < \infty$. Find the cdf of $X, Y$ and $Z$.

**Solution.**

$$F(x,y,z) = \Pr(X \le x, Y \le y, Z \le z) = \int_{t=-\infty}^{x}\int_{s=-\infty}^{y}\int_{q=-\infty}^{z} f(t,s,q)dtdsdq = \int_0^x\int_t^y\int_s^z e^{-(t+s+q)}dqdsdt$$

$$\therefore F(x,y,z) = \begin{cases} 0, & x \le 0,\ y \le 0,\ z \le 0 \\ 6\left[\frac{1}{2}e^{-(x+2y)} - e^{-(x+y+z)} - \frac{1}{6}e^{-3x} + \frac{1}{2}e^{-(2x+z)} - \frac{1}{2}e^{-2y} + e^{-(y+z)} - \frac{1}{2}e^{-z} + \frac{1}{6}\right], & 0 < x < y < z < \infty \\ 1, & x = y = z = \infty \end{cases}$$

◄

## 2.3 Transformation of Variables (cdf technique)

Assume that $X$ is a r.v defined on sample space $\mathcal{A}$ and has pdf $f(x)$ and cdf $F(x)$. Consider a new r.v $Y$ as a function of $X$, say $Y = \psi(X)$ defined on a sample space $\mathcal{B} = \{y : y = \psi(x),\ x \in \mathcal{A}\}$. The aim now is to find the distribution of r.v $Y$, let $G(y) = \Pr(Y \le y)$ and $g(x)$ represent the cdf and the pdf of $Y$. If the function $y = \psi(x)$ is a one-to-one transformation that maps the space $\mathcal{A}$ on to the space $\mathcal{B}$ ($y = \psi(x) : \mathcal{A} \xrightarrow[on\ to]{1-1} \mathcal{B}$), then the inverse function $x = \psi^{-1}(y)$ exist. Therefore the cdf of $Y$ can be written as:

$$G(y) = \Pr(Y \le y) = \Pr(\psi(X) \le y) = \Pr(X \le \psi^{-1}(y)) = F\big(\psi^{-1}(y)\big), \quad y \in \mathcal{B}.$$

Hence, the pdf $g(y)$ is:

$$g(y) = \begin{cases} G(y) - G(y-1), & \text{discrete} \\ G'(y), & \text{continuous} \end{cases}$$

**EXAMPLE 2.24**

Let the r.v $X$ has a pdf $f(x) = \frac{1}{2},\ -1 < x < 1$. Find the distribution of r.v $Y = X^2$.

**Solution.** **Case 1** Since we have the pdf $f(x) = \frac{1}{2},\ -1 < x < 1$, then:

$$F(x) = \Pr(X \le x) = \int_{-1}^{x} f(\tau)d\tau = \begin{cases} 0, & x \le -1 \\ \frac{1}{2}(x+1), & -1 < x < 1 \\ 1, & x \ge 1 \end{cases}$$

Let $G(y)$ and $g(y)$ represent the cdf and the pdf of $Y$ defined on sample space $\mathcal{B} = \{y : 0 < y < 1\}$, and

$$\begin{aligned} G(y) &= \Pr(Y \le y) = \Pr(X^2 \le y) = \Pr(-\sqrt{y} \le X \le \sqrt{y}) = \Pr(X \le \sqrt{y}) - \Pr(X \le -\sqrt{y}) \\ &= F(\sqrt{y}) - F(-\sqrt{y}) = \frac{1}{2}(\sqrt{y}+1) - \frac{1}{2}(-\sqrt{y}+1) = \frac{1}{2}2\sqrt{y} \end{aligned}$$

$$\therefore G(y) = \begin{cases} 0, & y \le 0 \\ \sqrt{y}, & 0 < y < 1 \\ 1, & y \ge 1 \end{cases}$$

Then, the pdf $g(y) = G'(y) = \frac{1}{2\sqrt{y}},\ 0 < y < 1$.

**Case 2** The pdf of r.v $X$ $f(x) = \frac{1}{2},\ -1 < x < 1$. Let $G(y)$ and $g(y)$ be the cdf and pdf of $Y = X^2$ with sample space $\mathcal{B} = \{y : 0 < y < 1\}$. Then:

$$G(y) = \Pr(Y \leq y) = \Pr(X^2 \leq y) = \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) = \int\limits_{-\sqrt{y}}^{\sqrt{y}} f(x)dx = \int\limits_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}dx = \sqrt{y}$$

$$\therefore G(y) = \begin{cases} 0, & y \leq 0 \\ \sqrt{y}, & 0 < y < 1 \\ 1, & y \geq 1 \end{cases}$$

◀

**EXAMPLE 2.25**

The function $f(x) = \frac{x}{6},\ x = 1, 2, 3$ is a pdf of r.v $X$ that is defined on sample space $\mathcal{A} = \{1, 2, 3\}$. Find the cdf of r.v $Y = 2X + 1$.

**Solution.** The cdf of r.v $X$ is:

$$F(x) = \Pr(X \leq x) = \sum_{\tau=1}^{x} f(\tau) = \frac{1}{6}\sum_{\tau=1}^{x} \tau = \frac{1}{6}(1 + 2 + \cdots + x)$$

$$\therefore F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{12}x(x+1), & 1 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

Let $G(y)$ is the cdf of r.v $Y = 2X + 1$ defined on sample space $\mathcal{B} = \{3.5.7\}$, then:

$$G(y) = \Pr(Y \leq y) = \Pr(2X + 1 \leq y) = \Pr\left(X \leq \frac{1}{2}(y-1)\right) = F\left(\frac{y-1}{2}\right)$$
$$= \frac{1}{12}\left(\frac{y-1}{2}\right)\left(\frac{y-1}{2} + 1\right)$$

$$\therefore G(y) = \begin{cases} 0, & y < 3 \\ \frac{1}{48}\left(y^2 - 1\right), & 3 \leq y < 7 \\ 1, & y \geq 7 \end{cases}$$

◀

**EXAMPLE 2.26**

Find the distribution of the r.v $Y = -\ln X$, where $X$ is defined on sample space $\mathcal{A} = \{x : 0 < x < 1\}$ with pdf $f(x) = 1,\ x \in \mathcal{A}$.

**Solution.** In order to evaluate the cdf of r.v $X$:

$$F(x) = \Pr(X \leq x) = \int_0^t f(\tau)d\tau = \int_0^x d\tau = \tau\Big|_0^x$$

$$\therefore F(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x < 1 \\ 1, & x \ge 1 \end{cases}$$

Assume $G(y)$ and $g(y)$ represent the cdf and pdf of r.v $Y = -\ln X$ defined on sample space $\mathcal{B} = \{y : 0 < y < \infty\}$,

$$\begin{aligned} G(y) &= \Pr(Y \le y) = \Pr(-\ln X \le y) = \Pr(\ln X \ge y) = \Pr(X \ge e^{-y}) \\ &= 1 - \Pr(X \le e^{-y}) = 1 - F\left(e^{-y}\right) \end{aligned}$$

$$\therefore G(y) = \begin{cases} 0, & y \le 0 \\ 1 - e^{-y}, & 0 < y < \infty \\ 1, & y = \infty \end{cases}$$

This leads to the pdf of r.v $Y$, $g(y) = G'(y) = e^{-y}, y \in \mathcal{B}$ ◀

## 2.4 Mathematical Expectation

We have observed that the probability distribution for a random variable is a theoretical model for the empirical distribution of data associated with a real population. If the model is an accurate representation of nature, the theoretical and empirical distributions are equivalent. Consequently, we attempt to find the mean and the variance for a random variable and thereby to acquire numerical descriptive measures, parameters, for the probability distribution.

***Definition***: Let $X$ be a r.v with probability function $f(x)$ and $u(X)$ be a real-valued function of $X$. Then the expected value of $u(X)$ is given by:

$$E[u(X)] = \begin{cases} \sum\limits_{x=-\infty}^{\infty} u(x)f(x), & \text{discrete} \\ \\ \int\limits_{x=-\infty}^{\infty} u(x)f(x)dx, & \text{continuous} \end{cases}$$

**EXAMPLE 2.27**

Let the random variable $X$ has a pdf $f(x) = \frac{x}{6},\ x = 1, 2, 3$. Find $E[X],\ E[X^2],\ E[X^3],\ E[(X-1)^3]$.

**Solution.**

-
$$E[X] = \sum_x xf(x) = \sum_{x=1}^{3} x\frac{x}{6} = \frac{1}{6}(1^2 + 2^2 + 3^2) = \frac{14}{6} = \frac{7}{3}$$

-
$$E[X^2] = \sum_x x^2 f(x) = \frac{1}{6}\sum_{x=1}^{3} x^2 = \frac{1}{6}(1^3 + 2^3 + 3^3) = \frac{36}{6} = 6$$

-
$$E[X^3] = \sum_x x^3 f(x) = \frac{1}{6}\sum_{x=1}^{3} x^3 = \frac{1}{6}(1^4 + 2^4 + 3^4) = \frac{36}{6} = \frac{98}{6} = \frac{49}{3}$$

-
$$E[(X-1)^3] = \sum_x (x-1)^3 f(x) = \frac{1}{6}\sum_{x=1}^{3} x(x-1)^3 = \frac{1}{6}(0 + 2 + 24) = \frac{26}{6} = \frac{13}{2}$$

◀

**EXAMPLE 2.28**

Let the random variable $X$ has a pdf $f(x) = \frac{1}{18}(X+2),\ -2 < X < 4$. Find $E[3X],\ E[(X-2)^3]$.

**Solution.**

-

$$E[3X] = \int_x 3xf(x) = \int_{x=-2}^{4} 3x\frac{1}{18}(x+2)dx = \frac{1}{6}\int_{-2}^{4}(x^2+2x)dx = \frac{1}{6}\left[\frac{x^3}{3}+x^2\right]_{-2}^{4}$$
$$= \frac{1}{6}\left[\left(\frac{64}{3}+16\right) - \left(-\frac{8}{3}+4\right)\right] = \frac{1}{6}\left[\frac{64}{3}+16 = \frac{8}{3}-4\right] = 6$$

-

$$E[(X+2)^3] = \int_{x=-2}^{4}(x+2)^3\frac{1}{18}(x+2)dx = \frac{1}{18}\int_{-2}^{4}(x+2)^4dx = \frac{1}{18}\frac{(x+2)^5}{5}\Big|_{-2}^{4}$$
$$= \frac{1}{90}\left[6^5 - 0\right] = 86.4$$

◄

***Notes***:

- If $c$ is constant, then $E[c] = c$.
- If $c$ is constant and $u$ is a function of $X$, then $E[cu(X)] = cE[u(X)]$.
- If $c_1, c_2, \ldots, c_n$ are constants, and $u_1, u_2, \ldots, u_n$ are functions, then $\sum_{i=1}^{n} E[c_i u_i] = \sum_{i=1}^{n} c_i E[u_i]$.

**EXAMPLE 2.29**

Let the r.v $X$ has a pdf $f(x) = 2(1-x),\ 0 < x < 1$. Find $E[X]$, $E[X^2]$ and $E[6X + 3X^2 - 4]$.

**Solution.** -

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = 2\int_0^1 x(x-1)dx = 2\left[\frac{x^2}{2}-\frac{x^3}{3}\right]_0^1 = \frac{1}{3}.$$

-

$$E[X^2] = \int_{-\infty}^{\infty} x^2f(x)dx = 2\int_0^1 x^2(x-1)dx = 2\left[\frac{x^3}{3}-\frac{x^4}{4}\right]_0^1 = \frac{1}{6}.$$

-

$$E[6X+3X^2-4] = 6E[X] + 3E[X^2] - 4 = 6\frac{1}{3} + 3\frac{1}{6} - 4 = \frac{-3}{2}.$$

◄

### 2.4.1 Some Special Mathematical Expectations

In this section, we will introduce some special mathematical expectation which are most common in the use of statistical problems.

1. The **Mean** (or the expected vale) of a r.v $X$ is the mathematical expectation $E[X]$ and denoted by $\mu$.

2. If $X$ is a r.v with mean $E(X) = \mu$, the **Variance** of r.v $X$, denoted by $\sigma^2$ or $Var(X)$, is defined to be the expected value of $(X\mu)^2$. That is,
$$\sigma^2 = Var(x) = E[(X - \mu)^2].$$
The standard deviation of $X$ is the positive square root of $\sigma^2$.

   ***Properties of Variance***

   **i** The variance $\sigma^2 = E[(x-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - \mu^2$.

   **ii** If $c$ is a constant, then $Var(c) = 0$.

   **iii** If $c$ is a constant and $X$ is a r.v, then $Var(cX) = c^2 Var(X)$

3. The mathematical expectation $\mu'_r = E[X^r]$ is called the $r^{th}$ moment about the origin.

4. The mathematical expectation $\mu_r = E[(X - \mu)^r]$ is called the $r^{th}$ moment about the mean.

5. The **Moment Generating Function** (mgf) of a r.v $X$ is the expectation of $E[e^{tX}]$ (if exist), and denoted by $M(t)$. The reason of the function $M(t)$ is called mgf can be explained by the following statement. We have
$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \dots.$$
Then, the expectation
$$\begin{aligned} E[e^{tx}] &= \sum_x e^{tx} f(x) = \sum_x \left[1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \dots\right] f(x) \\ &= \sum_x f(x) + t\sum_x x f(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \frac{t^3}{3!}\sum_x x^3 f(x) + \dots \\ &= 1 + t\mu'_1 + \frac{t^2}{2!}\mu'_2 + \frac{t^3}{3!}\mu'_3 + \dots. \end{aligned}$$
This argument involves an interchange of summations, which is justifiable if $M(t)$ exists. Thus, $E[e^{tX}]$ is a function of all the moments $\mu'_k$ about the origin, for $k = 1, 2, 3, \dots$. In particular, $\mu'_k$ is the coefficient of $t^k/k!$ in the series expansion of $M(t)$.

   ***Notes***:

   **i** $M(0) = 1$.

   **ii** If we can find $E[e^{tX}]$, we can find any of the moments for $X$. If $M(t)$ exist, then for any positive integer $k$,
$$\left.\frac{d^k M(t)}{dt^k}\right|_{t=0} = M^{(k)}(0) = \mu'_k.$$
In other words, if you find the $k^{th}$ derivative of $M(t)$ with respect to $t$ and then set $t = 0$, the result will be $\mu'_k$.

**iii** If we set $\xi(t) = \ln M(t)$, then

$$\xi'(t) = \frac{M'(t)}{M(t)} \Rightarrow \xi'(0) = \frac{M'(0)}{M(0)} = \frac{E[X]}{1} = \mu$$
$$\xi''(t) = \frac{M(t)M''(t) - M'(t)M'(t)}{[M(t)]^2} \Rightarrow \xi''(0) = \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2}$$
$$\Rightarrow \xi''(0) = \frac{E[X^2] - \mu^2}{1^2} = \sigma^2.$$

**EXAMPLE 2.30**

The probability distribution of a r.v $Y$ is given in the following table. Find the mean, variance and standard deviation of $Y$.

Probability distribution of $Y$

| y | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| p(y) | 1/8 | 1/4 | 3/8 | 1/4 |

**Solution.** By definition:

$$\mu = E[Y] = \sum_{y=0}^{3} yp(y) = 0(1/8) + 1(1/4) + 2(3/8) + 3(1/4) = 1.75$$

$$\sigma^2 = E[(Y-\mu)^2] = \sum_{y=0}^{3}(y-\mu)^2 p(y) = (0-1.75)^2(1/8)+(1-1.75)^2(1/4)+(2-1.75)^2(3/8)+(3-1.75)^2(1/4) = 0.9375$$

or

$$E[Y^2] = \sum_{y=0}^{3} y^2 p(y) = (0)^2(1/8) + (1)^2(1/4) + (2)^2(3/8) + (3)^2(1/4) = 4$$

$$\therefore \sigma^2 = E[Y^2] - \mu^2 = 4 - (1.75)^2 = 0.9275$$

and then

$$\sigma = +\sqrt{\sigma^2} = \sqrt{0.9375} = 0.97$$

◀

**EXAMPLE 2.31**

The manager of an industrial plant is planning to buy a new machine of either type A or type B. If $t$ denotes the number of hours of daily operation, the number of daily repairs $Y_1$ required to maintain a machine of type A is a random variable with mean and variance both equal to $0.10t$. The number of daily repairs $Y_2$ for a machine of type B is a random variable with mean and variance both equal to $0.12t$. The daily cost of operating A is $C_A(t) = 10t + 30Y_1^2$; for B it is $C_B(t) = 8t + 30Y_2^2$. Assume that the repairs take negligible time and that each night the machines are tuned so that they operate essentially like new machines at the start of the next day. Which machine minimizes the expected daily cost if a workday consists of (a) 10 hours and (b) 20 hours?

**Solution.** The expected daily cost for A is

$$\begin{aligned}E[C_A(t)] &= E[10t + 30Y_1^2] = 10t + 30E[Y_1^2]\\ &= 10t + 30\{Var(Y_1) + (E[Y_1])^2\} = 10t + 30[0.10t + (0.10t)^2]\\ &= 13t + 0.3t^2.\end{aligned}$$

In this calculation, we used the known values for $Var(Y_1)$ and $E(Y_1)$ and the fact that $Var(Y_1) = E(Y_1^2) - [E(Y_1)]^2$ to obtain that $E(Y_1^2) = Var(Y_1) + [E(Y_1)]^2 = 0.10t + (0.10t)^2$. Similarly,

$$\begin{aligned}E[C_B(t)] &= E[8t + 30Y_2^2] = 8t + 30E[Y_2^2]\\ &= 8t + 30V(Y_2) + (E[Y_2])^2 = 8t + 30[0.12t + (0.12t)^2]\\ &= 11.6t + 0.432t^2.\end{aligned}$$

Thus, for scenario (a) where $t = 10$,

$$E[C_A(10)] = 160 \text{ and } E[C_B(10)] = 159.2,$$

which results in the choice of machine B.

For scenario (b), $t = 20$ and

$$E[C_A(20)] = 380 \text{ and } E[C_B(20)] = 404.8,$$

resulting in the choice of machine A.

In conclusion, machines of type B are more economical for short time periods because of their smaller hourly operating cost. For long time periods, however, machines of type A are more economical because they tend to be repaired less frequently. ◀

**EXAMPLE 2.32**

A retailer for a petroleum product sells a random amount $X$ each day. Suppose that $X$ , measured in thousands of gallons, has the probability density function $f(x) = \frac{3}{8}x^2,\ 0 \le x \le 2$. The retailer's profit turns out to be 100$for each 1000 gallons sold if $X \le 1$ and 40$extra per 1000 gallons if $X > 1$. Find the retailer's expected profit for any given day.

**Solution.** Let $p(x)$ denote the retailer's daily profit. Then

$$p(x) = \begin{cases} 100X, & 0 \le X \le 1,\\ 140X, & 1 \le X \le 2.\end{cases}$$

We want to find expected profit; by definition, the expectation is:

$$\begin{aligned}E[p(X)] &= \int_x p(x)f(x)dx = \int_0^1 100x\left[\frac{3}{8}x^2\right]dx + \int_1^2 140x\left[\frac{3}{8}x^2\right]dx\\ &= \left[\frac{300}{(8)(4)}x^4\right]_0^1 + \left[\frac{420}{(8)(4)}x^4\right]_1^2 = 206.25\end{aligned}$$

Thus, the retailer can expect a profit of 206.25$on the daily sale of this particular product. ◀

**EXAMPLE 2.33**

Find the mean and variance, if exist, for each of the following distributions, with pdf's:

1. $f(x) = \frac{x}{15}, \quad x = 1, 2, 3, 4, 5.$
2. $f(x) = \frac{1}{2}(x+1), \quad -1 < x < 1.$
3. $f(x) = x^{-2}, \quad 1 < x < \infty.$
4. $f(x) = e^{-x}, \quad 0 < x < \infty.$
5. $f(x) = \frac{3}{8}x^2, \ 0 \leq x \leq 2$

**Solution.**

1.

$$\mu = E[X] = \sum_x x f(x) = \sum_{x=1}^{5} x\frac{x}{15} = \frac{1}{15}(1^2 + 2^2 + 3^2 + 4^2 + 5^2) = \frac{11}{3}.$$

$$E[X^2] = \sum_x x^2 f(x) = \sum_{x=1}^{5} x^2\frac{x}{15} = \frac{1}{15}(1^3 + 2^3 + 3^3 + 4^3 + 5^3) = 15.$$

$$\sigma^2 = Var(X) = E[X^2] - \mu^2 = 15 - \left(\frac{11}{3}\right)^2 = 15 - \frac{121}{9} = \frac{14}{9}.$$

2.

$$\mu = E[X] = \int_x x f(x)dx = \int_{-1}^{1} x\frac{1}{2}(x+1)dx = \frac{1}{2}\left[\frac{x^3}{3} + \frac{x^2}{2}\right]_{-1}^{1} = \frac{1}{2}\left[\left(\frac{1}{3} + \frac{1}{2}\right) - \left(\frac{-1}{3} + \frac{1}{2}\right)\right] = \frac{1}{3}.$$

$$E[X^2] = \int_x x^2 f(x)dx = \int_{-1}^{1} x\frac{1}{2}(x+1)dx = \frac{1}{2}\left[\frac{x^4}{4} + \frac{x^3}{3}\right]_{-1}^{1} = \frac{1}{3}.$$

$$\sigma^2 = Var(X) = E[X^2] - \mu^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}.$$

3.

$$\mu = E[X] = \int_x x f(x)dx = \int_1^{\infty} x x^{-2}dx = \int_1^{\infty} \frac{1}{x}dx = \ln x\Big|_1^{\infty} = \ln\infty - \ln 1 = \infty - 0 = \infty.$$

Therefore, the mean $\mu$ does not exist, hence the variance $\sigma^2$ does not exist neither.

4.

$$\mu = E[X] = \int_x x f(x)dx = \int_0^{\infty} xe^{-x}dx = -xe^{-x} - e^{-x}\Big|_0^{\infty} = (0-0) - (0-1) = 1.$$

$$E[X^2] = \int_x x^2 f(x)dx = \int_0^{\infty} xe^{-x}dx = -x^2e^{-x} - 2xe^{-x} - 2e^{-x}\Big|_0^{\infty} = 2.$$

$$\sigma^2 = E[X^2] - \mu^2 = 2 - (1)^2 = 1.$$

5.

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x)dx = \int_0^2 x\frac{3}{8}x^2dx = \frac{3}{8}\frac{x^4}{4}\Big|_0^2 = 1.5.$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^2 x^2 \frac{3}{8}x^2 dx = \frac{3}{8}\frac{x^5}{5}\bigg|_0^2 = 2.4$$

$$\sigma^2 = Var(X) = E[X^2] - (E[X])^2 = 2.4 - (1.5)^2 = 0.15$$

◄

**EXAMPLE 2.34**

Let the r.v $X$ has a pdf $f(x) = \left(\frac{1}{2}\right)^x,\ x = 1, 2, 3, \ldots$, then:

1. Find the mgf of $X$.
2. Evaluate the mean and variance of $X$ using the mgf.

**Solution.**

1. The mgf $M(t)$ is the expectation of the function $e^{tX}$, then

$$M(t) = E[e^{tX}] = \sum_x e^{tx} f(x) = \sum_{x=1}^{\infty} e^{tx}\left(\frac{1}{2}\right)^x = \sum_{x=1}^{\infty}\left(\frac{e^t}{2}\right)^x = \frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + \left(\frac{e^t}{2}\right)^3 + \cdots = \frac{(e^t/2)}{1-(e^t/2)}$$

$$\therefore M(t) = \frac{e^t}{2-e^t}, \quad t \neq \ln 2.$$

2. Check $M(0) = \frac{e^0}{2-e^0} = \frac{1}{2-1} = 1$. now

$$M'(t) = \frac{(2-e^t)e^t - e^t(-e^t)}{(2-e^t)^2} = \frac{2e^t - e^{2t} + e^{2t}}{(2-e^t)^2} = \frac{2e^t}{(2-e^t)^2}$$

$$M'(0) = \frac{2e^0}{(2-e^0)^2} = \frac{2}{(2-1)^2} = 2 = \mu$$

and that

$$M''(t) = \frac{(2-e^t)^2 2e^t - 2e^t 2(2-e^t)(-e^t)}{(2-e^t)^4} = \frac{(2-e^t)2e^t + 4e^2t}{(2-e^t)^3}$$

$$M''(0) = \frac{(2-1)2+4}{(2-1)^3} = 6 = E[X^2]$$

then

$$\sigma^2 = E[X^2] - \mu^2 = 6 - 4 = 2.$$

**<u>or</u>**, we consider the function $\xi(t) = \ln M(t) = t - \ln(2-e^t)$, then

$$\xi'(t) = 1 + \frac{e^t}{2-e^t} \Rightarrow \xi'(0) = 1 + \frac{1}{2-1} = 2 = \mu$$

and

$$\xi''(t) = \frac{(2-e^t)e^t - e^t(-e^t)}{(2-e^t)^2} = \frac{2e^t}{(2-e^t)^2} \Rightarrow \xi''(0) = \frac{2}{(2-1)^2} = 2 = \sigma^2.$$

◄

**EXAMPLE 2.35**

Find the mgf of a r.v $X$ that has a pdf $f(x) = xe^{-x},\ 0 < x < \infty$, then evaluate the mean and variance of $X$.

**Solution.** The mgf od r.v $X$ is the expectation of $e^{tX}$,

$$M(t) = E[e^{tX}] = \int_x e^{tx} f(x)dx = \int_0^\infty e^{tx} x e^{-x} dx = \int_0^\infty x e^{-(1-t)x} dx = \left[ -\frac{xe^{-(1-t)x}}{1-t} - \frac{e^{-(1-t)x}}{(1-t)^2} \right]_0^\infty$$

Therefore,

$$M(t) = \frac{1}{(1-t)^2},\ t < 1.$$

In order to evaluate $\mu$ and $\sigma^2$, we have the mgf $M(t) = (1-t)^{-2}$,

$$M'(t) = 2(1-t)^{-3} \Rightarrow M'(0) = 2 = \mu$$

$$M''(t) = 6(1-t)^{-4} \Rightarrow M''(0) = 6 = E[X^2]$$

$$\therefore \sigma^2 = E[X^2] - \mu^2 = 6 - 4 = 2$$

**or**, we consider the function $\xi(t) = \ln M(t) = -2\ln(1-t)$, then

$$\xi'(t) = \frac{2}{1-t} \Rightarrow \xi'(0) = 2 = \mu$$

$$\xi''(t) = 2(1-t)^{-2} \Rightarrow \xi''(0) = 2 = \sigma^2$$

◀

**EXAMPLE 2.36**

A manufacturing company ships its product in two different sizes of truck trailers. Each shipment is made in a trailer with dimensions 8 feet × 10 feet × 30 feet or 8 feet × 10 feet × 40 feet. If 30% of its shipments are made by using 30-foot trailers and 70% by using 40-foot trailers, find the mean volume shipped per trailer load. (Assume that the trailers are always full.)

**Solution.** Assume that the volume of the 30-foot trailers is $v_1$ and the 40-foot trailers is $v_2$, then:

$$v_1 = 8 \times 10 \times 30 = 2400 \text{ feet}^3.$$

$$v_2 = 8 \times 10 \times 40 = 3200 \text{ feet}^3.$$

since we have the probability of shipping throughout $v_1$ and $v_2$ are:

$$p(v_1) = 30\% = \frac{3}{10} \quad p(v_2) = 70\% = \frac{7}{10}.$$

Therefore, the expected shipping volume is:

$$E[V] = \sum_{i=1}^{2} v_i p(v_i) = 2400 \times \frac{3}{10} + 3200 \times \frac{7}{10} = 2990 \text{ feet}^3.$$

◀

**EXAMPLE 2.37**

In a gambling game a person draws a single card from an ordinary 52-card playing deck. A person is paid 15$for drawing a jack or a queen and 5$ for drawing a king or an ace. A person who draws any other card pays 4$. If a person plays this game, what is the expected gain?

**Solution.** Let the r.v $X$ represents the outcome of the draw. Then, the player gain could be represented as:

$$g = \begin{cases} 15, & x = J, Q \\ 5, & x = K, A \\ -4, & x = 2, 3, 4, 5, 6, 7, 8, 9, 10 \end{cases}$$

Since we have $\binom{52}{4} = \frac{1}{13}$ ways of drawing a card, then the probability of drawing any number or shape is equal to $\frac{1}{13}$, i.e:

Probability distribution of $X$

| x | 2 | 3 | ... | 10 | J | Q | K | A |
|---|---|---|---|---|---|---|---|---|
| p(x) | 1/13 | 1/13 | ... | 1/13 | 1/13 | 1/13 | 1/13 | 1/13 |

Then, the expected gain of the played is calculated by:

$$E[G] = \sum gp(x) = \left[9\left((-4) \times \frac{1}{13}\right) + 2\left((5) \times \frac{1}{13}\right) + 2\left((15) \times \frac{1}{13}\right)\right] = \frac{-36}{13} + \frac{10}{13} + \frac{30}{13} = \frac{4}{13} = 0.307.$$

◀

**EXAMPLE 2.38**

A builder of houses needs to order some supplies that have a waiting time $Y$ for delivery, with a continuous uniform distribution over the interval from 1 to 4 days $\left(p(y) = \frac{1}{3},\ 1 \le y \le 4\right)$. Because he can get by without them for 2 days, the cost of the delay is fixed at 100$ for any waiting time up to 2 days. After 2 days, however, the cost of the delay is 100$ + 20$ per day (prorated) for each additional day. That is, if the waiting time is 3.5 days, the cost of the delay is 100$ + 20$(1.5) = 130$. Find the expected value of the builder's cost due to waiting for supplies.

**Solution.** Assume that the cost of waiting the supplies $W_c$, and $Y$ is the r.v that represents the number of waiting days, then:

$$W_c = \begin{cases} 100, & 1 \le y \le 2 \\ 100 + 20(y-2), & 2 < y \le 4 \end{cases}$$

Therefore, expected value of the builder's cost due to waiting for supplies is

$$\begin{aligned} E[W_c] &= \int W_c p(y)dy = \int_1^2 100\frac{1}{3}dy + \int_2^4 (100 + 20(y-2))\frac{1}{3}dy \\ &= \frac{100}{3}y\Big|_1^2 + \frac{100}{3}y\Big|_2^4 + \frac{20}{3}\left[\frac{y^2}{2} - 2y\right]_2^4 = \frac{100}{3} + \frac{200}{3} + \frac{40}{3} = \frac{340}{3} = 113.33 \end{aligned}$$

◀

### 2.4.2 Tchebyshev's Inequality

In order to find the upper and lower bounds for certain probability, we will need to prove some theorems. These bounds are not necessarily close to the exact probability.

***Theorem*:** Let $u(X)$ be a non-negative function of a r.v $X$ whose pdf $f(x),\ -\infty < x < \infty$. If $E[u(X)]$ exist, then for all positive constant $c$,

$$\Pr[u(X) \geq c] \leq \frac{E[u(X)]}{c}.$$

***Theorem*:** *Tchebyshev Inequality*: Let $X$ be a r.v with mean $\mu$ and finite variance $\sigma^2$. Then, for any constant $k > 0$,

$$\Pr\big(|X-\mu| < k\sigma\big) \geq 1 - \frac{1}{k^2}, \text{ or } \Pr\big(|X-\mu| \geq k\sigma\big) \leq \frac{1}{k^2}$$

Two important aspects of this result should be pointed out. First, the result applies for any probability distribution. Second, the results of the theorem are very conservative in the sense that the actual probability that $X$ is in the interval $\mu \pm k\sigma$ usually exceeds the lower bound for the probability, $1 - 1/k^2$, by a considerable amount.

***Proof*:** Consider the previous theorem by taking $u(X) = (X-\mu)^2$ and $c^2 = k^2\sigma^2$, then

$$\Pr\big[(X-\mu)^2 \geq k^2\sigma^2\big] \leq \frac{E\big[(X-\mu)^2\big]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

Since $(x-\mu)^2 \geq k^2\sigma^2 \Leftrightarrow |x-\mu| \geq k\sigma$. It follows,

$$\Pr\big(|X-\mu| \geq k\sigma\big) \leq \frac{1}{k^2}$$

**EXAMPLE 2.39**

Let the r.v $X$ has a pdf $f(x) = \frac{1}{2\sqrt{3}},\ -\sqrt{3} < x < \sqrt{3}$. Find The exact value of $\Pr\big(|X-\mu| \geq \frac{3}{2}\sigma\big)$ and $\Pr\big(|X-\mu| \geq 2\sigma\big)$, and then compare those results with their upper bounds.

**Solution.** First of all, we need to find the mean $\mu$ and the variance $\sigma^2$. Then

$$\mu = E[X] = \int_x xf(x)dx = \frac{1}{2\sqrt{3}}\int_{-\sqrt{3}}^{\sqrt{3}} xdx = 0.$$

$$E[X^2] = \int_x x^2 f(x)dx = \frac{1}{2\sqrt{3}}\int_{-\sqrt{3}}^{\sqrt{3}} x^2dx = 1.$$

$$\sigma^2 = E[X^2] - \mu^2 = 1 \Rightarrow \sigma = 1.$$

The exact value of probability

$$\begin{aligned}\Pr\big(|X-\mu| \geq \frac{3}{2}\sigma\big) &= \Pr\big(|X| \geq \frac{3}{2}\big) = 1 - \Pr\big(|X| < \frac{3}{2}\big) = 1 - \Pr\big(-\frac{3}{2} < X < \frac{3}{2}\big)\\ &= 1 - \int_{-3/2}^{3/2} f(x)dx = 1 - \frac{1}{2\sqrt{3}}\int_{-3/2}^{3/2} dx = 1 - \frac{\sqrt{3}}{2} = 0.134\end{aligned}$$

To compare with the probability upper bound, we will use Tchebyshev inequality, to find this upper bound for probability $\Pr\big(|X-\mu| \geq \frac{3}{2}\sigma\big)$

$$\Pr\big(|X| \geq \frac{3}{2}\big) \leq \left(\frac{3}{2}\right)^2 = \frac{4}{9} = 0.44.$$

It is clear that the exact probability $(0.134)$ is less than the upper bound $(0.44)$.

For the next part, we do the same. The exact value of probability

$$\begin{aligned}\Pr\big(|X-\mu|\ge 2\sigma\big) &= \Pr\big(|X|\ge 2\big) = 1-\Pr\big(|X|<2\big) = 1-\Pr\big(-2<X<2\big)\\ &= 1-\int_{-2}^{2} f(x)dx = 1-\left[\int_{-2}^{\sqrt{3}} f(x)dx+\int_{-\sqrt{3}}^{\sqrt{3}} f(x)dx+\int_{\sqrt{3}}^{2} f(x)dx\right] = 1-(0+1+0)=0\end{aligned}$$

To compare with the probability upper bound, we will use Tchebyshev inequality, to find this upper bound for probability $\Pr\big(|X-\mu|\ge 2\sigma\big)$

$$\Pr\big(|X|\ge 2\big) \le \frac{1}{2^2} = 0.25.$$

It is clear that the exact probability (0) is less than the upper bound (0.25). ◀

***Note:*** We may have the mean $\mu$ and variance $\sigma^2$ for a distribution whose pdf is not available for some reason. In this case, to find a certain probability, we use Tchebyshev inequality to find the upper or lower bound for this probability.

**EXAMPLE 2.40**

Let the r.v $X$ has mean $\mu=3$ and variance $\sigma^2=4$. Use Tchebyshev inequality to determine a lower bound for $\Pr(-2<X<8)$.

**Solution.** To use the Tchebyshev inequality, we need to get to the form $\Pr\big[|X-\mu|<k\sigma\big] = 1-\frac{1}{k^2}$. Then

$$\begin{aligned}\Pr\big(-2<X<8\big) &= \Pr\big(-2-3<X-\mu<8-3\big) = \Pr\big(-5<X-\mu<5\big) = \Pr\big(|X-\mu|<5\big)\\ &= \Pr\big(|X-\mu|<\frac{5}{2}\sigma\big) \ge 1-\left(\frac{2}{5}\right)^2 = 1-\frac{4}{25} = \frac{21}{25} = 0.85\end{aligned}$$

◀

**EXAMPLE 2.41**

The number of customers per day at a sales counter, $Y$ , has been observed for a long period of time and found to have mean 20 and standard deviation 2. The probability distribution of $Y$ is not known. What can be said about the probability that, tomorrow, $Y$ will be greater than 16 but less than 24?

**Solution.** We want to find $\Pr(16<Y<24)$. From Tchebyshev inequality we know, for any $k\ge 0$, $\Pr(|Y-\mu|<k\sigma)\ge 1-1/k^2$, or

$$\Pr\big[(\mu-k\sigma)<Y<(\mu+k\sigma)\big] \ge 1-\frac{1}{k^2}.$$

Because $\mu=20$ and $\sigma=2$, it follows that $\mu-k\sigma=16$ and $\mu+k\sigma=24$ if $k=2$. Thus

$$\Pr\big(16<Y<24\big) = \Pr\big(\mu-2\sigma<Y<\mu+2\sigma\big) \ge 1-\frac{1}{2^2} = \frac{3}{4}.$$

In other words, tomorrow's customer total will be between 16 and 24 with a fairly high probability (at least $3/4$).

Notice that if $\sigma$ were 1, $k$ would be 4, and

$$\Pr\big(16<Y<24\big) = \Pr\big(\mu-4\sigma<Y<\mu+4\sigma\big) \ge 1-\frac{1}{4^2} = \frac{15}{16}.$$

Thus, the value of $\sigma$ has considerable effect on probabilities associated with intervals. ◀

**EXAMPLE 2.42**

Suppose that experience has shown that the length of time $T$ (in minutes) required to conduct a periodic maintenance check on a dictating machine follows a gamma distribution with mean $\mu = 6.2$ and variance $\sigma^2 = 12.4$. A new maintenance worker takes 22.5 minutes to check the machine. Does this length of time to perform a maintenance check disagree with prior experience?

**Solution.** We know that $\mu = 6.2$ and $\sigma^2 = 12.4 \Rightarrow \sigma = \sqrt{12.4} = 3.52$. We need to evaluate $\Pr(T \geq 22.5)$, then

$$\Pr\big(T - \mu \geq 22.5 - \mu\big)$$

. Notice that $t = 22.5$ minutes exceeds the mean $\mu = 6.2$ minutes by 16.3 minutes, or $k = 16.3/3.52 = 4.63$ standard deviations. Then from Tchebysheff's theorem,

$$\Pr\big(|T - 6.2| \geq 16.3\big) = \Pr\big(|T - \mu| \geq 4.63\sigma\big) \leq \frac{1}{(4.63)^2} = 0.0466.$$

This probability is based on the assumption that the distribution of maintenance times has not changed from prior experience. Then, observing that $\Pr(T \geq 22.5)$ is small, we must conclude either that our new maintenance worker has generated by chance a lengthy maintenance time that occurs with low probability or that the new worker is somewhat slower than preceding ones. Considering the low probability for $\Pr(T \geq 22.5)$, we favour the latter view. ◄

## CHAPTER 3

# SOME SPECIAL MATHEMATICAL DISTRIBUTIONS

As stated in Chapter **??**, a random variable is a real-valued function defined over a sample space. Consequently, a random variable can be used to identify numerical events that are of interest in an experiment. For any r.v, we can define many distribution functions in order to be able to calculate probabilities for certain events. In this chapter, we will introduce some special probability distribution for some discrete and continuous r.v's.

### 3.1 Discrete Distributions

In this section, some of the most important and popular distributions for discrete r.v are presented. Both the pdf and cdf is derived and some important properties and mathematical expectation for these distribution are obtained.

#### 3.1.1 Binomial Distribution

Some experiments consist of the observation of a sequence of identical and independent trials, each of which can result in one of two outcomes.For instance, each item leaving a manufacturing production line is either defective or non-defective. Each shot in a sequence of firings at a target can result in a hit or a miss, and each of $n$ persons questioned prior to a local election either favors candidate Jones or does not. In this section we are concerned with experiments, known as binomial experiments, that exhibit the following characteristics.

***Definition*:** A binomial experiment possesses the following properties:

1. The experiment consists of a fixed number, $n$, of identical trials.

2. Each trial results in one of two outcomes: success, $S$, or failure, $F$.



Please enter `\offprintinfo{(Title, Edition)}{(Author)}` `at the beginning of your document.`

3. The probability of success on a single trial is equal to some value $p$ and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.

4. The trials are independent.

5. The random variable of interest, $X$, the number of successes observed during the $n$ trials.

Determining whether a particular experiment is a binomial experiment requires examining the experiment for each of the characteristics just listed. Notice that the random variable of interest is the number of successes observed in the n trials. It is important to realize that a success is not necessarily "good" in the everyday sense of the word. In our discussions, success is merely a name for one of the two possible outcomes on a single trial of an experiment.

***Definition*:** A random variable $X$ is said to have a binomial distribution based on $n$ trials with success probability $p$ if and only if

$$f(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n \text{ and } 0 \le p \le 1,$$

and is denoted by $X \sim B(n, p)$.

The term binomial experiment derives from the fact each trial results in one of two possible outcomes and that the probabilities $p(y), y = 0, 1, 2, \ldots, n$, are terms of the binomial expansion

$$(p+q)^n = \binom{n}{0} p^n q^0 + \binom{n}{1} p^{n-1} q^1 + \binom{n}{2} p^{n-2} q^2 + \cdots + \binom{n}{n} p^0 q^n.$$

Now, in order to verify that $f(x)$ is a valid pdf, one can easily prove that

1. $f(x) > 0, \ \forall x \in \mathcal{A} = \{x : x = 0, 1, 2, \ldots, n\}$.

2. Since $f(x)$ satisfies the binomial expansion, and that $p + q = 1$, then

$$\sum_x f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1^n = 1.$$

The binomial probability distribution has many applications because the binomial experiment occurs in sampling for defectives in industrial quality control, in the sampling of consumer preference or voting populations, and in many other physical situations. We will illustrate with a few examples.

**EXAMPLE 3.1**

Suppose that a lot of 5000 electrical fuses contains 5% defectives. If a sample of 5 fuses is tested, find the probability of observing at least one defective.

**Solution.** It is reasonable to assume that $X$, the number of defectives observed, has an approximate binomial distribution with $p = 0.5$, and $q = 0.95$. Thus,

$$\begin{aligned}\Pr(\text{at least one defective}) &= 1 - f(0) = 1 - \binom{5}{0} p^0 q^5 \\ &= 1 - (0.95)^5 = 1 - 0.774 = 0.226\end{aligned}$$

Notice that there is a fairly large chance of seeing at least one defective, even though the sample is quite small. ◀

***The Cumulative Distribution Function*:** The cdf of $X$ that has a binomial pdf is defined as:

$$F(x) = \Pr(X \le x) = \sum_{\tau=0}^{x} f(\tau) = \sum_{\tau=0}^{x} \binom{n}{\tau} p^\tau q^{n-\tau}.$$

In practice, it is not easy or convenient to use the above form of the cfd to calculate the probability at certain point $F(x)$. Instead, we use Table 1 (p: 839-841)

**EXAMPLE 3.2**

The large lot of electrical fuses of the last example is supposed to contain only 5% defectives. If $n = 20$ fuses are randomly sampled from this lot, find the probability that at least four defectives will be observed.

**Solution.** Letting $X$ denote the number of defectives in the sample, we assume the binomial model for $X$ , with $p = 0.05$. Thus,

$$\Pr(X \geq 4) = 1 - \Pr(X \leq 3),$$

and using Table 1, we obtain

$$\Pr(X \leq 3) = \sum_{x=0}^{3} f(x) = 0.984$$

The value 0.984 is found in the table labelled $n = 20$ in Table 1. Then, the probability of getting at least 4 defective fuses is

$$\Pr(X \geq 4) = 1 - 0.984 = 0.016.$$

This probability is quite small. If we did indeed observe more than three defectives out of 20 fuses, we might suspect that the reported 5% defect rate is erroneous. ◀

***The Moment Generating Function*:** The mgf of $X$ could be evaluated as:

$$\begin{aligned} M(t) = E\left[e^{tX}\right] &= \sum_{x} e^{tx} f(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(pe^t\right)^x q^{n-x} \\ &= \left(pe^t + q\right)^n, \quad -\infty < t < \infty \end{aligned}$$

***Mean and Variance*:** Let $X$ be a binomial random variable based on $n$ trials and success probability $p$. Then

$$\mu = E[X] = np \text{ and } \sigma^2 = Var(X) = npq.$$

The derivation of the mean and variance could be done in three different ways; (1) by the direct approach of $E[X]$, (2) differentiating the mgf $M(t)$, or (3) the differentiation of the function $\ln(M(t))$. We will use the second way to evaluate the mean and variance as:

$$M'(t) = npe^t \left(pe^t + q\right)^{n-1} \Rightarrow M'(0) = np = E[X] = \mu.$$

$$M''(t) = np\left[(n-1)pe^2t \left(pe^t + q\right)^{n-2} + e^t \left(pe^t + q\right)^{n-1}\right]$$

$$M''(0) = np[(n-1)p + 1] = np(np - p + 1) = np(np + q) = n^2p^2 + npq = E[X^2]$$

$$\therefore Var(X) = \sigma^2 = E[X^2] - \mu^2 = n^2p^2 + npq - n^2p^2 = npq$$

**EXAMPLE 3.3**

Let the r.v $X \sim b(7, 0.5)$. Write down the pdf and mgf of $X$. Then find $\mu$ and $\sigma^2$. Evaluate $\Pr(X \leq 2)$, $\Pr(3 < X \leq 5)$, $\Pr(X = 5)$.

**Solution.**

- Since $X \sim b(7, 0.5)$, then the pdf of $X$

$$f(x) = \Pr(X = x) = \binom{7}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{7-x} = \binom{7}{x}\left(\frac{1}{2}\right)^7, \quad x = 0, 1, 2, \ldots, 7$$

- The mgf of $X$

$$M(t) = \left(\frac{1}{2}e^t + \frac{1}{2}\right)^7$$

- Mean and Variance: $\mu = np = \frac{7}{2}$ , and $\sigma^2 = npq = \frac{7}{4}$.
- $\Pr(X \leq 2) = 0.2266$
- $\Pr(3 < X \leq 5) = \Pr(X \leq 5) - \Pr(X < 3) = \Pr(X \leq 5) - \Pr(X \leq 2) = 0.9375 - 0.2266 = 0.7109$.
- $\Pr(X = 5) = \Pr(X \leq 5) - \Pr(X \leq 4) = 0.9375 - 0.7734 = 0.1641$.

◀

**EXAMPLE 3.4**

A die is tossed 5 times. What is the probability of obtaining exactly three two's?

**Solution.** The probability of getting two's from a die tossing is $p = \frac{1}{6}$. Then, the probability of getting anything else is $q = \frac{5}{6}$. If we assume that a r.v $X$ represents the number of two's in 5 tosses, then $X \sim b(5, \frac{1}{6})$ with pdf $f(x) = \binom{5}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{5-x}$ , $x = 0, 1, 2, 3, 4, 5$. Therefore:

$$\Pr(X = 3) = f(3) = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = 0.0322$$

. ◀

**EXAMPLE 3.5**

Let the r.v $X \sim b(7, p)$ and that $\Pr(X = 3) = \Pr(X = 4)$. Find $\mu$ and $\sigma^2$, and then evaluate $\Pr(X = 3)$.

**Solution.** Since $X$ has a binomial pdf, $f(x) = \Pr(X = x) = \binom{7}{x}p^x q^{7-x}$, $x = 0, 1, \ldots, 7$ and since we have $\Pr(X = 3) = \Pr(X = 4)$, then

$$f(3) = f(4) \Rightarrow \binom{7}{3}p^3q^4 = \binom{7}{4}p^4q^3 \Rightarrow p = q \Rightarrow p = 1 - p \Rightarrow p = \frac{1}{2}.$$

Therefore,

$$\mu = np = \frac{7}{2} \text{ and } \sigma^2 = npq = \frac{7}{4}.$$

Then,

$$\Pr(X = 3) = \Pr(x \leq 3) - \Pr(x \leq 2) = 0.5 - 0.2266 = 0.2734.$$

◀

### 3.1.2 The Geometric Distribution

The random variable with the geometric probability distribution is associated with an experiment that shares some of the characteristics of a binomial experiment. This experiment also involves identical and independent trials, each of which can result in one of two outcomes: success or failure. The probability of success is equal to $p$ and is constant from trial to trial. However, instead of the number of successes that occur in $n$ trials, the geometric random variable $X$ is the number of the trial on which the first success occurs. Thus, the experiment consists of a series of trials that concludes with the first success. Consequently, the experiment could end with the first trial if a success is observed on the very first trial, or the experiment could go on indefinitely.

***Definition*:** A random variable $X$ is said to have a geometric probability distribution if

$$f(x) = q^{x-1}p, \quad x = 1, 2, 3, \dots, \; 0 \le p \le 1,$$

or

$$f(x) = q^{x}p, \quad x = 0, 1, 2, \dots, \; 0 \le p \le 1,$$

We denote to $X$ as $X \sim Geo(p)$.

To prove that $f(x)$ is a valid pdf,

1. $f(x) > 0$, $\forall x \in \mathcal{A} = \{x : x = 1, 2, \dots\}$.
2. We need to evaluate $\sum_x f(x)$, then

$$\sum_{x=0}^{\infty} q^x p = p\left(1 + q + q^2 + q^3 + \dots\right) = p\frac{1}{1-q} = p\frac{1}{p} = 1$$

The geometric probability distribution is often used to model distributions of lengths of waiting times.

**EXAMPLE 3.6**

Suppose that the probability of engine malfunction during any one-hour period is $p = 0.02$. Find the probability that a given engine will survive two hours.

**Solution.** Letting $Y$ denote the number of one-hour intervals until the first malfunction, we have

$$\Pr(\text{survive two hours}) = \Pr(Y \ge 3) = \sum_{y=3}^{\infty} f(y).$$

Because $\sum_{y=1}^{\infty} f(y) = 1$,

$$\begin{aligned}\Pr(\text{survive two hours}) &= 1 - \sum_{y=1}^{2} f(y) \\ &= 1 - p - qp = 1 - 0.02 - (0.98)(0.02) = 0.9604.\end{aligned}$$

◀

***The Cumulative Distribution Function*:** The cdf of r.v $X$ which has a Geometric pdf is:

$$\begin{aligned}F(x) = \Pr(X \le x) &= \sum_{\tau=0}^{x} f(\tau) = \sum_{\tau=0}^{x} q^{\tau}p = p(1 + q + q^2 + q^3 + \dots + q^x) \\ &= \begin{cases} 0, & x < 0 \\ 1 - q^{x+1}, & 0 \le x < \infty \\ 1, & x = \infty \end{cases}\end{aligned}$$

***The Moment Generating Function*:** We can obtain the mgf of r.v $X$ by evaluating the expectation $E\left[e^{tX}\right]$, that is

$$M(t) = E\left[e^{tX}\right] = \sum_x e^{tx} f(x) = \sum_{x=0}^{\infty} e^{tX} q^x p = p\left[1 + (qe^t) + (qe^t)^2 + \dots\right]$$
$$= \frac{p}{1 - qe^t}, \quad t \neq \ln\left(\frac{1}{q}\right).$$

***Mean and Variance*:** If $X$ is a r.v with a geometric distribution, then we define the mean $\mu$ and variance $\sigma^2$ as

$$\mu = E[X] = \frac{q}{p} \text{ and } \sigma^2 = Var(X) = \frac{1-p}{p^2}.$$

As it was mention previously, we can prove that in three different ways. This time, we will use the property of $\psi(t) = \ln(M(t))$. Hence $\psi(t) = \ln M(t) = \ln p - \ln(1 - qe^t)$, and the derivative of of $\psi$ gives

$$\psi'(t) = \frac{qe^t}{1 - qe^t} \Rightarrow \psi'(0) = \frac{q}{1-q} = \frac{q}{p} = \mu$$

and,

$$\psi''(t) = \frac{(1-qe^t)qe^t - (qe^t)(qe^t)}{(1-qe^t)^2} \Rightarrow \psi''(0) = \frac{(1-q)(q) + q^2}{(1-q)^2} = \frac{q}{p^2} = \sigma^2$$

**EXAMPLE 3.7**

Let the r.v $X \sim Geo(\frac{2}{3})$. Answer the following:

1. Write down the pdf, cdf and mgf of $X$.
2. Find $\mu$ and $\sigma^2$.
3. Evaluate $\Pr(X \geq 3)$ by using both the pdf and cdf.
4. Evaluate $\Pr(X \geq 5 | X \geq 2)$.

**Solution.**

1. ▪ pdf: $f(x) = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^x, \;\; x = 0, 1, 2, \dots.$
   ▪ cdf:
   $$F(x) = \begin{cases} 0, & x < 0 \\ 1 - \left(\frac{1}{3}\right)^{x+1}, & 0 \leq x < \infty \\ 1, & x = \infty \end{cases}$$
   ▪ mgf: $M(t) = \dfrac{2/3}{1 - \frac{1}{3}e^t} = \dfrac{2}{3 - e^t}$.
2. $\mu = \dfrac{q}{p} = \dfrac{1/3}{2/3} = \dfrac{1}{2}$ and $\sigma^2 = \dfrac{q}{p^2} = \dfrac{1/3}{4/9} = \dfrac{3}{2}$.
3. ▪ Using pdf: $\Pr(X \geq 3) = 1 - \Pr(X < 3) = 1 - \Pr(X \leq 2)$, then
   $$\Pr(X \geq 3) = 1 - \sum_{x=0}^{2} f(x) = 1 - [f(0) + f(1) + f(2)]1 - \frac{2}{3}\left[1 + \frac{1}{3} + \frac{1}{9}\right] = \frac{1}{27}.$$

- Using cdf: $\Pr(X \geq 3) = 1 - \Pr(X \leq 2) = 1 - F(2) = 1 - 1 + \left(\frac{1}{3}\right)^3 = \frac{1}{27}$.

4. According to the conditional probability,

$$\Pr(X \geq 5|X \geq 2) = \frac{\Pr(X \geq 5 \cap X \geq 2)}{\Pr(X \geq 2)} = \frac{\Pr(X \geq 5)}{\Pr(X \geq 5)} = \frac{1 - \Pr(X \leq 4)}{1 - \Pr(X \leq 1)} = \frac{(1/3)^5}{(1/3)^2} = \frac{1}{27}.$$

◀

### 3.1.3 Negative Binomial Distribution

A random variable with a negative binomial distribution originates from a context much like the one that yields the geometric distribution. Again, we focus on independent and identical trials, each of which results in one of two outcomes: success or failure. The probability $p$ of success stays the same from trial to trial. The geometric distribution handles the case where we are interested in the number of the trial on which the first success occurs. What if we are interested in knowing the number of the trial on which the second, third, or fourth success occurs? The distribution that applies to the random variable $X$ equal to the number of the trial on which the $r^th$ success occurs ($r = 2, 3, 4$, etc.) is the negative binomial distribution.

***Definition*:** A r.v $X$ is said to have negative binomial distribution, denoted by $X \sim Nb(r, p)$, Is $X$ has the pdf $f(x)$, such that:

$$f(x) = \binom{x-1}{r-1} p^r q^{x-r}, \quad x = r, r+1, r+2, \ldots, \quad 0 \leq p \leq 1.$$

Or

$$f(x) = \binom{x+r-1}{x} p^r q^x, \quad x = 0, 1, 2, \ldots, \quad 0 \leq p \leq 1.$$

***Note*:** The Maclaurian's series expansion of:

$$\begin{aligned}(1-a)^{-r} =& 1 + ra + \frac{r(r+1)}{2!}a^2 + \frac{r(r+1)(r+2)}{3!}a^3 + \ldots \\ =& \binom{r-1}{0}a^0 + \binom{r}{1}a^1 + \binom{r+1}{2}a^2 + \binom{r+2}{3}a^3 + \ldots \\ =& \sum_{x=0}^{\infty} \binom{x+r-1}{x} a^x, \quad r = 1, 2, 3, \ldots\end{aligned}$$

In order to verify that $f(x)$ is a valid pdf, we note that

1. $f(x) > 0, \forall x \in \mathcal{A} = \{x : x = 0, 1, 2, \ldots\}$.
2. $\sum_x f(x) = \sum_{x=0}^{\infty} \binom{x+r-1}{x} p^r q^x = p^r \sum_{x=0}^{\infty} \binom{x+r-1}{x} q^x = p^r(1-q)^{-r} = 1$.

**EXAMPLE 3.8**

A geological study indicates that an exploratory oil well drilled in a particular region should strike oil with probability 0.2. Find the probability that the third oil strike comes on the fifth well drilled.

**Solution.** Assuming independent drilling and probability 0.2 of striking oil with any one well, let $X$ denote the number of the trial on which the third oil strike occurs. Then it is reasonable to assume that $X$ has a negative binomial distribution with $p = 0.2$. Because we are interested in $r = 3$ and $x = 5$,

$$\Pr(X = 5) = f(5) = \binom{4}{2}(0.2)^3(0.8)^2 = 6(0.008)(0.64) = 0.0307.$$

◀

***The Moment Generating Function*:** The mgf of $X$ could be presented as:

$$M(t) = E\left[e^{tX}\right] = \sum_{x=0}^{\infty} e^{tx}\binom{x+r-1}{x}p^r q^x = p^r \sum_{x=0}^{\infty}\binom{x+r-1}{x}(qe^t)^x = p^r(1-qe^t)^{-r}$$
$$M(t) = \left(\frac{p}{1-qe^t}\right)^r, \quad t \neq -\ln q.$$

***Mean and Variance*:** If $X$ is a random variable with a negative binomial distribution,

$$\mu = E[X] = \frac{r}{p} \text{ and } \sigma^2 = Var(X) = \frac{rq}{p^2}$$

**EXAMPLE 3.9**

A large stockpile of used pumps contains20% that are in need of repair. A maintenance worker is sent to the stockpile with three repair kits. She selects pumps at random and tests them one at a time. If the pump works, she sets it aside for future use. However, if the pump does not work, she uses one of her repair kits on it. Suppose that it takes 10 minutes to test a pump that is in working condition and 30 minutes to test and repair a pump that does not work. Find the mean and variance of the total time it takes the maintenance worker to use her three repair kits.

**Solution.** Let X denote the number of the trial on which the third nonfunctioning pump is found. It follows that $X$ has a negative binomial distribution with $p = 0.2$. Thus, $E(X) = 3/(0.2) = 15$ and $Var(X) = 3(0.8)/(0.2)2 = 60$. Because it takes an additional 20 minutes to repair each defective pump, the total time necessary to use the three kits is

$$T = 10X + 3(20).$$

Therefore, the expected time is,

$$E[T] = 10E[X] + 60 = 10(15) + 60 = 210,$$

and

$$Var(T) = 10^2 Var(X) = (100)(60) = 6000.$$

Thus, the total time necessary to use all three kits has mean 210 and standard deviation $\sqrt{6000} = 77.46$.

◀

### 3.1.4 The Hypergeometric Probability Distribution

Suppose that a population contains a finite number $N$ of elements that possess one of two characteristics. Thus, $r$ of the elements might be red and $b = N - r$, black. A sample of $n$ elements is randomly selected from the population, and the random variable of interest is $X$, the number of red elements in the sample. This random variable has what is known as the hypergeometric probability distribution. For example, the number of workers who are women, $X$, in Example **??** has the hypergeometric distribution.

***Definition*:** A random variable $X$ is said to have a hypergeometric probability distribution if and only if

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, n \text{ and } x \leq r,\ n-y \leq N-r.$$

1. We can easily notice that $f(x) \geq 0, \forall x$, since $\binom{a}{b} > 0, if a > b$, and $\binom{a}{b} = 0$, if $b > a$.

2. Notice that:

$$\sum_{i=0}^{n} \binom{r}{i}\binom{N-r}{n-i} = \binom{N}{n}.$$

Therefore,

$$\sum_{x=0}^{n} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1.$$

This verifies that $f(x)$ is a valid pdf of the r.v $X$.

**EXAMPLE 3.10**

An important problem encountered by personnel directors and others faced with the selection of the best in a finite set of elements is exemplified by the following scenario. From a group of 20 Ph.D. engineers, 10 are randomly selected for employment. What is the probability that the 10 selected include all the 5 best engineers in the group of 20?

**Solution.** For this example $N = 20$, $n = 10$, and $r = 5$. That is, there are only 5 in the set of 5 best engineers, and we seek the probability that $Y = 5$, where $Y$ denotes the number of best engineers among the ten selected. Then

$$f(y) = \frac{\binom{5}{5}\binom{15}{5}}{\binom{20}{10}} = \left(\frac{15!}{5!10!}\right)\left(\frac{10!10!}{20!}\right) = \frac{21}{1292} = 0.0162.$$

◀

***Mean and Variance*:** If $X$ is a random variable with a hypergeometric distribution,

$$\mu = E[X] = \frac{nr}{N}, \text{ and } \sigma^2 = Var(X) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

**EXAMPLE 3.11**

An industrial product is shipped in lots of 20. Testing to determine whether an item is defective is costly, and hence the manufacturer samples his production rather than using a 100% inspection plan. A sampling plan, constructed to minimize the number of defectives shipped to customers, calls for sampling five items from each lot and rejecting the lot if more than one defective is observed. (If the lot is rejected, each item in it is later tested.) If a lot contains four defectives, what is the probability that it will be rejected? What is the expected number of defectives in the sample of size 5? What is the variance of the number of defectives in the sample of size 5?

**Solution.** Let $X$ equal the number of defectives in the sample. Then $N = 20$, $r = 4$, and $n = 5$. The lot will be rejected if $X = 2, 3$, or 4. Then

$$\begin{aligned} \Pr(\text{rejecting the lot}) &= \Pr(X \geq 2) = f(2) + f(3) + f(4) \\ &= 1 - f(0) - f(1) = 1 - \frac{\binom{4}{0}\binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1}\binom{16}{4}}{\binom{20}{4}} \\ &= 1 - 0.2817 - 0.4694 = 0.2487 \end{aligned}$$

The mean and variance of the number of defectives in the sample of size 5 are

$$\mu = \frac{(5)(4)}{20} = 1 \text{ and } \sigma^2 = 5\left(\frac{4}{20}\right)\left(\frac{20-4}{20}\right)\left(\frac{20-5}{20-1}\right) = 0.632.$$

◀

### 3.1.5 The Poisson Distribution

The Poisson probability distribution often provides a good model for the probability distribution of the number $X$ of rare events that occur in space, time, volume, or any other dimension, where $\lambda$ is the average value of $X$. Examples of random variables with approximate Poisson distributions are the number of telephone calls handled by a switchboard in a time interval, the number of radioactive particles that decay in a particular time period, the number of errors a typist makes in typing a page, and the number of automobiles using a freeway access ramp in a ten-minute interval.

***Definition*:** A random variable $X$ is said to have a Poisson probability distribution if and only if

$$f(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x = 0, 1, 2, \ldots, \ \lambda > 0,$$

and is denoted by $X \sim P(\lambda)$

1. $f(x) > 0, \forall x$.
2. Since $e^{\lambda} = \sum \lim_{x=0}^{\infty} \frac{\lambda^x}{x!}$, then

$$\sum_x f(x) = \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

**EXAMPLE 3.12**

Suppose that a random system of police patrol is devised so that a patrol officer may visit a given beat location $X = 0, 1, 2, 3, \ldots$ times per half-hour period, with each location being visited an average of once per time period. Assume that $X$ possesses, approximately, a Poisson probability distribution. Calculate the probability that the patrol officer will miss a given location during a half-hour period. What is the probability that it will be visited once? Twice? At least once?

**Solution.** For this example the time period is a half-hour, and the mean number of visits per half-hour interval is $\lambda = 1$. Then

$$f(x) = \frac{(1)^x e^{-1}}{x!} = \frac{e^{-1}}{x!}, \quad x = 0, 1, 2, \ldots.$$

The event that a given location is missed in a half-hour period corresponds to $(X = 0)$, and

$$\Pr(X = 0) = f(0) = \frac{e^{-1}}{0!} = e^{-1} = 0.368.$$

Similarly,

$$\Pr(X = 1) = f(1) = \frac{e^{-1}}{1!} = e^{-1} = 0.368.$$

and

$$\Pr(X = 2) = f(2) = \frac{e^{-1}}{2!} = \frac{e^{-1}}{2} = 0.184.$$

The probability that the location is visited at least once is the event $(X \geq 1)$. Then

$$\Pr(X \geq 1) = \sum_{x=1}^{\infty} f(x) = 1 - f(0) = 1 - e^{-1} = 0.632.$$

◀

***The Cumulative Distribution Function*:** The case in Poisson distribution is similar to the Binomial distribution. The cdf of $X$ is very complicated, as

$$F(x) = \Pr(X \leq x) = \sum_{\tau=0}^{x} \frac{\lambda^\tau}{\tau!} e^{-\lambda}.$$

The approximated value of Piosson cdf could be found in tables (2).

***The Moment Generating Function*:** The mgf of r.v $X$ is:

$$M(t) = E\left[e^{tX}\right] = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t}$$
$$M(t) = e^{\lambda(e^t-1)}$$

***Mean and Variance*:** If $X$ is a r.v possessing a Poisson distribution with parameter $\lambda$, then

$$\mu = E[X] = \lambda, \ \ and \ \ \sigma^2 = Var(X) = \lambda.$$

In order to prove that, set $\psi(t) = \ln M(t) = \lambda(e^t - 1)$. Then

$$\psi'(t) = \lambda e^t \Rightarrow \psi'(0) = \lambda e^0 = \lambda = \mu$$
$$\psi''(t) = \lambda e^t \Rightarrow \psi''(0) = \lambda e^0 = \lambda = \sigma^2$$

**EXAMPLE 3.13**

Let the r.v $X \sim P(2)$.Find $\mu$ and $\sigma^2$. Evaluate $\Pr(X \geq 1)$, $\Pr(2 \leq X \leq 5)$ and $\Pr(X = 3)$.

**Solution.** Since $\lambda = 2$, then: $\mu = \sigma^2 = 2$.

- $\Pr(X \geq 1) = 1 - \Pr(X \leq 0) = 1 - \sum_{x=0}^{1} f(x) = 1 - (f(0) + f(1)) = 1 - 0.1353 = 0.8646.$
- $\Pr(2 \leq X \leq 5) = F(5) - f(1) = 0.9834 - 0.4060 = 0.5774.$
- $\Pr(X = 3) = f(3) = e^{-2}\frac{2^3}{3!} = e^{-2}\frac{4}{3} = 0.1804.$

◀

**EXAMPLE 3.14**

Industrial accidents occur according to a Poisson process with an average of three accidents per month. During the last two months, ten accidents occurred. Does this number seem highly improbable if the mean number of accidents per month, $\mu$, is still equal to 3?

**Solution.** The number of accidents in two months, $X$, has a Poisson probability distribution with mean $\lambda^* = 2(3) = 6$. The probability that $X$ is as large as 10 is

$$\Pr(X \geq 10) = \sum_{x=10}^{\infty} \frac{6^x}{x!} e^{-6}.$$

The tedious calculation required to find $\Pr(X \geq 10)$ can be avoided by using Table 2,

$$\Pr(X \geq 10) = 1 - \Pr(X \leq 9) = 1 - 0.9161 = 0.0839$$

It is not highly improbable to happen. ◀

***Theorem*:** Let the r.v $X \sim b(n,p)$. For large $n \to \infty$ and small $p \to 0$. Then $X$ could be approximated as $\lim_{n\to\infty} b(n,p) = P(\lambda)$, where $\lambda = np$.

**EXAMPLE 3.15**

Let the r.v $X \sim b(3000, 0.001)$. Find $\Pr(X = 5)$.

**Solution.** Since $X \sim b(n,p)$ with a large value of $n = 3000$ and small value of $p = 0.001$. Then according to the above theorem, we can approximate that $X \sim \lambda$, where $\lambda = np = (3000)(0.001) = 3$. Then:

$$\Pr(X = 5) = \Pr(X \leq 5) - \Pr(X \leq 4) = 0.9161 - 0.8153 = 0.1008.$$

◀

## 3.2 Continuous Distribution

In this section, some of the most important and popular distributions for continuous r.v are presented. Both the pdf and cdf is derived and some important properties and mathematical expectation for these distribution are obtained.

### 3.2.1 The Uniform Distribution

Suppose that a bus always arrives at a particular stop between 8:00 and 8:10 A.M. and that the probability that the bus will arrive in any given subinterval of time is proportional only to the length of the subinterval. That is, the bus is as likely to arrive between 8:00 and 8:02 as it is to arrive between 8:06 and 8:08. Let $X$ denote the length of time a person must wait for the bus if that person arrived at the bus stop at exactly 8:00. If we carefully measured in minutes how long after 8:00 the bus arrived for several mornings, we could develop a relative frequency histogram for the data. From the description just given, it should be clear that the relative frequency with which we observed a value of $X$ between 0 and 2 would be approximately the same as the relative frequency with which we observed a value of $X$ between 6 and 8. The random variable $X$ just discussed is an example of a random variable that has a uniform distribution. The general form for the density function of a random variable with a uniform distribution is as follows.

***Definition*:** If $a < b$, a r.v $X$ is said to have a continuous uniform pdf on the interval $(a, b)$ if and only if the density function of $X$ is:

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

If $X$ has a uniform distribution, then $X$ is denoted by $X \sim U(a, b)$.

In order to check that $f(X)$ is a pdf, we note that:

1. $f(x) > 0, \forall x \in (a, b)$.

2. $\int_x f(x)dx = \int_a^b \frac{dx}{b-a} = \frac{x}{b-a}\Big|_a^b = \frac{b-a}{b-a} = 1$.

**EXAMPLE 3.16**

Arrivals of customers at a checkout counter follow a Poisson distribution. It is known that, during a given 30-minute period, one customer arrived at the counter. Find the probability that the customer arrived during the last 5 minutes of the 30-minute period.

**Solution.** As just mentioned, the actual time of arrival follows a uniform distribution over the interval of $(0, 30)$. If $X$ denotes the arrival time, then

$$\Pr(25 < X < 30) = \int_{25}^{30} \frac{1}{30} dx = \frac{30 - 25}{30} = \frac{5}{30} = \frac{1}{6}.$$

The probability of the arrival occurring in any other 5-minute interval is also 1/6. ◀

***The Cumulative Distribution Function*:** The cdf of $X$ is derived as

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(\tau) d\tau = \int_{a}^{x} \frac{d\tau}{b-a} = \frac{\tau}{b-a}\Big|_a^x$$

Therefore:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

***Mean and Variance*:** If $a < b$ and $X$ is a r.v uniformly distributed on the interval $(a, b)$, then

$$\mu = E[X] = \frac{a+b}{2} \text{ and } \sigma^2 = Var(X) = \frac{(b-a)^2}{12}.$$

It is easy to evaluate the mean and variance of $X$ by direct definition as

$$E[X] = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}.$$

also

$$E[X^2] = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3},$$

therefore

$$\sigma^2 = E[X^2] - \mu^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

**EXAMPLE 3.17**

Let the r.v $X \sim U(a, b)$ with $\mu = 1$ and $\sigma^2 = \frac{4}{3}$. Find $\Pr(X < 0)$.

**Solution.** Since the mean $\mu = 1$ and

$$\mu = \frac{a+b}{2} = 1 \Rightarrow a + b = 2 \tag{3.1}$$

also, the variance $\sigma^2 = \frac{4}{3}$, and

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{4}{3} \Rightarrow (b-a)^2 = 16 \Rightarrow b - a = \pm 4 \tag{3.2}$$

Consider $a + b = 2$ and $b - a = -4$. Hence $b = -1$ and $a = 3$. This is non applicable because $X \sim U(a, b)$ and therefore $a < b$. Then, $a + b = 2$ and $b - a = 4$. Hence, $b = 3$ and $a = -1$.

$$\therefore f(x) = \frac{1}{b-a} = \frac{1}{3-(-1)} = \frac{1}{4}, \quad -1 < x < 3.$$

Therefore,

$$\Pr(x < 0) = \int_{-1}^{0} f(x)dx = \frac{1}{4}\int_{-1}^{0} dx = \frac{1}{4}.$$

◀

**EXAMPLE 3.18**

If a parachutist lands at a random point on a line between markers $A$ and $B$, find the probability that he is closer to $A$ than to $B$. Find the probability that his distance to $A$ is more than three times his distance to $B$.

**Solution.** Consider the landing point $X$ is a r.v and $X \sim (A, B)$, then for the first part of the question, we assume that $X - A < B - X$, and

$$\Pr(X - A < B - X) = \Pr(2X < B + A) = \Pr\left(X < \frac{B+A}{2}\right) = F\left(\frac{B+A}{2}\right)$$
$$= F\left(\frac{B+A}{2}\right) = \frac{\frac{B+A}{2} - A}{B - A} = \frac{\frac{A+B-2A}{2}}{B-A} = \frac{1}{2}.$$

For the second part, assuming that $X - A > 3(B - X)$, and that

$$\Pr(X - A > 3(B - X)) = \Pr(4X > 3B + A) = \Pr\left(X > \frac{3B+A}{4}\right) = 1 - F\left(\frac{3B+A}{4}\right)$$
$$= 1 - F\left(\frac{3B+A}{4}\right) = 1 - \frac{\frac{3B+A}{4} - A}{B-A} = 1 - \frac{\frac{3B+A-4A}{4}}{B-A} = 1 - \frac{\frac{3(B-A)}{4}}{B-A} = 1 - \frac{3}{4} = \frac{1}{4}.$$

◀

### 3.2.2 Gamma Distribution

Some random variables are always non-negative and for various reasons yield distributions of data that are skewed (non-symmetric) to the right. That is, most of the area under the density function is located near the origin, and the density function drops gradually as $x$ increases. A skewed probability density function is shown in Figure **??**

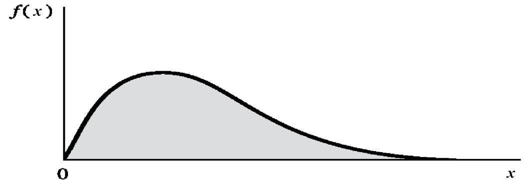


**Figure 3.1** A skewed probability density function

A family of pdf's that yields a wide variety of skewed distributional shapes is the gamma family. To define the family of gamma distributions, we first need to introduce a function that plays an important role in many branches of mathematics.

***Gamma Function*:** For any $\alpha > 0$, the gamma function $\Gamma(\alpha)$ is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx.$$

The most important properties of the gamma function are the following:

1. For any $\alpha > 0$, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. (via integration by parts).
2. For any positive integer, n, $\Gamma(n) = (n-1)!$.
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

***Probability Density Function*:** A random variable $X$ is said to have a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ if and only if the density function of $X$ is

$$f(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}, \qquad 0 \le x < \infty$$

and $X$ is denoted by $X \sim G(\alpha, \beta)$. To prove that $f(x)$ is a valid pdf, we need to check:

1. $f(x) > 0, \forall x \in [0, \infty)$.
2. to prove the unity of integration, we use the gamma function, then

$$\int_x f(x)dx = \int_0^\infty \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{\beta}}dx$$

Let $y = \frac{x}{\beta} \Rightarrow x = \beta y \Rightarrow dx = \beta dy$. Then

$$\int_0^\infty f(x)dx = \int_0^\infty \frac{1}{\beta^\alpha\Gamma(\alpha)}(\beta y)^{\alpha-1}e^{-y}\beta dy = \frac{1}{\Gamma(\alpha)}\int_0^\infty y^{\alpha-1}e^{-y}dy = \frac{1}{\Gamma(\alpha)}\Gamma(\alpha) = 1.$$

Therefore, $f(x)$ is a pdf.

***Cumulative Distribution Function*:** In the special case when $\alpha$ is an integer, cdf of a gamma distributed r.v can be expressed as a sum of certain Poisson probabilities. However, if $\alpha$ is not an integer, it is impossible to give a closed-form expression for

$$\int_0^x \frac{1}{\beta^\alpha \Gamma(\alpha)} \tau^{\alpha-1} e^{-\frac{\tau}{\beta}} d\tau$$

This integral is called an incomplete gamma function, and except when $\alpha = 1$ (an exponential distribution), it is impossible to obtain areas under the gamma density function by direct integration. There are not many tables of $F(x)$ available; in Table A.4, we present a small tabulation for $\alpha = 1, 2, \ldots, 10$, $\beta = 1$ and $x = 1, 2, \ldots, 15$.

***Moment Generating Function*:** The moment generating function of a gamma random variable is

$$M(t) = \frac{1}{(1-\beta t)^\alpha}, \quad t < \frac{1}{\beta}.$$

We use the definition of the mgf to prove it, such as

$$M(t) = E[e^{tX}] = \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\frac{1}{\beta}(1-\beta t)x} dx$$

Set $y = \left(\frac{1-\beta t}{\beta}\right) x \Rightarrow x = \left(\frac{\beta}{1-\beta t}\right) y \Rightarrow dx = \left(\frac{\beta}{1-\beta t}\right) dy$, then

$$M(t) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} e^{-y} \left(\frac{\beta}{1-\beta t}\right) dy = \frac{1}{\beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{1-\beta t}\right)^\alpha \int_0^\infty y^{\alpha-1} e^{-y} dy = \frac{1}{(1-\beta t)^\alpha}.$$

***Mean and Variance*:** If $X$ has a gamma distribution with parameters $\alpha$ and $\beta$, then

$$\mu = E[X] = \alpha\beta, \text{ and } \sigma^2 = Var(X) = \alpha\beta^2.$$

We can prove that either with direct integration or using the logarithm for the mgf.

$$\psi(t) = \ln(M(t)) = -\alpha \ln(1-\beta t) \Rightarrow \psi'(t) = \frac{\alpha\beta}{(1-\beta t)} \Rightarrow \psi'(0) = \alpha\beta = \mu.$$

and

$$\psi''(t) = \frac{\alpha\beta^2}{(1-\beta t)^2} \Rightarrow \psi''(0) = \alpha\beta^2 = \sigma^2.$$

**EXAMPLE 3.19**

Let the r.v $X \sim G(3, 2)$. Write down the pdf and mgf of $X$. Find the mean $\mu$ and variance $\sigma^2$. Evaluate the probability $\Pr(12.6 < x < 16.8)$.

**Solution.** Since we have $\alpha = 3$ and $\beta = 2$, then the pdf of $X$ is

$$f(x) = \frac{1}{2^3 \Gamma(3)} x^{3-1} e^{-\frac{x}{2}} = \frac{1}{16} x^2 e^{-\frac{x}{2}}, 0 < x < \infty.$$

then, the mgf is

$$M(t) = \frac{1}{(1-2t)^3}.$$

also, $\mu = \alpha\beta = 6$ and $\sigma^2 = \alpha\beta^2 = 12$. Finally,

$$\Pr(12.6 < X < 16.8) = \int_{12.6}^{16.8} f(x) = \frac{1}{16} \int_{12.6}^{16.8} x^2 e^{-\frac{x}{2}} dx = 0.04.$$

◀

**EXAMPLE 3.20**

Find the constant $c$ so that the function $f(x) = cx^5 e^{-3x}$, $0 < x < \infty$ is a pdf.

**Solution.** Since $f(x)$ is a pdf, then

$$1 = \int_0^\infty f(x)dx = c\int_0^\infty x^5 e^{-3x}dx = c\frac{\Gamma(6)}{3^6}\int_0^\infty \frac{3^6}{\Gamma(6)}x^{6-1}e^{-3x}dx = \frac{5!c}{3^6}$$

Therefore, $c = \frac{3^6}{5!}$. ◀

**EXAMPLE 3.21**

Suppose the survival time $X$ in weeks of a randomly selected male mouse exposed to 240 rads of gamma radiation has a gamma distribution with $\alpha = 8$ and $\beta = 1$. What is the expected survival time for the mouse? what is the probability that the mouse could survive between 6 and 12 weeks and the probability of it survives at least 3 weeks?

**Solution.** Since $X \sim G(8, 1)$, then the expected survival time in weeks is $E[X] = \alpha\beta = (8)(1) = 8$ weeks.
The probability that a mouse survives between 6 and 12 weeks is

$$\Pr(6 \le X \le 12) = \Pr(X \le 12) - \Pr(X \le 6) = 0.911 - 0.256 = 0.655.$$

The probability that a mouse survives at least 3 weeks is

$$\Pr(X \ge 3) = 1 - \Pr(X \le 3) = 1 - 0.012 = 0.998.$$

◀

### 3.2.3 The Exponential Distribution

The family of exponential distributions provides probability models that are widely used in engineering and science disciplines. For example, the exponential density function is often useful for modelling the length of life of electronic components.

***Definition*:** A r.v $X$ is said to have an exponential distribution with parameter $\lambda > 0$ if the pdf of $X$ is

$$f(x) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}, \quad x \ge 0.$$

The r.v is denoted by $X \sim Exp(\lambda)$. In order to prove that $f(x)$ is a valid pdf, we can simply prove

1. $f(x) > 0, \forall x \in (0, \infty)$.
2. It is easy to prove that the integral

$$\int_x f(x)dx = \int_0^\infty \frac{1}{\lambda}e^{-\frac{x}{\lambda}}dx = -e^{-\frac{x}{\lambda}}\Big|_0^\infty = -(e^{-\infty} - e^0) = 1.$$

***Note*:** The exponential pdf is a special case of the general gamma pdf in which $\alpha = 1$ and $\beta$ has been replaced by $1/\lambda$.

***The Cumulative Distribution Function*:** The cdf of a r.v that is exponentially distributed could be evaluate as

$$F(x) = \Pr(X \le x) = \int_{-\infty}^x f(\tau)d\tau = \int_0^x \frac{1}{\lambda}e^{-\frac{\tau}{\lambda}}d\tau = -e^{-\frac{\tau}{\lambda}}\Big|_0^x = 1 - e^{-\frac{x}{\lambda}},$$

therefore

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\frac{x}{\lambda}}, & 0 < x < \infty \\ 1, & x = \infty \end{cases}$$

***The Moment Generating Function*:** The mgf of $X$ is also evaluated for exponential distributed r.v as

$$M(t) = E[e^{tX}] = \int_x e^{tx} f(x)dx = \int_0^\infty e^{tx} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx = \int_0^\infty \frac{1}{\lambda} e^{-\frac{1}{\lambda}(1-\lambda t)x} = -\frac{e^{-\frac{1}{\lambda}(1-\lambda t)x}}{(1-\lambda t)}\Big|_0^\infty$$
$$= \frac{1}{1-\lambda t}, \quad t < \frac{1}{\lambda}.$$

***Mean and Variance*:**If $X$ is an exponential r.v with parameter $\lambda$, then

$$\mu = E[X] = \lambda \text{ and } \sigma^2 = Var(X) = \lambda^2.$$

The proof simply follows directly from the mean and variance of the gamma distribution with $\alpha = 1$.

***Note*:** When $\lambda = 1$, then r.v $X \sim Exp(1)$ is said to have a standard exponential distribution with pdf $f(x) = e^{-x},\ 0 < x < \infty$.

**EXAMPLE 3.22**

Let the r.v $X \sim Exp(\lambda)$. If $\Pr(X \leq 1) = \Pr(X > 1)$. Find $\mu$ and $\sigma^2$.

**Solution.** Since $\Pr(X \leq 1) = \Pr(x > 1) = 1 - \Pr(X \leq 1) \Rightarrow 2\Pr(X \leq 1) = 1 \Rightarrow \Pr(X \leq 1) = \frac{1}{2}$. Therefore, using the cdf $F(x) = 1 - e^{-\frac{x}{\lambda}},\ 0 < x < \infty$, then

$$\Pr(X \leq 1) = F(1) = 1 - e^{-\frac{1}{\lambda}} = \frac{1}{2} \Rightarrow e^{-\frac{1}{\lambda}} = \frac{1}{2} \Rightarrow -\frac{1}{\lambda} = \ln\left(\frac{1}{2}\right) = \ln(1) - \ln(2) = 0 - \ln(2) = -\ln(2)$$
$$\Rightarrow \lambda = \frac{1}{\ln(2)}.$$

Hence, $\mu = \lambda = \frac{1}{\ln(2)}$ and $\sigma^2 = \lambda^2 = \left(\frac{1}{\ln(2)}\right)^2$. ◀

**EXAMPLE 3.23**

If a r.v $X \sim Exp(2)$. Find $\Pr(X \leq 1 | X \leq 2)$.

**Solution.**

$$\Pr(X \leq 1 | X \leq 2) = \frac{\Pr(X \leq 1 \cap X \leq 2)}{\Pr(X \leq 2)} = \frac{\Pr(X \leq 1)}{\Pr(X \leq 2)} = \frac{F(1)}{F(2)} = \frac{1 - e^{-\frac{1}{2}}}{1 - e^{-1}}.$$

◀

**EXAMPLE 3.24**

The response time $X$ at an on-line computer terminal (the elapsed time between the end of a user's inquiry and the beginning of the system's response to that inquiry) has an exponential distribution with expected response time equal to 5 s. Then $E[X] = \lambda = 5$ s. Find the probability that the response time is at most 10 s and the probability that response time is between 5 and 10 s.

**Solution.** The probability that the response time is at most 10 s is

$$\Pr(X \leq 10) = F(10) = 1 - e^{-\frac{10}{5}} = 1 - e^{-2} = 0.865.$$

The probability that response time is between 5 and 10 s is

$$\Pr(5 \leq X \leq 10) = F(10) - F(5) = (1 - e^{-2}) - (1 - e^{-1}) = 0.233.$$

◀

**EXAMPLE 3.25**

One-hour carbon monoxide concentrations in air samples from a large city have an approximately exponential distribution with mean 3.6 ppm (parts per million).

**a** Find the probability that the carbon monoxide concentration exceeds 9 ppm during a randomly selected one-hour period.

**b** A traffic-control strategy reduced the mean to 2.5 ppm. Now find the probability that the concentration exceeds 9 ppm.

**Solution.** Since $X \sim Exp(\lambda)$, and the mean of carbon monoxide concentrations in air is 3.6, then

$$\Pr(X \geq 9) = 1 - \Pr(X \leq 9) = 1 - F(9) = 1 - (1 - e^{-\frac{9}{3.6}}) = 0.082$$

If the mean of carbon monoxide concentrations in air is reduced to 2.5, then

$$\Pr(X \geq 9) = 1 - \Pr(X \leq 9) = 1 - F(9) = 1 - (1 - e^{-\frac{9}{2.5}}) = 0.027.$$

◀

### 3.2.4 Chi-Square Distribution

Let $\nu$ be a positive integer. A random variable $X$ is said to have a chi-square distribution with $\nu$ degrees of freedom (dof) if $X$ is a gamma-distributed r.v with parameters $\alpha = \nu/2$ and $\beta = 2$. Such random variables occur often in statistical theory. The motivation behind calling the parameter $\nu$ the degrees of freedom of the Chi-square distribution rests on one of the major ways for generating a random variable with this distribution.

***The Probability Density Function*:** The r.v $X$ is said to have a Chi-square distribution if it has a pdf

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2}, \quad x \geq 0.$$

and is denoted by $X \sim \chi^2(\nu)$.

***The Cumulative Distribution Function*:** As the case of the gamma distribution, the cdf of chi-square distribution is indeterminable as the integral

$$F(x) = \Pr(X \leq x) = \int_0^x \frac{1}{2^{(\nu/2)}\Gamma(\nu/2)} \tau^{(\nu/2)-1} e^{-(\tau/2)} d\tau$$

is almost impossible to solve. However, values of this integral corresponding to specific values of $x$ and $\nu$ are given in the tables.

***The Moment Generating Function*:** The mgf of the r.v $X \sim \chi^2(\nu)$ is

$$M(t) = E[e^{tX}] = \frac{1}{(1-2t)^{\nu/2}}, \quad t < \frac{1}{2}.$$

***Mean and Variance*:** Since $X \sim \chi^2(\nu)$ is equal to $X \sim G(\frac{\nu}{2}, 2)$, then the mean and variance are

$$\mu = E[X] = \nu \text{ and } \sigma^2 = Var(X) = 2\nu.$$

**EXAMPLE 3.26**

Let the r.v $X$ has an mgf $M(t) = (1-2t)^{-8}$. What is the distribution of $X$? Find $\mu$ and $\sigma^2$. Evaluate $\Pr(6.97 < X < 26.3)$.

**Solution.** The given mgf is for a r.v $X \sim G(8, 2)$ or $X \sim \chi^2(16)$, with dof $\nu = 16$. Therefore $\mu = \nu = 16$ and $\sigma^2 = 2\nu = 32$.

The probability of

$$\Pr(7.96 < X < 26.3) = \Pr(X < 26.3) - \Pr(X < 7.96) = 0.95 - 0.05 = 0.9.$$

◄

### 3.2.5 The Beta Distribution

The beta density function is a two-parameter density function defined over the closed interval $0 \leq x \leq 1$. It is often used as a model for proportions, such as the proportion of impurities in a chemical product or the proportion of time that a machine is under repair.

***The Probability Density Function*:** A r.v $X$ is said to have a beta probability distribution with parameters $\alpha > 0$ and $\beta > 0$, denoted by $X \sim Be(\alpha, \beta)$, if the density function of $X$ is

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \leq x \leq 1,$$

where $B(\alpha, \beta)$, the beta function, is defined as

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

To prove that $f(x)$ is a pdf, we need to check that

1. $f(x) > r, \forall x \in (0, 1)$.

2. To show that $\int\limits_x f(x)dx = 1$. Consider the integral $I = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$. Successive integration by parts gives

$$\begin{array}{ccc}
(1-x)^{\beta-1} & & x^{\alpha-1} \\
-(\beta-1)(1-x)^{\beta-2} & \searrow & \frac{x^{\alpha}}{\alpha} \\
(\beta-1)(\beta-2)(1-x)^{\beta-3} & \searrow & \frac{x^{\alpha+1}}{\alpha(\alpha+1)} \\
\vdots & \vdots & \frac{x^{\alpha+2}}{\alpha(\alpha+1)(\alpha+2)} \\
\pm(\beta-1)(\beta-2)\ldots[\beta-(\beta-1)](1-x)^{\beta-\beta} & \searrow & \vdots \\
0 & & \frac{x^{\alpha+\beta-1}}{\alpha(\alpha+1)\ldots(\alpha+\beta-1)}
\end{array}$$

$$\begin{aligned}
\therefore I &= \frac{1}{\alpha}x^{\alpha}(1-x)^{\beta-1} + \frac{(\beta-1)}{\alpha(\alpha+1)}x^{\alpha+1}(1-x)^{\beta-2} + \frac{(\beta-1)(\beta-2)}{\alpha(\alpha+1)(\alpha+2)}x^{\alpha+2}(1-x)^{\beta-3} + \ldots \\
&+ \frac{(\beta-1)(\beta-2)\ldots 3\times 2\times 1}{\alpha(\alpha+1)(\alpha+2)\ldots(\alpha+\beta-1)}x^{\alpha+\beta-1}\Big|_0^1 \\
&= \frac{(\beta-1)(\beta-2)\ldots 3\times 2\times 1}{\alpha(\alpha+1)(\alpha+2)\ldots(\alpha+\beta-1)} = \frac{(\beta-1)!}{\alpha(\alpha+1)(\alpha+2)\ldots(\alpha+\beta-1)} \times \frac{(\alpha-1)(\alpha-2)\ldots 3\times 2\times 1}{(\alpha-1)(\alpha-2)\ldots 3\times 2\times 1} \\
\therefore I &= \frac{(\beta-1)!(\alpha-1)!}{(\alpha+\beta-1)!} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.
\end{aligned}$$

Thus, the integral of $f(x)$ is

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = 1$$

***Note*:** In the case of $\alpha = \beta = 1$, the beta distribution reduces to be a standard uniform, that is $Be(1,1) \equiv U(0,1)$.

***The Cumulative Distribution Function*:** The cdf for the beta r.v is commonly called the incomplete beta function and is denoted by

$$F(x) = \int_0^x \frac{\tau^{\alpha-1}(1-\tau)^{\beta-1}}{B(\alpha,\beta)}d\tau.$$

A tabulation of $F(x)$ is given in Tables of the Incomplete Beta Function (Pearson, 1968). When $\alpha$ and $\beta$ are both positive integers, $F(x)$ is related to the binomial probability function.

***Mean and Variance*:** If $X$ is a beta distributed r.v with parameters $\alpha > 0$ and $\beta > 0$, then

$$\mu = E[X] = \frac{\alpha}{\alpha+\beta} \text{ and } \sigma^2 = Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

We can derive the r$^{th}$ moment by the direct definition and by the use of the same technique we used to prove the unity of the pdf integration, then we can find that

$$E[X^r] = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha)\Gamma(\alpha+\beta+r)}, \qquad r = 1,2,3,\ldots$$

In order to derive the mean and variance, set $r = 1, 2$ to get

$$r = 1 \Rightarrow \mu = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)} = \frac{(\alpha+\beta-1)!\,\alpha!}{(\alpha-1)!\,(\alpha+\beta)!} = \frac{\alpha}{\alpha+\beta}.$$

$$r = 2 \Rightarrow \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} \Rightarrow \sigma^2 = E[X^2] - \mu^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

**EXAMPLE 3.27**

Find the constant $c$ in the pdf: $f(x) = c\,x^{19}(1-x)^{29},\ \ 0 < x < 1.$

**Solution.** Since $f(x)$ is a pdf, then

$$\begin{aligned} 1 &= c\int_0^1 x^{19}(1-x)^{29}dx = c\frac{\Gamma(20)\Gamma(30)}{\Gamma(50)}\int_0^1 \frac{\Gamma(50)}{\Gamma(20)\Gamma(30)}x^{20-1}(1-x)^{30-1}dx \\ \Rightarrow c &= \frac{\Gamma(50)}{\Gamma(20)\Gamma(30)}. \end{aligned}$$

◀

**EXAMPLE 3.28**

A gasoline wholesale distributor has bulk storage tanks that hold fixed supplies and are filled every Monday. Of interest to the wholesaler is the proportion of this supply that is sold during the week. Over many weeks of observation, the distributor found that this proportion could be modelled by a beta distribution with $\alpha = 4$ and $\beta = 2$. Find the probability that the wholesaler will sell at least 90% of his stock in a given week.

**Solution.** If $Y$ denotes the proportion sold during the week, then

$$f(y) = \frac{\Gamma(4+2)}{\Gamma(4)\Gamma(2)}y^3(1-y), \quad 0 \le y \le 1.$$

and therefore,

$$\Pr(Y > 0.9) = \int_{0.9}^1 f(y)dy = 20\int_{0.9}^1 y^3(1-y)dy = 20\left\{\frac{y^4}{4}\Big|_{0.9}^1 - \frac{y^5}{5}\Big|_{0.9}^1\right\} = (20)(0.004) = 0.08.$$

It is not very likely that 90% of the stock will be sold in a given week. ◀

### 3.2.6 The Normal Distribution

The normal distribution with the familiar bell shape is the most important one in all of probability and statistics. Many numerical populations have distributions that can be fit very closely by an appropriate normal curve. Examples include heights, weights, and other physical characteristics, measurement errors in scientific experiments, measurements on fossils, reaction times in psychological experiments, measurements of intelligence and aptitude, scores on various tests, and numerous economic measures and indicators. Even when the underlying distribution is discrete, the normal curve often gives an excellent approximation. In addition, even when individual variables themselves are not normally distributed, sums and averages of the variables will under suitable conditions have approximately a normal distribution; this is the content of the Central Limit Theorem discussed in the following chapters.

***Probability Density Function*:** A continuous r.v $X$ is said to have a normal distribution with parameters $\mu$ and $\sigma^2$, where $-\infty < \mu < \infty$ and $\sigma > 0$, if the pdf of $X$ is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \infty < x < \infty.$$

The r.v $X$ is denoted by $X \sim N(\mu, \sigma^2)$. Here is a proof that the normal curve satisfies the pdf requirement:

1. $f(x) > 0$, $\forall x \in (-\infty, \infty)$.

2. In order to satisfy $\int_x f(x)dx = 1$, we will need to use some special technique. First of all consider the integral $I = \int_{-\infty}^{\infty} e^{-x^2/2}dx$. Then

$$I^2 = \left\{ \int_{-\infty}^{\infty} e^{-x^2/2}dx \right\} \left\{ \int_{-\infty}^{\infty} e^{-y^2/2}dy \right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy$$

Use the Polar transformation by setting, $x = r\cos\theta,\ 0 \le r < \infty$ and $y = r\sin\theta,\ 0 \le \theta \le 2\pi$. Also, the Jacobian for the transformation is

$$J = \frac{\partial(x,y)}{\partial(r,\theta)} = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

Then, $dxdy = |J|drd\theta = rdrd\theta$. This implies that the integral $I^2$ can be written as

$$I^2 = \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} e^{-r^2/2} rdrd\theta = 2\pi \int_{r=0}^{\infty} re^{-r^2/2}dr = -(2\pi) \left[e^{-r^2/2}\right]_0^{\infty} = 2\pi$$

Hence, $I = \int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}$. Back to the Normal distribution function,

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

set $y = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma y + \mu$, then $\Rightarrow dx = \sigma dy$ and

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-y^2/2}\sigma dy = \frac{1}{\sqrt{2\pi}}\sqrt{2\pi} = 1.$$

This proves that $f(x)$ is a pdf.

***Cumulative Distribution Function*:** There is no a closed form expression for the cdf of a normal distributed r.v because the integral $F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(\tau)d\tau$ has no analytic solution. hence, its evaluation requires the use of numerical integration techniques. However, in literature, tables are approximated for the case of standard Normal distribution, that is $X \sim N(0,1)$.

***Definition*:** The normal distribution with mean $\mu = 0$ and $\sigma^2 = 1$ is called the standard normal distribution. A r.v that has a standard normal distribution is called a standard normal random variable and will be denoted by $Z$. The pdf of $Z$ is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$

The cdf of $Z$ is $F(z) = \Pr(Z \le z) = \int_{-\infty}^{z} f(\tau)d\tau$, which we will denote by $\Phi(z)$.

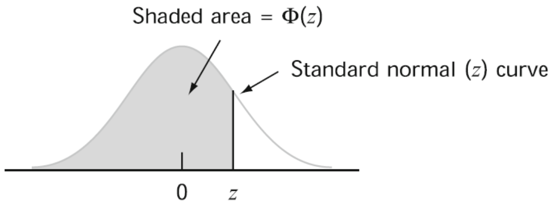


**Figure 3.2** Standard Normal Distribution

**EXAMPLE 3.29**

Let $Z$ denote a normal random variable with mean 0 and standard deviation 1.

1. Find $\Pr(Z > 1.25)$.
2. Find $\Pr(-0.38 \leq Z \leq 1.25)$.

**Solution.**

1. $\Pr(Z > 1.25) = 1 - \Pr(Z \leq 1.25) = 1 - \Phi(1.25)$, the probability $\Phi(z)$ can be found in tables and is $\Phi(1.25) = 0.8944$. Therefore $\Pr(Z > 1.25) = 1 - 0.8944 = 0.1056$.
2. $\Pr(-0.38 \leq Z \leq 1.25) = \Phi(1.25) - \Phi(-0.38) = 0.8944 - 0.3520 = 0.5424$.

◀

***Nonstandard Normal Distributions*:** When $X \sim N(\mu, \sigma^2)$, probabilities involving $X$ are computed by "standardizing." The standardized variable is $(X - \mu)/\sigma$. If $X$ has a normal distribution with mean $\mu$ and standard deviation $\sigma$, then

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

has a standard normal distribution. Thus

$$\begin{aligned}\Pr(a \leq X \leq b) &= \Pr\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)\\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).\end{aligned}$$

Therefore,

$$\Pr(X \leq a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \quad \text{and} \quad \Pr(X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right).$$

**EXAMPLE 3.30**

The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in avoiding rear-end collisions. It is suggested that reaction time for an in-traffic response to a brake signal from standard brake lights can be modelled with a normal distribution having mean value 1.25 s and standard deviation of 0.46 s. What is the probability that reaction time is between 1.00 and 1.75 s?

**Solution.** Let $X$ denote reaction time, then standardizing gives

$$1 \leq X \leq 1.75$$

if and only if

$$\frac{1-1.25}{0.46} \leq \frac{x-1.25}{0.46} \leq \frac{1.75-1.25}{0.46}.$$

Thus

$$\begin{aligned}\Pr(1 \leq X \leq 1.75) &= \Pr\left(\frac{1-1.25}{0.46} \leq Z \leq \frac{1.75-1.25}{0.46}\right)\\ &= \Pr(-0.54 \leq Z \leq 1.09) = \Phi(1.09) - \Phi(-0.54)\\ &= 0.8621 - 0.2946 = 0.5675.\end{aligned}$$

◄

### EXAMPLE 3.31

If the r.v $X \sim N(\mu, \sigma^2)$, such that $\Pr(X \leq 60) = 0.1$ and $\Pr(X > 90) = 0.05$. Find $\mu$ and $\sigma^2$.

**Solution.** Since $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$. Hence

$$0.1 = \Pr(X \leq 60) = \Pr\left(\frac{X-\mu}{\sigma} \leq \frac{60-\mu}{\sigma}\right) = \Pr\left(Z \leq \frac{60-\mu}{\sigma}\right).$$

From $\Phi(z)$ tables, we can find that $\frac{60-\mu}{\sigma} = -1.282$.

Similarly,

$$0.05 = \Pr(X > 90) = \Pr\left(\frac{X-\mu}{\sigma} > \frac{90-\mu}{\sigma}\right) = \Pr\left(Z > \frac{90-\mu}{\sigma}\right) = 1 - \Pr\left(Z \leq \frac{90-\mu}{\sigma}\right).$$

From $\Phi(z)$ tables, we can find that $\frac{90-\mu}{\sigma} = 1.645$. Then, it is easy to find that $\mu = 73.1$ and $\sigma = 10.2$. ◄

***Note*:**

1. If a r.v $X \sim N(0,1)$, then the r.v $Y = X^2 \sim \chi^2(1)$.
2. If a r.v $Z \sim N(0,1)$, then $\Pr(Z \leq -a) = \Pr(Z > a)$.
3. If a r.v $Z \sim N(0,1)$, then $\Pr(-a \leq Z \leq a) = 2\Pr(Z \leq a) - 1$

***Moment Generating Function*:** The mgf of a r.v $X \sim N(\mu, \sigma^2)$ is

$$M(t) = E[e^{tX}] = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

To prove that,

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx)} dx,$$

take

$$\begin{aligned}x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx &= x^2 - 2(\mu + \sigma^2 t)x + (\mu + \sigma^2 t)^2 - (\mu + \sigma^2 t)^2 + \mu^2\\ &= [x - (\mu + \sigma^2 t)]^2 - \mu - 2\mu\sigma^2 t - \sigma^4 t^2 + \mu^2\\ &= [x - (\mu + \sigma^2 t)]^2 - 2(\mu t + \frac{1}{2}\sigma^2 t^2)\sigma^2.\end{aligned}$$

Therefore,

$$M(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[x-(\mu+\sigma^2 t)]^2} \; e^{\mu t + \frac{1}{2}\sigma^2 t^2} dx = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}[x-(\mu+\sigma^2 t)]^2} \, dx.$$

That gives $M(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ since the integral is equal to one.

**EXAMPLE 3.32**

Let the r.v $X$ has an mgf of $M(t) = e^{5t+2t^2}$. What is the distribution of $X$? Find $\Pr(3 \leq X \leq 7)$

**Solution.** We can rewrite the mgf as $M(t) = e^{5t+\frac{1}{2}4t^2}$, which means that $X \sim N(5, 4)$ with $\mu = 5$ and $\sigma^2 = 4$. To find the probability,

$$\begin{aligned} \Pr(3 \leq X \leq 7) &= \Pr\left(\frac{3-5}{2} \leq Z \leq \frac{7-5}{2}\right) = \Pr(-1 \leq Z \leq 1) = 2\Pr(Z \leq 1) - 1 \\ &= 2(0.841) - 1 = 1.682 - 1 = 0.682. \end{aligned}$$

◄

# CHAPTER 4

# MULTIVARIATE PROBABILITY DISTRIBUTIONS

## 4.1 Introduction

In Chapters **??** and **??**, we studied probability models for a single random variable. Many problems in probability and statistics lead to models involving several random variables simultaneously. For example, a gambler playing blackjack is interested in the event of drawing both an ace and a face card from a 52-card deck. A biologist, observing the number of animals surviving in a litter, is concerned about the intersection of these events:

**A:** The litter contains $n$ animals.

**B:** $x$ animals survive.

Similarly, observing both the height and the weight of an individual represents the intersection of a specific pair of events associated with height–weight measurements. In this chapter, we first discuss probability models for the joint behaviour of several random variables, putting special emphasis on the case in which the variables are independent of each other. We then study expected values of functions of several random variables, including covariance and correlation as measures of the degree of association between two variables. Then we will considers conditional distributions, the distributions of random variables given the values of other random variables.

## 4.2 Bivariate and Multivariate Probability Distributions

There are many experimental situations in which more than one r.v will be of interest to an investigator. We shall first consider joint probability distributions for two discrete r.v's, then for two continuous variables, and finally for more than two variables.

### 4.2.1 Joint Probability Density Function

We already defined the pdf of more that one r.v in Chapter **??** as: Let $X_1$ and $X_2$ be two r.vs, then the ***Joint*** pdf of $X_1$ and $X_2$ is given by:

$$f(x_1, x_2) = \Pr(X_1 = x_1, X_2 = x_2) = \begin{cases} \sum\limits_{x_1}\sum\limits_{x_2} f(x_1, x_2), & \text{discrete} \\ \int\limits_{x_1}\int\limits_{x_2} f(x_1, x_2)dx_1dx_2, & \text{continuous} \end{cases}$$

**EXAMPLE 4.1**

A local supermarket has three checkout counters. Two customers arrive at the counters at different times when the counters are serving no other customers. Each customer chooses a counter at random, independently of the other. Let $X_1$ denote the number of customers who choose counter 1 and $X_2$, the number who select counter 2. Find the joint probability function of $X_1$ and $X_2$.

**Solution.** We might proceed with the derivation in many ways. The most direct is to consider the sample space associated with the experiment. Let the pair $\{i, j\}$ denote the simple event that the first customer chose counter $i$ and the second customer chose counter $j$ , where $i, j = 1, 2,$ and 3. Using the $mn$ rule, the sample space consists of $3 \times 3 = 9$ sample points. Under the assumptions given earlier, each sample point is equally likely and has probability $1/9$. The sample space associated with the experiment is

$$S = [\{1, 1\}, \{1, 2\}, \{1, 3\}, \{2, 1\}, \{2, 2\}, \{2, 3\}, \{3, 1\}, \{3, 2\}, \{3, 3\}].$$

Notice that sample point $\{1, 1\}$ is the only sample point corresponding to $(X_1 = 2, X_2 = 0)$ and hence $\Pr(X_1 = 2, X_2 = 0) = 1/9$. Similarly, $\Pr(X_1 = 1, X_2 = 1) = \Pr(\{1, 2\}$ or $\{2, 1\}) = 2/9$. Table 5.1 contains the probabilities associated with each possible pair of values for $X_1$ and $X_2$—that is, the joint probability function for $X_1$ and $X_2$.

Probability function for $X_1$ and $X_2$

| $x_2 \setminus x_1$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 1/9 | 2/9 | 1/9 |
| 1 | 2/9 | 2/9 | 0 |
| 2 | 1/9 | 0 | 0 |

◀

**EXAMPLE 4.2**

A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let $X =$ the proportion of time that the drive-up facility is in use (at least one customer is being served or waiting to be served) and $Y =$ the proportion of time that the walk-up window is in use. Then the set of possible values for $(X, Y)$ is the rectangle $D\{(x, y) : 0 \le x \le 1;\ 0 \le y \le 1\}$. Suppose the joint pdf of $(X, Y)$ is $f(x, y) = \frac{6}{5}(x+y^2), 0 \le x \le 1, 0 \le y \le 1$. Verify that $f(x, y)$ is a valid pdf. Evaluate $\Pr\left(0 \le x \le \frac{1}{4},\ 0 \le Y \le \frac{1}{4}\right)$.

**Solution.** To verify that this is a legitimate pdf, note that $f(x, y) > 0$ and

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy &= \int_0^1\int_0^1 \frac{6}{5}(x + y^2)dxdy = \int_0^1\int_0^1 \frac{6}{5}xdxdy + \int_0^1\int_0^1 \frac{6}{5}y^2dxdy \\ &= \int_0^1 \frac{6}{5}xdx(1 - 0) + \int_0^1 \frac{6}{5}y^2dy(1 - 0) = \frac{6}{10}x^2\Big|_0^1 + \frac{6}{15}y^3\Big|_0^1 = \frac{6}{10} + \frac{6}{15} = 1 \end{aligned}$$

The probability that neither facility is busy more than one-quarter of the time is

$$\Pr\left(0 \le X \le \frac{1}{4},\ 0 \le Y \le \frac{1}{4}\right) = \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x+y^2)dxdy = \frac{6}{5}\int_0^{1/4}\int_0^{1/4} xdxdy + \frac{6}{5}\int_0^{1/4}\int_0^{1/4} y^2dxdy$$
$$= \frac{6}{5}\left(\frac{1}{4}-0\right)\int_0^{1/4} xdx + \frac{6}{5}\left(\frac{1}{4}-0\right)\int_0^{1/4} y^2dx$$
$$= \frac{6}{20}\frac{x^2}{2}\Big|_0^{1/4} + \frac{6}{20}\frac{y^3}{3}\Big|_0^{1/4} = \frac{7}{640} = 0.0109.$$

◄

**EXAMPLE 4.3**

Gasoline is to be stocked in a bulk tank once at the beginning of each week and then sold to individual customers. Let $Y_1$ denote the proportion of the capacity of the bulk tank that is available after the tank is stocked at the beginning of the week. Because of the limited supplies, $Y_1$ varies from week to week. Let $Y_2$ denote the proportion of the capacity of the bulk tank that is sold during the week. Because $Y_1$ and $Y_2$ are both proportions, both variables take on values between 0 and 1. Further, the amount sold, $y_2$, cannot exceed the amount available, $y_1$. Suppose that the joint density function for $Y_1$ and $Y_2$ is given by: $f(y_1, y_2) = 3y_1$, $0 \le y_2 \le y_1 \le 1$. Find the probability that less than one-half of the tank will be stocked and more than one-quarter of the tank will be sold.

**Solution.** We want to find $\Pr(0 \le Y_1 \le 0.5, Y_2 > 0.25)$. For any continuous random variable, the probability of observing a value in a region is the volume under the density function above the region of interest. We are interested only in values of $y_1$ and $y_2$ such that $0 \le y_1 \le 0.5$ and $y_2 > 0.25$. Thus we have

$$\Pr(0 \le Y_1 \le 0.5, Y_2 > 0.25) = \int_0^{1/2}\int_{1/4}^{y_1} 3y_1 dy_2 dy_1 = \int_0^{1/2} 3y_1 \left(y_2\Big|_{1/4}^{y_1}\right) dy_1$$
$$= \int_0^{1/2} 3y_1(y_1 - 1/4)dy_1 = \left[y_1^3 - (3/8)y_1^2\right]_0^{1/2} = 1/32.$$

◄

### 4.2.2 Joint Cumulative Distribution Function

Let $X_1$ and $X_2$ be any r.v's with joint pdf $f(x_1, x_2)$, then the joint distribution function $F(x_1, x_2)$, such that

$$F(x_1,x_2) = \Pr(X_1 \le x_1, X_2 \le x_2) = \begin{cases} \sum\limits_{\tau_1=-\infty}^{x_1}\sum\limits_{\tau_2=-\infty}^{x_2} f(\tau_1,\tau_2), & \text{discrete} \\ \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2} f(\tau_1,\tau_2)d\tau_1 d\tau_2, & \text{continuous} \end{cases},$$

for all $-\infty < x_1 < \infty$, $-\infty < x_2 < \infty$.

***Theorem*:** If $X_1$ and $X_2$ are r.v's with joint distribution function $F(x_1, x_2)$, then

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$.
2. $F(\infty, \infty) = 1$.
3. If $y_1^* \ge y_1$ and $y_2^* \ge y_2$, then

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \ge 0.$$

**EXAMPLE 4.4**

Consider the r.v's $X_1$ and $X_2$ of Example **??**. Find $F(1,2)$, $F(1.5,2)$, and $F(5,7)$.

**Solution.** Using the results in Example **??**, we find

$$F(-1,2) = \Pr(X_1 \leq -1, X_2 \leq 2) = \Pr(\phi) = 0.$$

Further,

$$\begin{aligned} F(1.5,2) &= \Pr(X_1 \leq 1.5, X_2 \leq 2) \\ &= f(0,0) + f(0,1) + f(0,2) + f(1,0) + f(1,1) + f(1,2) = 8/9. \end{aligned}$$

Similarly,

$$F(5,7) = \Pr(X_1 \leq 5, X_2 \leq 7) = 1.$$

Notice that $F(x_1,x_2) = 1$ for all $x_1, x_2$ such that $\min\{x_1,x_2\} \geq 2$. Also, $F(x_1,x_2) = 0$ if $\min\{x_1,x_2\} < 0$. ◀

### 4.2.3 Marginal Distribution

Recall that the distinct values assumed by a discrete random variable represent mutually exclusive events. Similarly, for all distinct pairs of values $x_1$, $x_2$, the bivariate events $(X_1 = x_1, X_2 = x_2)$, represented by $(x_1,x_2)$, are mutually exclusive events. It follows that the univariate event $(X_1 = x_1)$ is the union of bivariate events of the type $(X_1 = x_1, X_2 = x_2)$, with the union being taken over all possible values for $x_2$. For example, consider the experiment of tossing a pair of dice. The sample space contains 36 sample points, corresponding to the $mn = (6)(6) = 36$ ways in which numbers may appear on the faces of the dice. Consider the following r.v's

$X_1$**:** The number of dots appearing on die 1,

$X_2$**:** The number of dots appearing on die 2.

Then,

$$\begin{aligned} \Pr(X_1 = 1) &= p(1,1) + p(1,2) + p(1,3) + \cdots + p(1,6) \\ &= 1/36 + 1/36 + \cdots + 1/36 = 6/36 = 1/6 \\ \Pr(X_1 = 2) &= p(2,1) + p(2,2) + p(2,3) + \cdots + p(2,6) \\ &= 1/36 + 1/36 + \cdots + 1/36 = 6/36 = 1/6 \\ &\vdots \\ \Pr(X_1 = 6) &= p(6,1) + p(6,2) + p(6,3) + \cdots + p(6,6) \\ &= 1/36 + 1/36 + \cdots + 1/36 = 6/36 = 1/6 \end{aligned}$$

Expressed in summation notation, probabilities about the variable $X_1$ alone are

$$\Pr(X_1 = x_1) = p_1(x_1) = \sum_{y_2=1}^{6} p(x_1,x_2).$$

Similarly, probabilities corresponding to values of the variable $X_2$ alone are given by

$$\Pr(X_2 = x_2) = p_2(x_2) = \sum_{y_1=1}^{6} p(x_1,x_2).$$

Summation in the discrete case corresponds to integration in the continuous case, which leads us to the following definition.

***Definition*:** Let $X_1$ and $X_2$ be jointly r.v's with probability function $f(x_1, x_2)$. Then the marginal probability functions of $X_1$ and $X_2$, respectively, are given by

$$f_1(x_1) = \begin{cases} \sum\limits_{x_2} f(x_1, x_2), & discrete \\ \int\limits_{x_2} f(x_1, x_2)dx_2, & continuous \end{cases}$$

$$f_2(x_2) = \begin{cases} \sum\limits_{x_1} f(x_1, x_2), & discrete \\ \int\limits_{x_1} f(x_1, x_2)dx_1, & continuous \end{cases}$$

***Notes*:** Let $X$ and $Y$ are two r.v's with a joint pdf $f(x, y)$ and marginal pdf's $f_X(x)$ and $f_Y(y)$ respectively. Suppose we require to evaluate

1. $\Pr(a \leq X \leq b)$:

$$\Pr(a \leq X \leq b) = \begin{cases} \sum\limits_{x=a}^{b} f_X(x) = \sum\limits_{x=a}^{b} \sum\limits_{y} f(x, y), & discrete \\ \int\limits_{x=a}^{b} f_X(x)dx = \int\limits_{x=a}^{b} \int\limits_{y} f(x, y)dxdy, & continuous \end{cases}$$

2. $\Pr(c \leq Y \leq d)$:

$$\Pr(c \leq Y \leq d) = \begin{cases} \sum\limits_{y=c}^{d} f_Y(y) = \sum\limits_{y=c}^{d} \sum\limits_{x} f(x, y), & discrete \\ \int\limits_{y=c}^{d} f_Y(y)dy = \int\limits_{y=c}^{d} \int\limits_{x} f(x, y)dxdy, & continuous \end{cases}$$

3. $E[u(X)]$:

$$E[u(X)] = \begin{cases} \sum\limits_{x} u(x)f_X(x) = \sum\limits_{x} \sum\limits_{y} u(x)f(x, y), & discrete \\ \int\limits_{x} u(x)f_X(x)dx = \int\limits_{x} \int\limits_{y} u(x)f(x, y)dxdy, & continuous \end{cases}$$

4. $E[u(Y)]$:

$$E[u(Y)] = \begin{cases} \sum\limits_{y} u(y)f_Y(y) = \sum\limits_{y} \sum\limits_{x} u(y)f(x, y), & discrete \\ \int\limits_{y} u(y)f_Y(y)dy = \int\limits_{y} \int\limits_{x} u(y)f(x, y)dxdy, & continuous \end{cases}$$

5. $E[u(X, Y)]$:

$$E[u(X, Y) = \begin{cases} \sum\limits_{y} \sum\limits_{x} u(x, y)f(x, y), & discrete \\ \int\limits_{x} \int\limits_{y} u(x, y)f(x, y)dydx, & continuous \end{cases}$$

**EXAMPLE 4.5**

Let the joint pdf of r.v's $X$ and $Y$ be:

1. $f(x,y) = \frac{1}{21}(x+y)$, $x = 1, 2, 3$; $y = 1, 2$.
2. $f(x,y) = e^{-(x+y)}$, $0 < x < \infty$; $0 < y < \infty$.

**a.** Find the marginal pdf for both $X$ and $Y$.

**b.** Find $E[X]$, $Var(X)$, $E[Y]$, $Var(Y)$, $E[XY]$.

**Solution.** For number (1) we have:

**a.** The marginal pdf of $X$ is

$$f_x(x) = \sum_y f(x,y) = \sum_{y=1}^{2} \frac{1}{21}(x+y) = \frac{1}{21}[(x+1)+(x+2)] = \frac{1}{21}(2x+3), \quad x = 1, 2, 3.$$

Also, the marginal pdf of $Y$ is

$$f_y(y) = \sum_x f(x,y) = \sum_{x=1}^{3} \frac{1}{21}(x+y) = \frac{1}{21}[(y+1)+(y+2)+(y+3)] = \frac{1}{7}(y+2), \quad x = 1, 2.$$

**b.** To find the expectations,

$$\mu_x = E[X] = \sum_x f_x(x) = \frac{1}{21}\sum_{x=1}^{3} x(2x+3) = \frac{46}{21}.$$

$$E[X^2] = \sum_x x^2 f_x(x) = \frac{1}{21}\sum_{x=1}^{3} x^2(2x+3) = \frac{114}{21}.$$

$$\sigma_x^2 = Var(X) = E[X^2] - \mu_x^2 = \frac{114}{21} - \left(\frac{46}{21}\right)^2 = \frac{278}{441}.$$

Similarly,

$$\mu_y = E[Y] = \sum_y f_y(y) = \frac{1}{7}\sum_{y=1}^{2} y(y+2) = \frac{11}{7}.$$

$$E[Y^2] = \sum_y y^2 f_y(y) = \frac{1}{7}\sum_{y=1}^{2} y^2(y+2) = \frac{19}{7}.$$

$$\sigma_y^2 = Var(Y) = E[Y^2] - \mu_y^2 = \frac{19}{7} - \left(\frac{11}{7}\right)^2 = \frac{12}{49}.$$

Finally,

$$E[XY] = \sum_x \sum_y xyf(x,y) = \frac{1}{21}\sum_{x=1}^{3}\sum_{y=1}^{2} xy(x+y) = \frac{1}{21}\sum_{x=1}^{3}[x(x+1)+x(x+2)]$$

$$= \frac{1}{21}\sum_{x=1}^{3}(3x^2+5x) = \frac{1}{21}(8+29+42) = \frac{24}{7}.$$

For number (2) we have:

**a.** The marginal pdf of $X$ is

$$f_x(x) = \int_y f(x,y)dy = \int_0^\infty e^{-(x+y)}dy = -e^{-x}\left[e^{-y}\right]_0^\infty = e^{-x}, \ \ 0 < x < \infty.$$

Also, the marginal pdf of $Y$ is

$$f_y(y) = \int_x f(x,y)dx = \int_0^\infty e^{-(x+y)}dx = -e^{-y}\left[e^{-x}\right]_0^\infty = e^{-y}, \ \ 0 < y < \infty.$$

**b.** To find the expectations,

$$\mu_x = E[X] = \int_x x f_x(x)dx = \int_0^\infty x e^{-x}dx = 1 = E[Y] = \mu_y.$$

$$E[X^2] = \int_x x^2 f_x(x)dx = \int_0^\infty x^2 e^{-x}dx = 2 = E[Y^2].$$

$$\sigma_x^2 = Var(X) = E[X^2] - \mu_x^2 = 2 - (1)^2 = 1 = \sigma_y^2 = Var(Y)$$

Finally,

$$\begin{aligned} E[XY] &= \int_x \int_y xy f(x,y)dxdy = \int_0^\infty \int_0^\infty xy e^{-(x+y)}dxdy \\ &= \left(\int_0^\infty x e^{-x}dx\right)\left(\int_0^\infty y e^{-y}dy\right) = (1)(1) = 1. \end{aligned}$$

◀

***Definition*:** If $f(x,y)$ and $F(x,y)$ are the joint pdf and cdf of r.v's $X$ and $Y$. Then the marginal cdf of $X$ and $Y$, respectively , are

$$F_x(x) = F(x,\infty) \text{ and } F_y(y) = F(\infty, y).$$

and are defined as

$$F_x(x) = \begin{cases} \sum\limits_{\tau=-\infty}^{x} f_x(\tau), & discrete, \\ \int\limits_{-\infty}^{x} f_x(\tau)d\tau, & continuous. \end{cases}$$

and

$$F_y(y) = \begin{cases} \sum\limits_{\tau=-\infty}^{y} f_y(\tau), & discrete, \\ \int\limits_{-\infty}^{y} f_y(\tau)d\tau, & continuous. \end{cases}$$

**EXAMPLE 4.6**

Back to the last example and find the joint cdf of $X$ and $Y$, then find the marginal cdf for both $X$ and $Y$.

**Solution.** For number (1) we have $f(x, y) = \frac{1}{21}(x+y)$, $x = 1, 2, 3$; $y = 1, 2$. Then

$$\begin{aligned} F(x,y) &= \sum_{\tau=1}^{x}\sum_{\kappa=1}^{y} f(\tau,\kappa) = \frac{1}{21}\sum_{\tau=1}^{x}\sum_{\kappa=1}^{y}(\tau+\kappa) = \frac{1}{21}\sum_{\tau=1}^{x}[(\tau+1)+(\tau+2)+\cdots+(\tau+y)] \qquad (4.1)\\ &= \frac{1}{21}\sum_{\tau=1}^{x}\left[y\tau+\frac{y(y+1)}{2}\right] = \frac{1}{21}\left[\frac{yx(x+1)}{2}+\frac{xy(y+1)}{2}\right] = \frac{1}{42}xy(x+y+2), \end{aligned}$$

hence

$$F(x,y) = \begin{cases} 0, & x<1,\ y<1 \\ \frac{1}{42}xy(x+y+2), & 1\le x<3,\ 1\le y<2 \\ 1, & x\ge 3,\ y\ge 2 \end{cases}$$

The marginal cdf of $X$ is

$$F_x(x) = F(x,2) = \begin{cases} 0, & x<1 \\ \frac{1}{21}x(x+4), & 1\le x<3 \\ 1, & x\ge 3 \end{cases}$$

Similarly, the marginal cdf of $Y$ is

$$F_y(y) = F(3,y) = \begin{cases} 0, & y<1 \\ \frac{1}{14}y(y+5), & 1\le y<2 \\ 1, & y\ge 2 \end{cases}$$

For number (2) we have $f(x,y) = e^{-(x+y)}$, $0<x<\infty$; $0<y<\infty$. Then

$$\begin{aligned} F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(\tau,\kappa)d\tau d\kappa = \int_{0}^{x}\int_{0}^{y} e^{-(\tau+\kappa)}d\tau d\kappa = \left(\int_0^x e^{-\tau}d\tau\right)\left(\int_0^y e^{-\kappa}d\kappa\right)\\ &= \left[-e^{-\tau}\right]_0^x\left[-e^{-\kappa}\right]_0^y, \end{aligned}$$

hence

$$F(x,y) = \begin{cases} 0, & x\le 0,\ y\le 0 \\ (1-e^{-x})(1-e^{-y}), & 0<x<\inf,\ 0<y<\infty \\ 1, & x=y=\infty \end{cases}$$

The marginal cdf of $X$ is

$$F_x(x) = F(x,\infty) = \begin{cases} 0, & x\le 0 \\ 1-e^{-x}, & 0<x<\inf \\ 1, & x=\infty \end{cases}$$

The marginal cdf of $Y$ is

$$F_y(y) = F(\infty,y) = \begin{cases} 0, & y\le 0 \\ 1-e^{-y}, & 0<y<\inf \\ 1, & y=\infty \end{cases}$$

◄

***Remarks*:**

1. The joint moment generating function (mgf) of the two r.v's $X$ and $Y$ with a joint pdf $f(x,y)$ is defined as

$$M(t_1,t_2) = E[e^{t_1X+t_2Y}] = \begin{cases} \sum\limits_x \sum\limits_y e^{t_1X+t_2Y} f(x,y), & discrete \\ \int\limits_x \int\limits_y e^{t_1X+t_2Y} f(x,y)dxdy, & continuous \end{cases}$$

   and satisfy $M(0,0) = 1$.

2. The marginal mgf of $X$ and $Y$ can be obtained by

$$M_X(t_1) = E[e^{t_1X}] = M(t_1,0), \quad \text{and} \quad M_Y(t_2) = E[e^{t_2Y}] = M(0,t_2).$$

3. We can evaluate the mean and variance for both $X$ and $Y$ using the mgf as

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

### 4.2.4 Conditional Distribution

The distribution of $Y$ can depend strongly on the value of another variable $X$. For example, if $X$ is height and $Y$ is weight, the distribution of weight for men who are 180 cm tall is very different from the distribution of weight for short men. The conditional distribution of $Y$ given $X = x$ describes for each possible $x$ how probability is distributed over the set of possible $y$ values. We define the conditional distribution of $Y$ given $X$, but the conditional distribution of $X$ given $Y$ can be obtained by just reversing the roles of $X$ and $Y$. Both definitions are analogous to that of the conditional probability $\Pr(A|B)$ as the ratio $\Pr(A \cup B)/\Pr(B)$.

***Definition*:** If $X$ and $Y$ are jointly r.v's with joint density function $f(x,y)$, then the conditional pdf of $X$ given $Y = y$ is

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

also, conditional pdf of $Y$ given $X = x$ is

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

***Definition*:** If $X$ and $Y$ are jointly r.v's with joint density function $f(x,y)$, then the conditional distribution function of $X$ given $Y = y$ and $Y$ given $X = x$ are

$$F(x|y) = \Pr(X \le x|Y = y), \quad \text{and} \quad F(y|x) = \Pr(Y \le y|X = x).$$

**EXAMPLE 4.7**

Let the joint pdf of r.v's $X$ and $Y$ be $f(x,y) = e^{-y}$, $0 < x < y < \infty$. Find the following

1. The conditional pdf of both $X$ given $Y$ and $Y$ given $X$.
2. $E[X|Y = y]$, $Var(X|Y = y)$, $E[Y|X = x]$ and $Var[Y|X = x]$.
3. Evaluate the probabilities: $\Pr(1 \le X \le 2)$ and $\Pr(1 \le X \le 2|Y = 2)$.

**Solution.** We have the joint pdf $f(x,y) = e^{-y}$, $0 < x < y < \infty$, then the marginal pdf of $X$ and $Y$ are

$$f_X(x) = \int_y f(x,y)dy = \int_x^\infty e^{-y}dy = -e^{-y}\Big|_x^\infty = e^{-x}, \quad 0 < x < \infty.$$

and

$$f_Y(y) = \int_x f(x,y)dx = \int_0^y e^{-y}dy = e^{-y}x\Big|_0^y = ye^{-y}, \quad 0 < y < \infty.$$

1. The conditional pdf of $X$ given $Y = y$ is

$$f_1(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y}, \quad 0 < x < y, \text{ for any } 0 < y < \infty.$$

The conditional pdf of $Y$ given $X = x$ is

$$f_2(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}, \quad x < y < \infty, \text{ for any } 0 < x < \infty.$$

2. To evaluate the expectations:

$$E[X|Y = y] = \mu_{x|y} = \int_x xf_1(x|y)dx = \int_0^y x\frac{1}{y}dx = \frac{1}{y}\left[\frac{x^2}{2}\right]_0^y = \frac{y}{2}.$$

$$E[X^2|Y = y] = \int_x x^2 f_1(x|y)dx = \int_0^y x^2\frac{1}{y}dx = \frac{1}{y}\left[\frac{x^3}{3}\right]_0^y = \frac{y^2}{3}.$$

$$\therefore \sigma^2_{x|y} = E[X^2|Y = y] - \mu^2_{x|y} = \frac{y^2}{3} - \left(\frac{y}{2}\right)^2 = \frac{y^2}{12}.$$

Similarly,

$$E[Y|X = x] = \mu_{y|x} = \int_y yf_2(y|x)dy = \int_x^\infty ye^{-(y-x)}dx = e^x\int_x^\infty ye^{-y}dy$$
$$= e^x\left[-ye^{-y} - e^{-y}\right]_x^\infty = e^x(xe^{-x} + e^{-x}) = x + 1.$$

$$E[Y^2|X = x] = \int_y y^2 f_2(y|x)dy = \int_x^\infty y^2e^{-(y-x)}dx = e^x\int_x^\infty y^2e^{-y}dy$$
$$= e^x\left[-y^2e^{-y} - 2ye^{-y} - 2e^{-y}\right]_x^\infty = e^x(x^2e^{-x} + 2xe^{-x} + 2e^{-x}) = x^2 + 2x + 2.$$

$$\sigma_{y|x} = E[Y^2|X = x] - \mu^2_{y|x} = x^2 + 2x + 2 - (x+1)^2 = 1.$$

3. To Evaluate the probabilities:

$$\Pr(1 \le X \le 2) = \int_1^2 f_X(x)dx = \int_1^2 e^{-x}dx = -e^{-x}\Big|_1^2 = e^{-1} - e^{-2}.$$

$$\Pr(1 \le X \le 2|Y = 2) = \int_1^2 f_1(x|y = 2)dx = \int_1^2 \frac{1}{2}dx = \frac{x}{2}\Big|_1^2 = \frac{1}{2}(2-1) = \frac{1}{2}.$$

◄

**EXAMPLE 4.8**

A soft-drink machine has a random amount $Y$ in supply at the beginning of a given day and dispenses a random amount $X$ during the day (with measurements in gallons). It is not resupplied during the day, and hence $X \leq Y$. It has been observed that $X$ and $Y$ have a joint density function $f(x, y) = \frac{1}{2}$, $0 \leq x \leq y \leq 2$. That is, the points $(x, y)$ are uniformly distributed over the triangle with the given boundaries. Find the conditional density of $X$ given $Y = y$. Evaluate the probability that less than 1/2 gallon will be sold, given that the machine contains 1.5 gallons at the start of the day.

**Solution.** The marginal density of $Y$ is given by

$$f_Y(y) = \int_x f(x, y)dx = \int_0^y \frac{1}{2}dx = \frac{y}{2}, \quad 0 \leq y \leq 2.$$

Note that $f_Y(y) > 0$ if and only if $0 < y2$. Thus, for any $0 < y \leq 2$,

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{1/2}{y/2} = \frac{1}{y}, \quad 0 \leq x \leq y.$$

Also, $f(x|y)$ is undefined if $y \leq 0$ or $y > 2$. The probability of interest is

$$\Pr(X \leq 1|Y = 1.5) = \int_0^{1/2} f(x|y = 1.5)dx = \int_0^{1/2} \frac{1}{1.5}dx = \frac{1/2}{3/2} = \frac{1}{3}.$$

◀

**EXAMPLE 4.9**

Let the r.v $X$ has a pdf $f(x)$ and cdf $F(x)$. Define $f(x|X > x_0) = \frac{f(x)}{1-F(x_0)}$, $x_0 < x < \infty$ and $x_0$ is a fixed number.

1. Show that $f(x|X > x_0)$ be a conditional pdf of $X$ given $X > x_0$.
2. Consider $f(x) = e^{-x}$, $0 < x < \infty$. Compute $\Pr(X > 2|X > 1)$.

**Solution.** 1. To prove that $f(x|X > x_0)$ is a pdf, we check

- $f(x|X > x_0) \geq 0$, because $f(x) \geq 0$ and $1 - F(x_0) > 0$.
- The unity of integration,

$$\begin{aligned}\int_{x_0}^{\infty} f(x|X > x_0)dx &= \int_{x_0}^{\infty} \frac{f(x)}{1 - F(x_0)}dx = \frac{1}{1 - F(x_0)} \int_{x_0}^{\infty} f(x)dx = \frac{1}{1 - F(x_0)} [F(x)]_{x_0}^{\infty} \\ &= \frac{F(\infty) - F(x_0)}{1 - F(x_0)} = \frac{1 - F(x_0)}{1 - F(x_0)} = 1.\end{aligned}$$

2. Since we have the pdf $f(x) = e^{-x}$ and the cdf is defined as: $F(x) = 1 - e^{-x}$ (exponential distribution). Therefore,

$$f(x|X > 1) = \frac{f(x)}{1 - F(1)} = \frac{e^{-x}}{1 - 1 + e^{-1}} = e \cdot e^{-x}, \quad 1 < x < \infty.$$

Now, to compute the probability

$$\Pr(X > 2|X > 1) = \int_2^{\infty} f(x|X > 1)dx = \int_2^{\infty} e \cdot e^{-x}dx = -e\left[e^{-x}\right]_2^{\infty} = -e(0 - e^{-2}) = e^{-1}.$$

◀

### 4.2.5 Independent Random Variables

In many situations, information about the observed value of one of the two variables $X$ and $Y$ gives information about the value of the other variable. In Example **??** we saw two dependent r.v's, for which probabilities associated with $X$ depended on the observed value of $Y$. In other examples, this is not always the case. Probabilities associated with $X$ were the same, regardless of the observed value of $Y$. We now present a formal definition of independence of r.v's. Two events $A$ and $B$ are independent if $\Pr(A \cup B) = \Pr(A) \times \Pr(B)$.

***Definition*:** Two random variables $X$ and $Y$ with joint density function $f(x, y)$ and marginal density functions $f_X(x)$ and $f_Y(y)$, are said to be independent if for every pair of $x$ and $y$ values,

$$f(x, y) = f_X(x) \cdot f_Y(y), \quad \forall (x, y).$$

we can also define the independence in terms of the joint cdf and the marginal cdf's $F_X(x)$ and $F_Y(y)$ of $X$ and $Y$, then

$$F(x, y) = F_X(x) \cdot F_Y(y),$$

for every pair of real numbers $(x, y)$.

For the die-tossing problem of Section **??**, show that $X_1$ and $X_2$ are independent since each of the 36 sample points was given probability 1/36. Consider, for example, the point (1, 2). We know that $p(1, 2) = 1/36$. Also, $p_1(1) = \Pr(X_1 = 1) = 1/6$ and $p_2(2) = \Pr(X_2 = 2) = 1/6$. Hence, $p(1, 2) = p_1(1)p_2(2)$. The same is true for all other values for $x_1$ and $x_2$, and it follows that $X_1$ and $X_2$ are independent.

***Note*:** If two r.v's $X$ and $Y$ are independent, we have

$$f_1(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x) \cdot f_Y(y)}{f_Y(y)} = f_X(x).$$

and

$$f_2(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{f_X(x) \cdot f_Y(y)}{f_X(x)} = f_Y(y).$$

**EXAMPLE 4.10**

Back to example **??** and examine whether or not $X$ and $Y$ are independent.

**Solution.** In example **??** we have two cases:

1. The first case where $X$ and $Y$ are two discrete r.v's with joint pdf $f(x, y) = \frac{1}{21}(x + y)$, $x = 1, 2, 3$; $y = 1, 2$. and marginal pdf's $f_X(x) = \frac{1}{21}(2x + 3)$, $x = 1, 2, 3$ and $f_Y(y) = \frac{1}{7}(y + 2)$, $y = 1, 2$. In order to check if $X$ and $Y$ are independent,
$$f_X(x) \cdot f_Y(y) = \frac{1}{217}(2x + 3)(y + 2) \neq \frac{1}{21}(x + y) = f(x, y).$$
Therefore $X$ and $Y$ are dependent r.v's.

2. The second case where $X$ and $Y$ are two continuous r.v's with joint pdf $f(x, y) = e^{-(x+y)}$, $0 < x, y < \infty$, and marginal pdf's $f_X(x) = e^{-x}$, $0 < x < \infty$ and $f_Y(y) = e^{-y}$, $0 < y < \infty$. In order to check if $X$ and $Y$ are independent,
$$f_X(x) \cdot f_Y(y) = e^{-x} \cdot e^{-y} = e^{-(x+y)} = f(x, y),$$
this implies that $X$ and $Y$ are independent.

◀

***Properties*:** Let $X$ and $Y$ are two independent r.v's, then:

1. $\Pr(a \le X \le b, c \le Y \le d) = \Pr(a \le X \le b) \cdot \Pr(c \le Y \le d)$.
2. $E[u(X) \cdot v(Y)] = E[u(X)] \cdot E[v(Y)]$.
3. $M(t_1, t_2) = M_X(t_1) \cdot M_Y(t_2)$.

**EXAMPLE 4.11**

Let the joint mgf of r.v's $X$ and $Y$ is: $M(t_1, t_2) = \frac{e^{\frac{t_1^2}{1-2t_2}}}{1-2t_2}$. Find the marginal mgf's for both $X$ and $Y$ and test whether $X$ and $Y$ are independent.

**Solution.** We can simply find the marginal mgf's for $X$ and $Y$ as

$$M_X(t_1) = M(t_1, 0) = e^{t_1^2}, \quad \text{and} \quad M_Y(t_2) = M(0, t_2) = (1 - 2t_2)^{-1}.$$

In order to test if $X$ and $Y$ are independent,

$$M_X(t_1) \cdot M_Y(t_2) = \frac{e^{t_1^2}}{1 - 2t_2} \neq M(t_1, t_2),$$

therefore $X$ and $Y$ are dependent. ◀

***theorem*:** Let $X$ and $Y$ have a joint density $f(x, y)$, Then $X$ and $Y$ are independent r.v's if and only if

$$f(x, y) = g(x)h(y)$$

where $g(x)$ is a non-negative function of $x$ alone and $h(y)$ is a non-negative function of $y$ alone.

**EXAMPLE 4.12**

Let $X$ and $Y$ have a joint density given by: $f(x, y) = 2x$, $0 \le x \le 1$, $0 \le y \le 1$. Are $X$ and $Y$ independent variables?

**Solution.** Notice that $f(x, y)$ is positive for $0 \le x \le 1$ and $0 \le y \le 1$. Further, $f(x, y) = g(x)h(y)$, where

$$g(x) = x, \quad 0 \le x \le 1, \quad \text{and} \quad h(y) = 2, 0 \le y \le 1.$$

Therefore, $X$ and $Y$ are independent r.v's. Notice that $g(x)$ and $h(y)$, as defined here, are not density functions, although $2g(x)$ and $h(y2)/2$ are densities. ◀

## 4.3 Expected Values, Covariance, and Correlation

We previously saw that any function $u(X)$ of a single r.v $X$ is itself a r.v. However, to compute $E[u(X)]$, it was not necessary to obtain the probability distribution of $u(X)$; instead, $E[u(X)]$ was computed as a weighted average of $u(X)$ values, where the weight function was the pdf $f(x)$ of $X$. A similar result holds for a function $u(X, Y)$ of two jointly distributed r.v's.

**EXAMPLE 4.13**

Let $X$ and $Y$ have joint density given by: $f(x,y) = 2x, 0 \leq x \leq 1; 0 \leq y \leq 1$. Find $E[XY]$, $E[X]$ and $Var(X)$.

**Solution.** From expectation definition we obtain

- The expected value of $E[XY]$ is:

$$\begin{aligned} E[XY] &= \int_0^1 \int_0^1 xyf(x,y)dxdy = \int_0^1 \int_0^1 xy(2x)dxdy = \int_0^1 y\left[\frac{2x^3}{3}\right]_0^1 dy \\ &= \frac{2}{3}\int_0^1 ydy = \frac{2}{3}\left[\frac{y^2}{2}\right]_0^1 = \frac{1}{3}. \end{aligned}$$

- The expected value of $E[X]$ is:

$$\begin{aligned} E[X] &= \int_0^1 \int_0^1 xf(x,y)dxdy = \int_0^1 \int_0^1 x(2x)dxdy = \int_0^1 \left[\frac{2x^3}{3}\right]_0^1 dy \\ &= \frac{2}{3}\int_0^1 dy = \frac{2}{3}\left[y\right]_0^1 = \frac{2}{3}. \end{aligned}$$

- To find $Var(X)$ we need to find $E[X^2]$ first, then:

$$\begin{aligned} E[X^2] &= \int_0^1 \int_0^1 x^2f(x,y)dxdy = \int_0^1 \int_0^1 x^2(2x)dxdy = \int_0^1 \left[\frac{2x^4}{4}\right]_0^1 dy \\ &= \frac{1}{2}\int_0^1 dy = \frac{1}{2}\left[y\right]_0^1 = \frac{1}{2}. \end{aligned}$$

Therefore,

$$Var(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

◀

**EXAMPLE 4.14**

A process for producing an industrial chemical yields a product containing two types of impurities. For a specified sample from this process, let $X$ denote the proportion of impurities in the sample and let $Y$ denote the proportion of type I impurities among all impurities found. Suppose that the joint distribution of $X$ and $Y$ can be modelled by the pdf: $f(x,y) = 2(1-x), 0 \leq x \leq 1; 0 \leq y \leq 1$. Find the expected value of the proportion of type I impurities in the sample.

**Solution.** Because $X$ is the proportion of impurities in the sample and $Y$ is the proportion of type I impurities among the sample impurities, it follows that $XY$ is the proportion of type I impurities in the entire sample. Thus, we want to find $E[XY]$:

$$\begin{aligned} E[XY] &= \int_0^1 \int_0^1 2xy(1-x)dxdy = 2\int_0^1 x(1-x)\left(\frac{1}{2}\right)dx = \int_0^1 (x - x^2)dx \\ &= \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}. \end{aligned}$$

Therefore, we would expect 1/6 of the sample to be made up of type I impurities. ◀

### 4.3.1 The Covariance and Correlation Coefficient

Intuitively,we think of the dependence of two random variables $X$ and $Y$ as implying that one variable, say $X$, either increases or decreases as $Y_2$ changes. We will confine our attention to two measures of dependence: the covariance between two random variables and their correlation coefficient.

***4.3.1.1 The Covariance*** When two random variables $X$ and $Y$ are not independent, it is frequently of interest to assess how strongly they are related to each other. The average value of $(X-\mu_x)(Y-\mu_y)$ provides a measure of the linear dependence between $X$ and $Y$. This quantity, $E[(X-\mu_x)(Y-\mu_y)]$, is called the covariance of $X$ and $Y$.

***Definition*:** If $X$ and $Y$ are r.v's with means $\mu_x$ and $\mu_y$ and variances $\sigma_x^2$ and $\sigma_y^2$, respectively, the covariance of $X$ and $Y$ is defined as

$$Cov(X,Y)=E[(X-\mu_x)(Y-\mu_y)].$$

The larger the absolute value of the covariance of $X$ and $Y$, the greater the linear dependence between $X$ and $Y$. Positive values indicate that $X$ increases as $Y$ increases; negative values indicate that $X$ decreases as $Y$ increases. A zero value of the covariance indicates that the variables are uncorrelated and that $X$ and $Y$ are independent.

***Note*:** From the definition of the covariance, we can find the relation between the variances and covariance of the two r.v's $X$ and $Y$ as the following:

$$\begin{aligned}Cov(X,Y)&=E[(X-\mu_x)(Y-\mu_y)]=E[XY-\mu_xY-\mu_yX+\mu_x\mu_y]=E[XY]-\mu_xE[Y]-\mu_yE[X]+\mu_x\mu_y\\&=E[XY]-\mu_x\mu_y-\mu_y\mu_x+\mu_x\mu_y=E[XY]-\mu_x\mu_y\end{aligned}$$

**EXAMPLE 4.15**

Let the joint pdf of r.v's $X$ and $Y$ is defined as $f(x,y)=x+y$, $0<x<1$, $0<y<1$. Compute the covariance of $X$ and $Y$.

**Solution.** In order to evaluate the covariance of $X$ and $Y$, we need to calculate first $\mu_x$ and $\mu_y$. So lets start with computing the marginal pdf's for both $X$ and $Y$. Then,

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

and

$$f_Y(y)=y+\frac{1}{2},\quad 0<y<1.$$

Secondly, we have to evaluate $\mu_x$ and $\mu_y$, thus

$$\mu_x=E[X]=\int_x xf_X(x)dx=\int_0^1 x\,(x+1/2)\,dx=\left[\frac{x^3}{3}+\frac{x^2}{4}\right]_0^1=\frac{7}{12}.$$

Similarly, $\mu_y=\frac{7}{12}$. Now the expected value of $XY$ is calculated by

$$\begin{aligned}E[XY]&=\int_x\int_y xyf(x,y)dxdy=\int_0^1\int_0^1 xy(x+y)dxdy=\int_0^1\int_0^1(x^2y+xy^2)dxdy\\&=\int_0^1\left[\frac{x^2y^2}{2}+\frac{xy^3}{3}\right]_0^1dx=\int_0^1\left(\frac{x^2}{2}+\frac{x}{3}\right)dx=\left[\frac{x^3}{6}+\frac{x^2}{6}\right]_0^1=\frac{1}{3}.\end{aligned}$$

Therefore, the covariance of $X$ and $Y$ is

$$Cov(X,Y)=E[XY]-\mu_x\mu_y=\frac{1}{3}-\left(\frac{7}{12}\right)\left(\frac{7}{12}\right)=-\frac{1}{144}.$$

◀

**EXAMPLE 4.16**

A nut company markets cans of deluxe mixed nuts containing almonds, cashews, and peanuts. Suppose the net weight of each can is exactly 1 lb, but the weight contribution of each type of nut is random. Because the three weights sum to 1, a joint probability model for any two gives all necessary information about the weight of the third type. Let $X =$ the weight of almonds in a selected can and $Y =$ the weight of cashews. Then the region of positive density is $D = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1; x + y \le 1\}$. Now let the joint pdf for $(X, Y)$ be $f(x, y) = 24xy$, $0 \le x \le 1$, $0 \le y \le 1$; $x + y \le 1$. Find the probability that the two types of nuts together make up at most 50% of the can. Find The marginal pdf for almonds. Evaluate the expected amount of almonds and of cashews at each can. Then find the covariance of $X$ and $Y$.

**Solution.** - Let $A = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1; x + y \le 0.5\}$, and we need to calculate the probability of event $A$, that is

$$\Pr\{(X, Y) \in A\} = \iint_A f(x, y)dxdy = \int_0^{0.5} \int_0^{0.5-x} 24xydydx = 0.0625.$$

- The marginal pdf for almonds is obtained by holding $X$ fixed at $x$ and integrating $f(x, y)$ along the vertical line through $x$:

$$f_X(x) = \int_0^{1-x} f(x, y)dy = \int_0^{1-x} 24xydy = 12x(1 - x)^2, \quad 0 \le x \le 1.$$

- The expected amount of almond at each can is obtained as:

$$\mu_x = E[X] = \int_0^1 xf_X(x)dx = \int_0^1 12x^2(1 - x)^2dx = \frac{2}{5}.$$

Similarly, the expected amount of cashews at these cans can be easily evaluated by by replacing $x$ by $y$ in $f_X(x)$. Then we get $\mu_y = \frac{2}{5}$.

- Finally, to examine how those two r.v's are related to each other, we calculate the covariance of $X$ and $Y$. Then

$$E[XY] = \int_0^1 \int_0^{1-x} xy24xydydx = \frac{2}{15}.$$

Thus,

$$Cov(X, Y) = E[XY] - \mu_x\mu_y = \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = \frac{2}{15} - \frac{4}{25} = -\frac{2}{75}.$$

A negative covariance is reasonable here because more almonds in the can implies fewer cashews.

◀

***Note*:** If $X$, $Y$ and $Z$ are r.v's and $a$ and $b$ are constants, then

$$Cov(aX + bY, Z) = aCov(X, Z) + bCov(Y, Z).$$

***4.3.1.2 Correlation Coefficient*** It is difficult to employ the covariance as an absolute measure of dependence because its value depends upon the scale of measurement. This problem can be eliminated by standardizing its value and using the correlation coefficient, $\rho$, a quantity related to the covariance and defined as

$$\rho = \frac{Cov(X, Y)}{\sigma_x\sigma_y}.$$

where $\sigma_x$ and $\sigma_y$ are the standard deviations of $X$ and $Y$, respectively. The sign of the correlation coefficient is the same as the sign of the covariance. Thus, $\rho > 0$ indicates that $Y$ increases as $X$ increases, and $\rho = +1$ implies perfect correlation, with all points falling on a straight line with positive slope. A value of $\rho = 0$ implies zero covariance and no correlation. A negative coefficient of correlation implies a decrease in $Y$ as $X$ increases, and $\rho = -1$ implies perfect correlation, with all points falling on a straight line with negative slope.

**EXAMPLE 4.17**

From example **??**, find the correlation coefficient.

**Solution.** To calculate the correlation coefficient, we need to calculate the standard deviation for both $X$ and $Y$, and therefore $E[X^2]$ and $E[Y^2]$, then

$$E[X^2] = \int_x x^2 f_X(x)dx = \int_0^1 x^2\left(x+\frac{1}{2}\right)dx = \left[\frac{x^4}{4}+\frac{x^3}{6}\right]_0^1 = \frac{5}{12}.$$

Hence

$$\sigma_x^2 = E[X^2] - \mu_x^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

Similarly, $\sigma_y^2 = \frac{11}{144}$. Therefore,

$$\rho = \frac{Cov(X,Y)}{\sigma_x\sigma_y} = \frac{-1/144}{\sqrt{11/144}\sqrt{11/144}} = -\frac{1}{11}.$$

◀

***Theorem*:** If $X$ and $Y$ are independent r.v's, the $\rho = 0$. However, if but $\rho = 0$ does not imply independence.

**EXAMPLE 4.18**

Let $X$ and $Y$ be discrete r.v's with joint probability distribution as shown in the table below. Show that $X$ and $Y$ are dependent but have zero covariance.

Probability function for $X$ and $Y$

| $y \setminus x$ | -1 | 0 | 1 |
|---|---|---|---|
| -1 | 1/16 | 3/16 | 1/16 |
| 0 | 3/16 | 0 | 3/16 |
| 1 | 1/16 | 3/16 | 1/16 |

**Solution.** Calculating the marginal pdf's yields, $f_X(-1) = f_X(1) = 5/16 = f_Y(-1) = f_Y(1)$, and $f_X(0) = 6/16 = f_Y(0)$. Then, we obtain the following marginal pdf's for both $X$ and $Y$:

Marginal pdf of $X$ and $Y$

| $x$ | $f_X(x)$ | | $y$ | $f_Y(y)$ |
|---|---|---|---|---|
| -1 | 5/16 | | -1 | 5/16 |
| 0 | 6/16 | | 0 | 6/16 |
| 1 | 5/16 | | 1 | 5/16 |

The value $f(0,0) = 0$ in the centre cell stands out. Obviously, $f(0,0) \neq f_X(0)f_Y(0)$, and this is sufficient to show that $X$ and $Y$ are dependent.

Again looking at the marginal probabilities, we see that $E[X] = E[Y] = 0$. Also,

$$\begin{aligned}
E[XY] &= \sum_x\sum_y xyf(x,y) \\
&= (-1)(0)(3/16) + (-1)(1)(1/16) + (0)(-1)(3/16) \\
&+ (0)(0)(0) + (0)(1)(3/16) + (1)(-1)(1/16) + (1)(0)(3/16) + (1)(1)(1/16) \\
&= (1/16) - (1/16) - (1/16) + (1/16) = 0
\end{aligned}$$

Thus,

$$Cov(X, y) = E[XY] - E[X]E[Y] = 0 - (0)(0) = 0$$

This example shows that if the covariance of two r.v's is zero, the variables need not be independent. ◀

**EXAMPLE 4.19**

Annie and Alvie have agreed to meet for lunch between noon (0:00 p.m.) and 1:00 p.m. Denote Annie's arrival time by $X$, Alvie's by $Y$, and suppose $X$ and $Y$ are independent with pdf's:

$$f_X(x) = 3x^2, \quad 0 \le x \le 1: \quad \text{and} \quad f_Y(y) = 2y, \quad 0 \le y \le 1.$$

What is the expected amount of time that the one who arrives first must wait for the other person?

**Solution.** First of all, the joint pdf of $X$ and $Y$ is found by:

$$f(x, y) = f_X(x) \cdot f_Y(y) = 3x^2 \times 2y = 6x^2 y$$

Assuming that $X$ and $Y$ are the arrival time of Annie and Alive respectively give as that the waiting time of the one who comes first is $u(X, Y) = |X - Y|$, then the expected waiting time is $E[|X - Y|]$, and since

$$u(X,Y) = |X - Y| = \begin{cases} X - Y, & x - y \ge 0 \Rightarrow x \ge y \Rightarrow 0 \le y \le x \le 1, \\ Y - X, & x - y < 0 \Rightarrow x < y \Rightarrow 0 < x < y < 1, \end{cases}$$

and then

$$\begin{aligned} E[|X - Y|] &= \int_0^1 \int_0^x (x - y)(6x^2 y) dy dx + \int_0^1 \int_0^y (y - x)(6x^2 y) dx dy \\ &= \frac{1}{6} + \frac{1}{12} = \frac{1}{4}. \end{aligned}$$

Therefore, the expected waiting time for anyone who comes first is 1/4 of an hour. ◀

# CHAPTER 5

# FUNCTION OF RANDOM VARIABLES

## 5.1 Introduction

In Chapter **??** we discussed the problem of starting with a single r.v $X$, forming some function of $X$, such as $X^2$ or $e^X$, to obtain a new r.v $Y = h(X)$, and investigating the distribution of this new r.v. We now generalize this scenario by starting with more than a single r.v. Consider as an example a system having a component that can be replaced just once before the system itself expires. Let $X_1$ denote the lifetime of the original component and $X_2$ the lifetime of the replacement component. Then any of the following functions of $X_1$ and $X_2$ may be of interest to an investigator:

1. The total lifetime $X_1 + X_2$.

2. The ratio of lifetimes $X_1/X_2$; for example, if the value of this ratio is 2, the original component lasted twice as long as its replacement.

3. The ratio $X_1/(X_1 + X_2)$, which represents the proportion of system lifetime during which the original component operated.

To determine the probability distribution for a function of n random variables, $X_1, X_2, \ldots, X_n$, we must find the joint probability distribution for the r.v's themselves. We generally assume that observations are obtained through random sampling. We will assume throughout the remainder of this text that populations are large in comparison to the sample size and consequently that the random variables obtained through a random sample are in fact independent of one another. We will present three methods for finding the probability distribution for a function of random variables.

Consider r.v's $X_1, X_2, \ldots, X_n$ and a function $U(X_1, X_2, \ldots, X_n)$, denoted simply as $U$. We will present two method to find the probability function of the function of r.v's.

## 5.2 The Joint pdf of a Function of Random Variables

Given set of r.v's $X_1, X_2, \dots, X_n$, consider forming new r.v's $Y_i = ui(X_1, X_2, \dots, X_n)$, $i = 1, \dots, k$, $k \leq n$.

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \text{ the joint pdf of the original r.v's,} \\ g(y_1, y_2, \dots, y_k) &= \text{ the joint pdf of the new r.v's.} \end{aligned}$$

We now focus on finding the joint distribution of these new variables. For our next study, we will use the term random sample which is defined as:

***Definition*: (Random Sample):** Let $X_1, X_2, \dots, X_n$ be independent r.v's each of which has the same pdf $f(x)$, that is $f_i(x_i) = f(x_i)$, $i = 1, 2, \dots, n$, then the joint pdf of these r.v's $f(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i)$. Hence, the r.v's $X_1, X_2, \dots, X_n$ are said to constitute a random sample (r.s) of size $n$ from a distribution $F(x)$.

For instant, if the two r.v's $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent with joint pdf

$$f(x, y) = \prod_{i=1}^{2} f_i(x_i) = \left[ \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2} \right] \left[ \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2} \right].$$

Therefore, $X_1$ and $X_2$ are said to be a r.s of size two.

***Definition*:** A function of one or more r.v's which is not depend on any unknown parameter is called a ***Statistic***. For example, $Y = \sum X_i$ is a statistic, while $Y = X - \mu/\sigma$ is not a statistic unless $\mu$ and $\sigma$ is known. A statistic is any quantity whose value can be calculated from sample data. Therefore, a statistic is a random variable and will be denoted by an upper case letter; a lower case letter is used to represent the calculated or observed value of the statistic.

***Definitions*:** Let $X_1, X_2, \dots, X_n$ be a r.s of size $n$ from any distribution, we define

1. The statistic $\bar{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ is called the sample mean.
2. The statistic $S^2 = \frac{1}{n-1} \left[ \sum\limits_{i=1}^{n} \left( X_i^2 - n\bar{X} \right) \right]$ is called the sample variance.

### 5.2.1 The Method of Transformations

We will illustrate the method of transformations in two cases. The first is when the r.v's $X_1, X_2, \dots, X_n$ are discrete, and the second case for continuous set of r.v's $X_1, X_2, \dots, X_n$.

***5.2.1.1 Discrete r.v's*** Let $X_1, X_2, \dots, X_n$ be $n$ r.v's of discrete type defined on the n-dimensional space $\mathcal{A}$ with joint pdf $f(x_1, x_2, \dots, x_n)$. Let the r.v's $Y_i = u_i(X_1, X_2, \dots, X_n)$, $i = 1, 2, \dots, k$ and $k \leq n$ be functions of $X_1, X_2, \dots, X_n$. The joint pdf of $Y_1, Y_2, \dots, Y_k$ is required. If $k < n$, we will introduce additional new r.v's $Y_{k+1} = u_{k+1}(X_1, X_2, \dots, X_n)$, $Y_{k+2} = u_{k+2}(X_1, X_2, \dots, X_n)$, $dots Y_n = u_n(X_1, X_2, \dots, X_n)$, so that the function $y_i = u_i(X_1, X_2, \dots, X_n)$, $i = 1, 2, \dots, n$ are defined to be one-to-one transformation that maps the space $\mathcal{A}$ of $\{X_i\}$ onto the space $\mathcal{B}$ of $\{Y_i\}$ with inverse transforms $x_i = \omega_i(y_1, y_2, \dots, y_n)$, $i = 1, 2, \dots, n$. Then the joint pdf of $Y_1, Y_2, \dots, Y_n$ is defined as:

$$g(y_1, y_2, \dots, y_n) = f(\omega_1(y_1, y_2, \dots, y_n), \omega_2(y_1, y_2, \dots, y_n), \dots, \omega_n(y_1, y_2, \dots, y_n)).$$

Therefore, the joint pdf of $Y_1, Y_2, \dots, Y_k$ is

$$g^*(y_1, y_2, \dots, y_k) = \sum_{y_{k+1}} \sum_{y_{k+2}} \cdots \sum_{y_n} g(y_1, y_2, \dots, y_n).$$

We can also evaluate the marginal pdf of each r.v $Y_i$ using the definition of the marginal density function.

**EXAMPLE 5.1**

Let the r.v $X$ has a pdf $f(x) = \frac{x}{15}$, $x = 1, 2, 3, 4, 5$. Find the pdf of the r.v $Y = 2X + 1$.

**Solution.** The function $y = 2x + 1$ defines a one-to-one transformation that maps the space $\mathcal{A} = \{x : x = 1, 2, 3, 4, 5\}$ onto the space $\mathcal{B} = \{y : y = 3, 5, 7, 9, 11\}$ with inverse $x = \frac{1}{2}(y - 1)$. Then the pdf of $Y$ is

$$g(y) = f\left(\frac{y-1}{2}\right) = \frac{1}{30}(y-1), \quad y = 3, 5, 7, 9, 11.$$

◀

**EXAMPLE 5.2**

Let the r.v $X \sim b(n, p)$. What is the distribution of r.v $Y = n - X$?

**Solution.** The pdf of $X$ is given by: $f(x) = \binom{n}{x} p^x q^{n-x}$, $x = 0, 1, \ldots, n$. The function $y = n - x$ defines a one-to-one transform that maps the space $\mathcal{A} = \{x : x = 0, 1, \ldots, n\}$ onto the space $\mathcal{B} = \{y : y = 0, 1, \ldots, n\}$ with inverse function of $x = n - y$. Then the pdf of $Y$ is

$$g(y) = f(n-y) = \binom{n}{n-y} p^{n-y} q^{n-(n-y)} = \binom{n}{y} q^y p^{n-y}, \quad y = 0, 1, \ldots, n.$$

Hence, the r.v $Y \sim b(n, q)$. ◀

**EXAMPLE 5.3**

Suppose that $X_1$ and $X_2$ are two independent r.v's where $X_1 \sim P(\lambda_1)$ and $X_2 \sim P(\lambda_2)$ (i.e $X_1$ and $X_2$ are a r.s of size 2). Find the distribution of the r.v $Y = X_1 + X_2$.

**Solution.** Since $X_1$ and $X_2$ are a r.s, then the joint pdf of $X_1$ and $X_2$ can be determined as

$$f(x_1, x_2) = f_1(x_1) \times f_2(x_2) = \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!} = \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{x_1}\lambda_2^{x_2}}{x_1! x_2!}, \quad x_i = 0, 1, 2, \ldots;\ i = 1, 2.$$

◀

Consider, beside $Y = X_1 + X_2$, $Z = X_2$. The functions $y = x_1 + x_2$ and $z = x_2$ define one-to-one transformation that maps the space $\mathcal{A} = \{(x_1, x_2) : x_i = 0, 1, 2, \ldots;\ i = 1, 2\}$ onto the space $\mathcal{B} = \{(y, z) : 0 \le z \le y \le \infty\}$, with inverse functions $x_1 = y - z$ and $x_2 = z$. Then the joint pdf of $Y$ and $Z$ is

$$g(y, z) = f(y - z, z) = \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{y-z}\lambda_2^{z}}{(y-z)! z!}, \quad (y, z) \in \mathcal{B}.$$

Now, the marginal pdf of $Y$ is obtained by

$$\begin{aligned} g_Y(y) &= \sum_z g(y, z) = e^{-(\lambda_1+\lambda_2)} \sum_{z=0}^{y} \frac{\lambda_1^{y-z}\lambda_2^{z}}{(y-z)! z!} = \frac{e^{-(\lambda_1+\lambda_2)}}{y!} \sum_{z=0}^{y} \frac{y!}{(y-z)! z!} \lambda_1^{y-z}\lambda_2^{z} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{y!} \sum_{z=0}^{y} \binom{y}{z} \lambda_2^{z}\lambda_1^{y-z} = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^y}{y!}, \quad y = 0, 1, 2, \ldots. \end{aligned}$$

Therefore $Y \sim P(\lambda_1 + \lambda_2)$.

***Note*:** We can prove that if $X_1, X_2, \ldots, X_n$ are r.s of size $n$ and $X_i \sim P(\lambda_i)$, then $Y = \sum_{i=1}^{n} X_i \sim P(\sum_{i=1}^{n} \lambda_i)$.

**EXAMPLE 5.4**

Let $X_1$ and $X_2$ be independent r.v's of discrete type with joint pdf $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$, where $f_i(x_i)$ is the marginal pdf of $X_i$, $i = 1, 2$. Consider $Y_i = u_i(X_i)$, $i = 1, 2$ define one-to-one transformation from $\mathcal{A}$ onto $\mathcal{B}$. Show that the r.v's $Y_1 = u_1(X_1)$ and $Y_2 = u_2(X_2)$ are also independent.

**Solution.** Given $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ where $f(x_1, x_2)$ is the joint pdf of $X_1$ and $X_2$ with $f_i(x_i)$ is the marginal pdf of $X_i$, $i = 1, 2$.

Also the functions $y_i = u_i(x_i)$, $i = 1, 2$ maps $\mathcal{A}$ onto $\mathcal{B}$, then the inverse functions $x_i = \omega_i(y_i)$, $i = 1, 2$. Then the joint pdf of $Y_1$ and $Y_2$ is:

$$g(y_1, y_2) = f(\omega_1(y_1), \omega_2(y_2)) = f_1(\omega_1(y_1)) \cdot f_2(\omega_2(y_2)).$$

◀

Then, the marginal pdf of $Y_1$ is obtained by

$$g_1(y_1) = \sum_{y_2} g(y_1, y_2) = f_1(\omega_1(y_1)) \sum_{y_2} f_2(\omega_2(y_2)) = c_1 f_1(\omega_1(y_1)),$$

similarly, the marginal pdf of $Y_2$ is

$$g_2(y_2) = \sum_{y_1} g(y_1, y_2) = f_2(\omega_2(y_2)) \sum_{y_1} f_1(\omega_1(y_1)) = c_2 f_2(\omega_2(y_2)).$$

In order to verify the independence of $Y_1$ and $Y_2$, we need to prove that $g(y_1, y_2) = g_1(y_1) \cdot g_2(y_2)$. Therefore

$$\begin{aligned} g_1(y_1) \cdot g_2(y_2) &= c_1 f_1(\omega_1(y_1)) \cdot c_2 f_2(\omega_2(y_2)) = c_1 c_2 f_1(\omega_1(y_1)) \cdot f_2(\omega_2(y_2)) \\ &= c_1 c_2 g(y_1, y_2), \\ \Rightarrow g(y_1, y_2) &= \frac{1}{c_1 c_2} g_1(y_1) \cdot g_2(y_2). \end{aligned}$$

Since $g(y_1, y_2)$ is the joint pdf of $Y_1$ and $Y_2$, then

$$\begin{aligned} 1 &= \sum_{y_1} \sum_{y_2} g(y_1, y_2) = \sum_{y_1} \sum_{y_2} \frac{1}{c_1 c_2} g_1(y_1) \cdot g_2(y_2) = \frac{1}{c_1 c_2} \left[ \sum_{y_1} g_1(y_1) \right] \left[ \sum_{y_2} g_2(y_2) \right] \\ &= \frac{1}{c_1 c_2} \Rightarrow c_1 c_2 = 1 \\ &\therefore g(y_1, y_2) = g_1(y_1) \cdot g_2(y_2). \end{aligned}$$

This implies that $Y_1$ and $Y_2$ are independent r.v's.

***5.2.1.2 Continuous r.v's*** In the case of continuous r.v's we assume the same environment as it was with the discrete case. Let $X_1, X_2, \ldots, X_n$ be $n$ r.v's of continuous type defined on the n-dimensional space $\mathcal{A}$ with joint pdf $f(x_1, x_2, \ldots, x_n)$. Let the r.v's $Y_i = u_i(X_1, X_2, \ldots, X_n)$, $i = 1, 2, \ldots, k$ $(k \leq n)$ be functions of $X_1, X_2, \ldots, X_n$. The joint pdf of $Y_1, Y_2, \ldots, Y_k$ is required. If $k < n$, we will introduce additional new r.v's $Y_{k+1} = u_{k+1}(X_1, X_2, \ldots, X_n)$, $Y_{k+2} = u_{k+2}(X_1, X_2, \ldots, X_n)$, $dotsY_n = u_n(X_1, X_2, \ldots, X_n)$, so that the function $y_i = u_i(X_1, X_2, \ldots, X_n)$, $i = 1, 2, \ldots, n$ are defined to be one-to-one transformation that maps the space $\mathcal{A}$ of $\{X_i\}$ onto the space $\mathcal{B}$ of $\{Y_i\}$ with inverse transforms $x_i = \omega_i(y_1, y_2, \ldots, y_n)$, $i = 1, 2, \ldots, n$. In this case, we will need to calculate the Jacobian $n \times n$ determinant of the first partial derivatives as:

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

Then the joint pdf of $Y_1, Y_2, \ldots, Y_n$ is defined as:

$$g(y_1, y_2, \ldots, y_n) = |J| \times f(\omega_1(y_1, y_2, \ldots, y_n), \omega_2(y_1, y_2, \ldots, y_n), \ldots, \omega_n(y_1, y_2, \ldots, y_n)),$$

and finally, the joint pdf of $Y_1, Y_2, \ldots, Y_k$ is

$$g^*(y_1, y_2, \ldots, y_k) = \int_{y_{k+1}} \int_{y_{k+2}} \cdots \int_{y_n} g(y_1, y_2, \ldots, y_n) dy_{k+1} dy_{k+2} \ldots dy_n.$$

We can also evaluate the marginal pdf of each r.v $Y_i$ using the definition of the marginal density function.

**EXAMPLE 5.5**

Let the r.v $X$ has a pdf $f(x) = 3x^2$, $0 < x < 1$. Find the pdf of the r.v $Y = X^3$ and evaluate $\Pr(\frac{1}{2} < Y < \frac{3}{4})$.

**Solution.** The function $y = x^3$ is a one-to-one function from $\mathcal{A} = \{x : 0 < x < 1\}$ onto $\mathcal{B} = \{y : 0 < y < 1\}$ with inverse function $x = y^{1/3}$. The Jacobian in the 1-dimensional case is obtained as

$$J = \frac{\partial x}{\partial y} = \frac{1}{3} y^{-2/3}.$$

Therefore, the pdf of the r.v $Y$ is

$$g(y) = f(y^{1/3})|J| = 3(y^{1/3})^2 \frac{1}{3} y^{-2/3} = 1, \quad 0 < y < 1.$$

Now, to evaluate the probability $\Pr(\frac{1}{2} < Y < \frac{3}{4})$,

$$\Pr\left(\frac{1}{2} < Y < \frac{3}{4}\right) = \int_{1/2}^{3/4} dy = y\Big|_{1/2}^{3/4} = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}.$$

◄

**EXAMPLE 5.6**

Let the r.v $X \sim U(0, 1)$. Find and name the distribution of the r.v $Y = -2 \ln X$.

**Solution.** The function $y = -2 \ln x$ is a one-to-one function that maps $\mathcal{A} = \{x : 0 < x < 1\}$ onto $\mathcal{B} = \{y : 0 < y < \infty\}$, with inverse function $x = e^{-\frac{1}{2}y}$. The Jacobian,

$$J = \frac{\partial x}{\partial y} = -\frac{1}{2} e^{-\frac{1}{2}y}.$$

Then, the pdf of $Y$ can be obtained as

$$g(y) = |J| f(e^{-\frac{1}{2}y}) = (1)\frac{1}{2} e^{-\frac{1}{2}y} = \frac{1}{2} e^{-\frac{1}{2}y}, \quad 0 < y < \infty$$

Hence, the r.v $Y \sim Exp(2)$. ◄

### EXAMPLE 5.7

Let $X_1$ and $X_2$ be a r.s of size 2 from $N(0,1)$. Define the r.v's $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find the joint pdf of $Y_1$ and $Y_2$ and show that they are independent.

**Solution.** The joint pdf of $X_1$ and $X_2$ is:

$$f(x_1, x_2) = f_1(x_1) \times f_2(x_2) = \frac{1}{2\pi} e^{-\frac{1}{2\pi}(X_1^2+X_2^2)}, \quad -\infty < x_i < \infty;\ i = 1, 2.$$

The functions $y_1 = x_1 + x_2$, $y_2 = x_1 - x_2$ define as one-to-one functions that map the space $\mathcal{A} = \{(x_1, x_2) : -\infty < x_i < \infty;\ i = 1, 2\}$ onto the space $\mathcal{B} = \{(y_1, y_2) : -\infty < y_i < \infty;\ i = 1, 2\}$, with inverses $x_1 = \frac{1}{2}(y_1 + y_2)$ and $x_2 = \frac{1}{2}(y_1 - y_2)$. The Jacobian can be calculated as

$$J = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}.$$

Therefore, the joint pdf of $y_1$ and $y_2$ is

$$g(y_1, y_2) = f\left(\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)\right) \times |J|.$$

taking the term $x_1^2 + x_2^2$:

$$x_1^2 + x_2^2 = \frac{1}{4}(y_1 + y_2)^2 + \frac{1}{4}(y_1 - y_2)^2 = \frac{1}{4}\left[y_1^2 + 2y_1y_2 + y_2^2 + y_1^2 - 2y_1y_2 + y_2^2\right] = \frac{1}{4}(2y_1^2 + 2y_2^2) = \frac{1}{2}(y_1^2 + y_2^2).$$

Therefore,

$$g(y_1, y_2) = \frac{1}{4\pi} e^{-\frac{1}{4}(y_1^2+y_2^2)}, \quad -\infty < y_i < \infty;\ i = 1, 2.$$

To verify if $Y_1$ and $Y_2$ are independent or not, we need to find the marginal pdf's for both $Y_1$ and $Y_2$. Then, the marginal pdf of $Y_1$ is

$$\begin{aligned} g_1(y_1) &= \int_{y_2} g(y_1, y_2) dy_2 = \frac{1}{4\pi} e^{-\frac{1}{4}y_1^2} \int_{-\infty}^{\infty} e^{-\frac{1}{4}y_2^2} dy_2 = \frac{\sqrt{2\pi}\sqrt{2}}{4\pi} e^{-\frac{1}{4}y_1^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{4}y_2^2} dy_2 \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_1^2}{2}} \quad -\infty < y_1 < \infty. \end{aligned}$$

That is $Y_1 \sim N(0, 2)$. Similarly, the marginal pdf of $Y_2$ is $g_2(y_2) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_2^2}{2}}$, $-\infty < y_1 < \infty$, which means that $Y_2 \sim N(0, 2)$. Then,

$$g_1(y_1) \cdot g_2(y_2) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_1^2}{2}} \times \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_2^2}{2}} = \frac{1}{4\pi} e^{-\frac{1}{4}(y_1^2+y_2^2)} = g(y_1, y_2).$$

Therefore, $Y_1$ and $Y_2$ are independent r.v's. ◀

### EXAMPLE 5.8

Let $X_1$ and $X_2$ be a r.s of size 2 from $U(0,1)$. Define the r.v's $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find the joint pdf of $Y_1$ and $Y_2$. Also, find marginal pdf's for both $Y_1$ and $Y_2$.

**Solution.** The joint pdf of $X_1$ and $X_2$ is $f(x_1, x_2) = 0$, $0 < x_i < 1$, $i = 1, 2$.

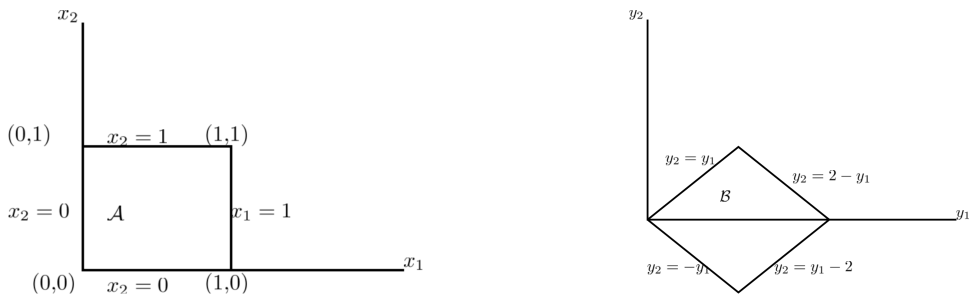


**Figure 5.1** The Sample Space $\mathcal{A}$ and $\mathcal{B}$

The functions $y_1 = x_1 + x_2$, $y_2 = x_1 - x_2$ define a one-tp-one transformation that maps $\mathcal{A} = \{(x_1, x_2) : 0 < x_1 < 1, 0 < x_2 < 1\}$ on to the space $\mathcal{B}$ in the $y_1 y_2$-plane with inverses:

$$x_1 = \frac{1}{2}(y_1 + y_2) \quad \text{and} \quad x_2 = \frac{1}{2}(y_1 - y_2).$$

We determine the boundaries of $\mathcal{B}$ as follows:

$$\begin{aligned}
x_1 = 0 &\Rightarrow \frac{1}{2}(y_1 + y_2) = 0 \Rightarrow y_2 = -y_1 \\
x_1 = 1 &\Rightarrow \frac{1}{2}(y_1 + y_2) = 1 \Rightarrow y_2 = 2 - y_1 \\
x_2 = 0 &\Rightarrow \frac{1}{2}(y_1 - y_2) = 0 \Rightarrow y_2 = y_1 \\
x_2 = 1 &\Rightarrow \frac{1}{2}(y_1 - y_2) = 1 \Rightarrow y_2 = y_1 - 2
\end{aligned}$$

and

$$J = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

Then, the joint pdf of $Y_1$ and $Y_2$ is

$$g(y_1, y_2) = f\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right)|J| = \frac{1}{2}, \ \ (y_1, y_2) \in \mathcal{B}.$$

The marginal pdf of $Y_1$ could be found as

$$g_1(y_1) = \int_{y_2} g(y_1, y_2) dy_2 = \begin{cases} \int\limits_{y_2=-y_1}^{y_1} \frac{1}{2} dy_2 = y_1, & 0 \leq y_1 < 1 \\ \\ \int\limits_{y_2=y_1-2}^{2-y_1} \frac{1}{2} dy_2 = 2 - y_1, & 1 \leq y_1 < 2 \end{cases}$$

and the marginal pdf of $Y_2$ is

$$g_2(y_2) = \int_{y_1} g(y_1, y_2) dy_1 = \begin{cases} \int\limits_{y_1=-y_2}^{y_2+2} \frac{1}{2} dy_1 = y_2 + 1, & -1 \leq y_2 < 0 \\ \\ \int\limits_{y_1=y_2}^{2-y_2} \frac{1}{2} dy_1 = 1 - y_2, & 0 \leq y_2 < 1 \end{cases}$$

◀

**EXAMPLE 5.9**

Consider $n = 3$ identical components with independent lifetimes $X_1$, $X_2$, $X_3$, each having an exponential distribution with parameter $1/\lambda$. If the first component is used until it fails, replaced by the second one which remains in service until it fails, and finally the third component is used until failure, then the total lifetime of these components is $Y = X_1 + X_2 + X_3$. Find the distribution of total lifetime.

**Solution.** The joint pdf of $X_1$, $X_2$ and $X_3$ is $f(x_1, x_2, x_3) = \lambda^3 e^{-\lambda(x_1+x_2+x_3)}$, $0 < x_i < \infty$, $i = 1, 2, 3$. We define two other new variables $Y_1 = X_1$ and $Y_2 = X_1 + X_2$ (so that $Y_1 < Y_2 < Y$). Then the functions $y = x_1 + x_2 + x_3$, $y_2 = x_1 + x_2$ and $y_1 = x_1$ define a one-tp-one transformation that maps $\mathcal{A} = \{(x_1, x_2, x_3) : 0 < x_i < \infty, i = 1, 2, 3\}$ onto the space $\mathcal{B} = \{(y_1, y_2, y_3) : 0 < y_1 < y_2 < y < \infty\}$, with inverses:

$$x_1 = y_1, \quad x_2 = y_2 - y_1 \quad \text{and} \quad x_3 = y - y_2.$$

The Jacobian is calculated as

$$J = \frac{\partial(x_1, x_2, x_3)}{\partial(y_1, y_2, y)} = \begin{vmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{vmatrix} = 1$$

Then, the joint pdf of $Y_1$, $Y_2$ and $Y$ is

$$g(y_1, y_2, y) = f(y_1, y_2 - y_1, y - y_2)|J| = \lambda^3 e^{-\lambda y}, \quad 0 < y_1 < y_2 < y < \infty.$$

Finally, the marginal pdf of $Y$ is

$$\begin{aligned} g^*(y) &= \int_{y_1}\int_{y_2} g(y_1, y_2, y)dy_1 dy_2 = \int_0^y \int_{y_1}^y \lambda^3 e^{-\lambda y} dy_1 dy_2 = \int_0^y \lambda^3 e^{-\lambda y} \Big[y_2\Big]_{y_1}^y dy_1 \\ &= \int_0^y \lambda^3 e^{-\lambda y}(y - y_1)dy_1 = \lambda^3 e^{-\lambda y}\left[yy_1 - \frac{y_1^2}{2}\right]_0^y = \lambda^3 e^{-\lambda y}\left[y^2 - \frac{y^2}{2}\right] \\ &= \frac{\lambda^3}{2} y^2 e^{-\lambda y}, \quad 0 < y < \infty. \end{aligned}$$

◀

## 5.3 Order Statistics

Many functions of random variables of interest in practice depend on the relative magnitudes of the observed variables. For instance, we may be interested in the fastest time in an automobile race or the heaviest mouse among those fed on a certain diet. Thus, we often order observed random variables according to their magnitudes. The resulting ordered variables are called order statistics.

Formally, let $X_1, X_2, \ldots, X_n$ denote independent continuous random variables with distribution function $F(x)$ and density function $f(x)$. We denote the ordered random variables $X_i$ by $Y_1, Y_2, \ldots, Y_n$, where $Y_1 \leq Y_2 \leq \cdots \leq Y_n$. where

$$\begin{aligned} Y_1 &= \text{the smallest among } X_1, X_2, \ldots, X_n \\ Y_2 &= \text{the second smallest among } X_1, X_2, \ldots, X_n \\ &\vdots \\ Yn &= \text{the largest among } X_1, X_2, \ldots, X_n \end{aligned}$$

i.e, $Y_1 = \min\{X_1, X_2, \dots, X_n\}$ and $Y_n = \max\{X_1, X_2, \dots, X_n\}$.

***Notes*:**

1. In most applications, the unordered set of r.v's $\{X_i\}$ is usually taken to be a r.s of size $n$.
2. We only consider the continuous type of r.v's.

### 5.3.1 The distribution of Order Statistics

Let $X_1, X_2, \dots, X_n$ be a r.s of size $n$ from a continuous distribution having a pdf of $f(x)$ and cdf $F(x)$, where the joint pdf of those r.v's is

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i), \quad -\infty < x_i < \infty;\ i = 1, 2, \dots, n.$$

Suppose that $Y_1, Y_2, \dots, Y_n$ is a ordered sample of the original r.s, that is $Y_1 = \min\{X_1, X_2, \dots, X_n\}$ and $Y_n = \max\{X_1, X_2, \dots, X_n\}$. The probability density functions for $Y_1$ and $Y_n$ can be found using the method of distribution functions. We will derive the density function of $Y_n$ first. Because $Y_n$ is the maximum of $X_1, X_2, \dots, X_n$, the event $(Y_n \le y)$ will occur if and only if the events $(X_i \le y)$ occur for every $i = 1, 2, \dots, n$. That is,

$$\Pr(Y_n \le y) = \Pr(X_1 \le y, X_2 \le y, \dots, X_n \le y).$$

Because the $X_i$ are independent and $\Pr(X_i \le y) = F(y)$ for $i = 1, 2, \dots, n$, it follows that the distribution function of $Y_n$ is given by

$$F_{Y_n}(y) = \Pr(Y_n \le y) = \Pr(X_1 \le y)\Pr(X_2 \le y)\dots\Pr(X_n \le y) = [F(y)]^n.$$

Letting $g_n(y)$ denote the density function of $Y_n$, we see that, on taking derivatives of both sides,

$$g_n(y) = n[F(y)]^{n-1} f(y).$$

The density function for $Y_1$ can be found in a similar manner. The distribution function of $Y_1$ is

$$F_{Y_1}(y) = \Pr(Y_1 \le y) = 1 - \Pr(Y_1 > y).$$

Because $Y_1$ is the minimum of $X_1, X_2, \dots, X_n$, it follows that the event $(Y_1 > y)$ occurs if and only if the events $(X_i > y)$ occur for $i = 1, 2, \dots, n$. Because the $X_i$ are independent and $\Pr(Y_i > y) = 1F(y)$ for $i = 1, 2, \dots, n$, we see that

$$\begin{aligned}
F_{Y_1}(y) &= \Pr(Y_1 \le y) = 1 - \Pr(Y_1 > y) \\
&= 1 - \Pr(X_1 > y, X_2 > y, \dots, X_n > y) \\
&= 1 - [\Pr(X_1 > y)\Pr(X_2 > y)\dots\Pr(X_n > y)] \qquad = 1 - [1 - F(y)]^n.
\end{aligned}$$

Thus, if $g_1(y)$ denotes the density function of $Y_1$, differentiation of both sides of the last expression yields

$$g_1(y) = n[1 - F(y)]^{n-1} f(y).$$

In general, The marginal pdf of the $k^{\text{th}}$ order statistic $Y_k$ could be derived as

$$g_k(y) = \frac{n!}{(k-1)!(n-k)!}[F(y)]^{k-1}[1 - F(y)]^{n-k} f(y), \quad -\infty < y < \infty.$$

**EXAMPLE 5.10**

Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a r.s of size 4 from a distribution having a pdf of $f(x) = 2x$, $0 < x < 1$. Compute $\Pr(Y_3 > 1/2)$ and $E[Y_3]$.

**Solution.** Since the pdf of the r.s is $f(x) = 2x$, $0 < x < 1$. Then we can find the cdf as

$$F(x) = \int_{-\infty}^{x} f(\tau)d\tau = \int_{0}^{x} 2\tau d\tau = \tau\Big|_0^x = x^2.$$

Now, since $n = 4$ and $k = 3$, then the pdf of $Y_3$ is

$$g_3(y) = \frac{4!}{2!1!}y^4(1-y^2)2y = 24y^5(1-y^2), \quad 0 < y < 1.$$

Hence, the probability

$$\Pr\left(Y_3 > \frac{1}{2}\right) = \int_{1/2}^{1} g_3(y)dy = \int_{1/2}^{1} 24y^5(1-y^2)dy = 24\left[\frac{y^6}{6} - \frac{y^8}{8}\right]_{1/2}^{1} = 0.949.$$

The mean value of $Y_3$ is

$$E[Y_3] = \int_0^1 yg_3(y)dy = 24\int_0^1 y^6(1-y^2)dy = 24\left[\frac{y^7}{7} - \frac{y^9}{9}\right]_0^1 = \frac{16}{21}.$$

◀

**EXAMPLE 5.11**

Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n$ from $Exp(\lambda)$. Find the distribution of the first order statistics $Y_1$.

**Solution.** Since $X \sim Exp(\lambda)$, then $f(x) = \frac{1}{\lambda}e^{-x/\lambda}$, $0 < x < \infty$, and the cdf $F(x) = 1 - e^{-x/\lambda}$. Therefore, the pdf of $Y_1$ is given by

$$g_1(y) = n[1-F(y)]^{n-1}f(y) = n\left[1 - 1 + e^{-\frac{y}{\lambda}}\right]^{n-1}\frac{1}{\lambda}e^{-\frac{y}{\lambda}} = \frac{n}{\lambda}\left[e^{-\frac{y}{\lambda}}\right]^{n-1}e^{-\frac{y}{\lambda}} = \frac{n}{\lambda}e^{-\frac{n}{\lambda}y}, \quad 0 < y < \infty.$$

◀

**EXAMPLE 5.12**

Electronic components of a certain type have a length of life $X$, with probability density given by: $f(x) = (1/100)e^{-x/100}$, $x > 0$. (Length of life is measured in hours.) Suppose that two such components operate independently and in series in a certain system (hence, the system fails when either component fails). Find the density function for $Y$, the length of life of the system.

**Solution.** Because the system fails at the first component failure, $Y = \min\{X_1, X_2\}$, where $X_1$ and $X_2$ are independent r.v's with the given density. Then, because $F(x) = 1 - e^{-x/100}$, for $x \geq 0$,

$$g_1(y) = n[1-F(y)]^{n-1}f(y) = 2e^{-\frac{y}{100}}\frac{e^{-\frac{y}{100}}}{100} = \frac{1}{50}e^{-\frac{y}{50}}, \quad y > 0.$$

Thus, the minimum of two exponentially distributed random variables has an exponential distribution. Notice that the mean length of life for each component is 100 hours, whereas the mean length of life for the system is $E[Y] = 50 = 100/2$. ◀

**EXAMPLE 5.13**

Let $X$ denote the contents of a one-gallon container, and suppose that its pdf is $f(x) = 2x$ for $0 < x < 1$ (and 0 otherwise) with corresponding cdf $F(x) = x^2$ in the interval of positive density. Consider a random sample of four such containers. Determine the expected value of $Y_4 - Y_1$, the difference between the contents of the most-filled container and the least-filled container.

**Solution.** To determine $Y_4 - Y_1$, the sample range, The pdf's of $Y_4$ and $Y_1$ are

$$g_4(y) = n[F(y)]^{n-1} f(y) = 4(y^2)^3 \cdot 2y = 8y^7, \quad 0 \le y \le 1,$$

and

$$g_1(y) = n[1 - F(y)]^{n1} f(y) = 4(1 - y^2)^3 \cdot 2y = 8y(1 - y^2)^3, \quad 0 \le y \le 1.$$

Therefore, the expected value of $Y_4 - Y_1$ is

$$E[Y_4 - Y_1] = E[Y_4] - E[Y_1] = \int_0^1 y \cdot 8y^7 dy - \int_0^1 y \cdot 8y(1 - y^2)^3 dy = 0.889 - 0.406 = 0.483.$$

If random samples of four containers were repeatedly selected and the sample range of contents determined for each one, the long run average value of the range would be 0.483. ◄

### 5.3.2 Joint Distribution of the Order Statistics

Consider a r.s $X_1, X_2, \ldots, X_n$ of size $n$ with a joint pdf

$$f(x_1, x_2, \ldots, x_n) = f(x_1) \cdot f(x_2) \ldots f(x_n),$$

and $f(x_i)$ is the marginal pdf of each r.v $X_i$. Then joint pdf of r.v's $Y_1, Y_2, \ldots, Y_n$ is presented as:

$$g(y_1, y_2, \ldots, y_n) = n! f(y_1, y_2, \ldots, y_n) = n! \prod_{i=1}^{n} f(y_i), \quad -\infty < y_1 < y_2 < \cdots < y_n < \infty.$$

We can also derive the joint pdf of any two order statistics, if $h$ and $k$ are two integers such that $1 \le h < k \le n$, the joint density of $Y_h$ and $Y_k$ is given by

$$g_{h,k}(y_h, y_k) = \frac{n!}{(h-1)!(k-h-1)!(n-k)!} [F(y_h)]^{h-1} [F(y_k) - F(y_h)]^{k-h-1} [1 - F(y_k)]^{n-k} f(y_h) f(y_k),$$

$$-\infty < x_h < x_k < \infty$$

**EXAMPLE 5.14**

If $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a r.s of size $n$ from $N(\mu, \sigma^2)$. Find the joint pdf of $Y_1, Y_2, \ldots, Y_n$.

**Solution.** The joint pdf of the order statistics $Y_1, Y_2, \ldots, Y_n$ is

$$g(y_1, y_2, \ldots, y_n) = n! \prod_{i=1}^{n} f(y_i) = n! (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2}, \quad -\infty < y_1 < y_2 < \cdots < y_n.$$

◄

**EXAMPLE 5.15**

Let $Y_1, Y_2, Y_3$ be the order statistics of a r.s of size 3 from $U(0,1)$.

1. Find the joint pdf of $Y_1$ and $Y_3$.
2. Find the marginal pdf of $Z = Y_3 - Y_1$.

**Solution.** Since $Y_1, Y_2, Y_3$ be the order statistics of a r.s of size 3 from $U(0,1)$, then: $f(x) = 1, 0 < x < 1$. and

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x < 1 \\ 1, & x \geq 1 \end{cases}$$

1. In order to get the joint pdf of $Y_1$ and $Y_3$, we have $n = 3$, $h = 1$ and $k = 3$. Therefore,

$$g_{1,3}(y_1, y_3) = \frac{3!}{0!1!0!}[F(y_1)]^0[F(y_3) - F(y_1)]^1[1 - F(y_3)]^0 f(y_1)f(y_3) = 6(y_3 - y_1), \quad 0 < y_1 < y_3 < 1.$$

2. We have the transformation $Z = Y_3 - Y_1$, set $W = Y_3$. Now, the one-to-one functions $z = y_3 - y_1$, $w = y_3$ map the space $\mathcal{A} = \{(y_1, y_3) : 0 < y_1 < y_3 < 1\}$ onto $\mathcal{B} = \{(z, w) : 0 < z < w < 1\}$, with inverses $y_1 = w - z$ and $y_3 = w$, and

$$J = \frac{\partial(y_1, y_3)}{\partial(z, w)} = \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} = -1.$$

Then, the joint pdf of $Z$ and $W$ is

$$h(z, w) = g_{1,3}(w - z, w)|J| = 6(w - (w - z)) = 6z, \quad 0 < z < w < 1.$$

The marginal pdf of $Z$ is

$$h_1(z) = \int_w h(z, w)dw = \int_z^1 6z dw = 6z\Big[w\Big]_z^1 = 6z(1 - z), \quad 0 < z < 1.$$

◀

# CHAPTER 6

# SAMPLING DISTRIBUTIONS

## 6.1 Introduction

This chapter helps make the transition between probability and inferential statistics. Given a sample of $n$ observations from a population, we will be calculating estimates of the population mean, median, standard deviation, and various other population characteristics (parameters). Prior to obtaining data, there is uncertainty as to which of all possible samples will occur. Because of that, estimates such as $\bar{x}$, $\tilde{x}$ and $s$ will vary from one sample to another. The behaviour of such estimates in repeated sampling is described by what are called sampling distributions. Any particular sampling distribution will give an indication of how close the estimate is likely to be to the value of the parameter being estimated. We will start with some important definition and theorems that are needed in what follows.

Recall the definitions from Chapter'**??**, which are

***Definition*:** A function of one or more r.v's which is not depend on any unknown parameter is called a ***Statistic***. For example, $Y = \sum X_i$ is a statistic, while $Y = X - \mu/\sigma$ is not a statistic unless $\mu$ and $\sigma$ is known. A statistic is any quantity whose value can be calculated from sample data. Therefore, a statistic is a random variable and will be denoted by an upper case letter; a lower case letter is used to represent the calculated or observed value of the statistic.

***Definitions*:** Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n$ from any distribution, we define

1. The statistic $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ is called the sample mean.

2. The statistic $S^2 = \frac{1}{n-1}\left[\sum\limits_{i=1}^{n}\left(X_i^2 - n\bar{X}\right)\right]$ is called the sample variance.



Please enter `\offprintinfo{(Title, Edition)}{(Author)}` at the beginning of your document.

**EXAMPLE 6.1**

The time that it takes to serve a customer at the cash register in a mini-market is a r.v having an exponential distribution with parameter $1/\lambda$. Suppose $X_1$ and $X_2$ are service times for two different customers, assumed independent of each other. Consider the total service time $T_o = X_1 + X_2$ for the two customers, also a statistic. The cdf of $T_o$ is, for $t \geq 0$,

$$F_{T_o}(t) = \Pr(T_o \leq t) = \Pr(X_1 + X_2 \leq t) = \iint\limits_{\{(x_1,x_2):x_1+x_2\leq t\}} f(x_1, x_2)dx_1dx_2$$

$$= \int_0^t \int_0^{t-x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} dx_2dx_1 = \int_0^t (\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t})dx_1 = 1 - e^{-\lambda t} - \lambda t e^{-\lambda x_1}$$

The pdf of $T_o$ is obtained by differentiating $F_{T_o}(t)$:

$$f_{T_o}(t) = \lambda^2 t e^{-\lambda t}, \quad t \geq 0. \tag{6.1}$$

This is a gamma pdf ($\alpha = 2$ and $\beta = 1/\lambda$).

The pdf of $\bar{X} = T_o/2$ can be obtained by the cdf method of Section as

$$f_{\bar{X}}(\bar{x}) = 4\lambda^2 \bar{x} e^{-2\lambda\bar{x}}, \quad \bar{x} \geq 0. \tag{6.2}$$

The mean and variance of the underlying exponential distribution are $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$. Using expressions‘(**??**) and (**??**), it can be verified that $E[\bar{X}] = 1/\lambda$, $Var(\bar{X}) = 1/(2\lambda^2)$, $E[T_o] = 2/\lambda$ and $Var(T_o) = 2/\lambda^2$. These results again suggest some general relationships between means and variances of $\bar{X}$, $T_o$, and the underlying distribution.

## 6.2 Sampling Distributions Related to the Normal Distribution

We have already noted that many phenomena observed in the real world have relative frequency distributions that can be modelled adequately by a normal probability distribution. Thus, in many applied problems, it is reasonable to assume that the observable random variables in a random sample, $X_1, X_2, \ldots, X_n$, are independent with the same normal density function. We present it formally in the following theorem.

***Theorem 1*:**

Let $X_1, X_2, \ldots, X_n$ be independent r.v's with mgf of $X_i$, $i = 1, 2, \ldots, n$ is $M_{X_i}(t_i)$. Then the r.v $Y = \sum_{i=1}^n k_i X_i$ has an mgf $M_Y(t) = \prod_{i=1}^n M_{X_i}(t_i)$.

***Proof*:**

In order to evaluate the mgf of $Y = \sum_{i=1}^n k_i X_i$, we consider

$$M_Y(t) = E\left[e^{tY}\right] = E\left[e^{t\sum_{i=1}^n k_i X_i}\right] = E\left[\prod_{i=1}^n e^{tk_iX_i}\right]$$

$$= \prod_{i=1}^n E\left[e^{tk_iX_i}\right] = \prod_{i=1}^n M_{X_i}(k_i t).$$

***Theorem 2*:**

Let $X_1, X_2, \ldots, X_n$ be independent r.v's where $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$. Then the r.v $Y = \sum\limits_{i=1}^{n} k_i X_i$ has a normal distribution, $N\left(\sum\limits_{i=1}^{n} k_i\mu_i, \sum\limits_{i=1}^{n} k_i^2\sigma_i^2\right)$.

***Proof*:**

Since each $X_i \sim N(\mu_i, \sigma_i^2)$, then the mgf of any $X_i$ is $M_{X_i}(t_i) = e^{\mu_i t_i + \frac{1}{2}\sigma_i^2 t_i^2}$, for each $i = 1, 2, \ldots, n$.
If $M_Y(t)$ is the mgf of $Y$, the according to theorem 1,

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(k_i t) = \prod_{i=1}^{n} e^{\mu_i k_i t + \frac{1}{2}\sigma_i^2 k_i^2 t^2} = e^{(\sum_{i=1}^{n} k_i\mu_i)t + \frac{1}{2}(\sum_{i=1}^{n} k_i^2\sigma_i^2)t^2},$$

Which is the mgf of the r.v $Y \sim N(\sum_{i=1}^{n} k_i\mu_i, \sum_{i=1}^{n} k_i^2\sigma_i^2)$

***Note*:** If $X_1, X_2, \ldots, X_n$ from theorem 2 represent a r.s of size $n$ from $N(\mu, \sigma^2)$, then the r.v

$$Y = \sum_{i=1}^{n} k_i X_i \sim N\left(\mu\sum_{i=1}^{n} k_i, \sigma^2\sum_{i=1}^{n} k_i^2\right)$$

.

***Theorem 3*: (The Distribution of $\bar{X}$)**

Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a normal distribution with mean $\mu$ and variance $\sigma^2$. Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is normally distributed with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma_{\bar{X}}^2 = \sigma^2/n$, i.e $\bar{X} \sim N(\mu, \sigma^2/n)$.

***Proof*:**

Since $X_i \sim N(\mu, \sigma^2)$, and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \frac{1}{n}X_i$. Then, according Theorem 2, taking $k_i = 1/2$,

$$\bar{X} \sim N\left(\sum_{i=1}^{n} \frac{1}{n}\mu, \sum_{i=1}^{n} \frac{1}{n^2}\sigma^2\right) \equiv N\left(\mu, \sigma^2/n\right).$$

**EXAMPLE 6.2**

$X_1, X_2, \ldots, X_{25}$ be a r.s of size 25 from $N(75, 100)$. Find $\Pr(71 < \bar{X} < 79)$.

**Solution.** We have $n = 25$, $\mu = 75$ and $\sigma^2 = 100$, and since each $X_i \sim N(75, 100)$, then according Theorem 2

$$\bar{X} \sim N(70, 100/25) \equiv N(75, 4).$$

Therefore,

$$\Pr(71 < \bar{X} < 79) = \Pr\left(\frac{71 - 75}{2} < \frac{\bar{X} - 75}{2} < \frac{79 - 75}{2}\right) = \Pr(-2 < Y < 2)$$
$$= 2\Pr(Y < 2) - 1 = 2(0.977) - 1 = 0.954$$

◀

**EXAMPLE 6.3**

A bottling machine can be regulated so that it discharges an average of $\mu$ ounces per bottle. It has been observed that the amount of fill dispensed by the machine is normally distributed with $\sigma = 1$ ounce. A sample of $n = 9$ filled bottles is randomly selected from the output of the machine on a given day (all bottled with the same machine setting), and the ounces of fill are measured for each. Find the probability that the sample mean will be within 0.3 ounce of the true mean $\mu$ for the chosen machine setting.

**Solution.** If $X_1, X_2, \ldots, X_9$ denote the ounces of fill to be observed, then we know that the $X_i$'s are normally distributed with mean $\mu$ and variance $\sigma^2 = 1$ for $i = 1, 2, \ldots, 9$. Therefore, by Theorem 2, $\bar{X}$ possesses a normal sampling distribution with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma^2_{\bar{X}} = \sigma^2/n = 1/9$. We want to find

$$\begin{aligned}\Pr(|\bar{X} - \mu| \leq 0.3) = \Pr(-0.3 \leq \bar{X} - \mu \leq 0.3) &= \Pr\left(-\frac{0.3}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right) \\ &= \Pr\left(-\frac{0.3}{1/\sqrt{9}} \leq Z \leq \frac{0.3}{1/\sqrt{9}}\right) = \Pr(-0.9 \leq Z \leq 0.9) = 2(0.1841) - 1 = 0.6318.\end{aligned}$$

Thus, the probability is only 0.6318 that the sample mean will be within 0.3 ounce of the true population mean. ◀

**EXAMPLE 6.4**

The time that it takes a randomly selected rat of a certain subspecies to find its way through a maze is a normally distributed r.v with $\mu = 1.5$ min and $\sigma = 0.35$ min. Suppose five rats are selected. Let $X_1, \ldots, X_5$ denote their times in the maze. Assuming the $X_i$'s to be a random sample from this normal distribution, what is the probability that the total time $T_o = X_1 + \cdots + X_5$ for the five is between 6 and 8 min? Determination of the probability that the sample average time $X$ is at most 2.0 min.

**Solution.** By Theorem 2, $T_o$ has a normal distribution with $\mu_o = n\mu = 5(1.5) = 7.5$ and variance $\sigma_o^2 = n\sigma^2 = 5(0.1225) = 0.6125$, so $\sigma_o = 0.783$. To standardise $T_o$, subtract $\mu_o$ and divide by $\sigma_o$:

$$\Pr(6 \leq T_o \leq 8) = \Pr\left(\frac{6 - 7.5}{0.783} \leq Z \leq \frac{8 - 7.5}{0.783}\right) = \Pr(-1.92 \leq Z \leq 0.64) = 0.7115$$

To determination of the probability that the sample average time X (a normally distributed variable) is at most 2 min, we require $\mu_{\bar{X}} = \mu = 1.5$ and $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.35/\sqrt{5} = 0.1565$. Then

$$\Pr(\bar{X} \leq 2) = \Pr\left(Z \leq \frac{2 - 1.5}{0.1565}\right) = \Pr(Z \leq 3.19) = 0.9993.$$

◀

***Theorem 4***

Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n$ from a normal distribution with mean $\mu$ and variance $\sigma^2$. Then $Z_i = (X_i - \mu)/\sigma$ are independent, standard normal random variables, $i = 1, 2, \ldots, n$, and

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$$

has a $\chi^2$ distribution with $n$ degrees of freedom (df).

**EXAMPLE 6.5**

If $Z_1, Z_2, \ldots, Z_6$ denotes a r.s from the standard normal distribution, find a number $b$ such that

$$\Pr\left(\sum_{i=1}^{n} Z_i^2 \leq b\right) = 0.95.$$

**Solution.** By Theorem 4,$\sum_{i=1}^{6} Z_i^2$ has a $\chi^2$ distribution with 6 df. Looking at the tables, in the row headed 6 df and the column headed $\chi^2_{0.05}$, we see the number 12.5916. Thus,

$$\Pr\left(\sum_{i=1}^{6} Z_i^2 > 12.9516\right) = 0.05 \equiv \Pr\left(\sum_{i=1}^{6} Z_i^2 \leq 12.9516\right) = 0.95,$$

and $b = 12.5916$ is the 0.95 quantile (95th percentile) of the sum of the squares of six independent standard normal random variables. ◀

***Theorem 5*: (The Distribution of $\frac{(n-1)S^2}{\sigma^2}$)**

Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a normal distribution with mean $\mu$ and variance $\sigma^2$. Then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

has a $\chi^2$ distribution with $(n-1)$ df. Also, $\bar{X}$ and $S^2$ are independent r.v's, i.e, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

**EXAMPLE 6.6**

Let $X_1, X_2, \ldots, X_6$ be a r.s of size 6 from $N(\mu, 10)$. Find $\Pr(2.3 < S^2 < 22.2)$

**Solution.** Since $X_i \sim N(\mu, 10)$, for $i = 1, 2, \ldots, 6$. Then, according Theorem 5, $\frac{(n-1)S^2}{\sigma^2} = \frac{5S^2}{10} = \frac{S^2}{2} \sim \chi^2(5)$. Therefore,

$$\begin{aligned}\Pr(2.3 < S^2 < 22.2) &= \Pr\left(\frac{2.3}{2} < \frac{S^2}{2} < \frac{22.2}{2}\right) = \Pr(1.15 < Y < 11.1)\\ &= \Pr(Y < 11.1) - Pr(Y < 1.15) = 0.95 - 0.05 = 0.9.\end{aligned}$$

◀

## 6.3 The Central Limit Theorem

If we sample from a normal population, Theorem 3 tells us that $\bar{X}$ has a normal sampling distribution. But what can we say about the sampling distribution of $\bar{X}$ if the variables $X_i$ are not normally distributed? Fortunately, $\bar{X}$ will have a sampling distribution that is approximately normal if the sample size is large. The formal statement of this result is called the central limit theorem.

***Theorem*:(The Central Limit Theorem)**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean $\mu$ and variance $\sigma^2$. Then, in the limit as $n \to \infty$, the standardised versions of $\bar{X}$ and $T_o$ have the standard normal distribution. That is,

$$\lim_{n\to\infty} \Pr\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Pr(Z \leq z)$$

or

$$\lim_{n\to\infty} \Pr\left(\frac{T_o - n\mu}{\sigma\sqrt{n}} \le z\right) = \Pr(Z \le z)$$

on another words,

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{T_o - n\mu}{\sigma\sqrt{n}} \sim N(0,1).$$

**EXAMPLE 6.7**

Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n = 75$ from $U(0,1)$. Approximate $\Pr(0.45 < \bar{X} < 0.55)$, where $\bar{X}$ is the sample mean.

**Solution.** Since $X \sim U(0,1)$, then $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$. Then according to the central limit theorem, the r.v $Y = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ has a limiting $N(0,1)$, that is

$$Y = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{\sqrt{75}(\bar{X} - \mu)}{\sqrt{1/12}} = \sqrt{(75)(12)}(\bar{X} - 0.5) = 30(\bar{X} - 0.5).$$

The approximate value of

$$\begin{aligned}\Pr(0.45 < \bar{X} < 0.55) &= \Pr[30(0.45 - 0.5) < 30(\bar{X} - 0.5) < 30(0.55 - 0.5)] \\ &= \Pr(-1.5 < Y < 1.5) = 2\Pr(Y \le 1.5) - 1 \\ &= 2(0.933) - 1 = 0.866.\end{aligned}$$

◄

**EXAMPLE 6.8**

The service times for customers coming through a checkout counter in a retail store are independent random variables with mean 1.5 minutes and variance 1.0. Approximate the probability that 100 customers can be served in less than 2 hours of total service time.

**Solution.** If we let $X_i$ denote the service time for the $i^{th}$ customer, then we want

$$\Pr\left(\sum_{i=1}^{100} X_i \le 120\right) = \Pr\left(\bar{X} \le \frac{120}{100}\right) = \Pr(\bar{X} \le 1.2).$$

Because the sample size is large, the central limit theorem tells us that $\bar{X}$ is approximately normally distributed with mean $\mu_{\bar{X}}\bar{X} = \mu = 1.5$ and variance $\sigma^2_{\bar{X}} = \sigma^2/n = 1/100$. Therefore, using the tables, we have

$$\begin{aligned}\Pr(\bar{X} \le 1.2) = \Pr\left(\frac{\bar{X} - 1.5}{1/\sqrt{100}} \le \frac{1.2 - 1.5}{1/\sqrt{100}}\right) \\ \approx \Pr[Z \le (1.2 - 1.5)10] = \Pr(Z \le -3) = 0.0013.\end{aligned}$$

Thus, the probability that 100 customers can be served in less than 2 hours is approximately 0.0013. This small probability indicates that it is virtually impossible to serve 100 customers in only 2 hours. ◄

# CHAPTER 7

# POINT ESTIMATION

## 7.1 Introduction

Given a parameter of interest, such as a population mean $\mu$ or population proportion $p$, the objective of point estimation is to use a sample to compute a number that represents in some sense a good guess for the true value of the parameter. The resulting number is called a *point estimate*. In Section 7.1, we present some general concepts of point estimation. In Section 7.2, we describe and illustrate two important methods for obtaining point estimates: the method of moments and the method of maximum likelihood.

Obtaining a point estimate entails calculating the value of a statistic such as the sample mean $\bar{X}$ or sample standard deviation $S$. We should therefore be concerned that the chosen statistic contains all the relevant information about the parameter of interest. The idea of no information loss is made precise by the concept of sufficiency.

## 7.2 General Concepts and Criteria

Statistical inference is frequently directed toward drawing some type of conclusion about one or more parameters (population characteristics). To do so requires that an investigator obtain sample data from each of the populations under study. Conclusions can then be based on the computed values of various sample quantities.

***Definition*: A Point Estimator** of a parameter $\theta$ is a single number that can be regarded as a sensible value for $\theta$. A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of $\theta$.

***Notes*:**

1. we will denote to the estimate of an unknown parameter $\theta$ by $\hat{\theta}$.

2. Many different estimators (rules for estimating) may be obtained for the same population parameter.
3. The sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is one possible point estimator of the population mean $\mu$.

**EXAMPLE 7.1**

An automotive manufacturer has developed a new type of bumper, which is supposed to absorb impacts with less damage than previous bumpers. The manufacturer has used this bumper in a sequence of 25 controlled crashes against a wall, each at 10 mph, using one of its compact car models. Let $X =$ the number of crashes that result in no visible damage to the automotive. The parameter to be estimated is $p =$ the proportion of all such crashes that result in no damage [alternatively, $p = \Pr(\text{no damage in a single crash})$]. If $X$ is observed to be $x = 15$, the most reasonable estimator and estimate are

$$\text{estimator } \hat{p} = \frac{X}{n} \qquad \text{estimate } = \frac{x}{n} = \frac{15}{25} = 0.6$$

If for each parameter of interest there were only one reasonable point estimator, there would not be much to point estimation. In most problems, though, there will be more than one reasonable estimator.

### 7.2.1 The Bias and Mean Square Error of Point Estimators

Suppose we have two measuring instruments; one instrument has been accurately calibrated, but the other systematically gives readings smaller than the true value being measured. When each instrument is used repeatedly on the same object, because of measurement error, the observed measurements will not be identical. However, the measurements produced by the first instrument will be distributed about the true value in such a way that on average this instrument measures what it purports to measure, so it is called an unbiased instrument. The second instrument yields observations that have a systematic error component or bias. In other words, we would like the mean or expected value of the distribution of estimates to equal the parameter estimated; that is, $E[\hat{\theta}] = \theta$. Point estimators that satisfy this property are said to be *unbiased*.

***Definition*:** Let $\hat{\theta}$ be a point estimator for a parameter $\theta$. Then $\hat{\theta}$ is an **unbiased estimator** if $E[\hat{\theta}] = \theta$. If $E[\hat{\theta}] \neq \theta$, $\hat{\theta}$ is said to be biased. the difference $B(\hat{\theta}) = E[\hat{\theta}] - \theta$ is called the bias of $\hat{\theta}$.

In the best of all possible worlds, we could find an estimator $\hat{\theta}$ for which $\hat{\theta} = \theta$ always. However, $\hat{\theta}$ is a function of the sample $X_i$'s, so it is a random variable. For some samples, $\hat{\theta}$ will yield a value larger than $\theta$, whereas for other samples $\hat{\theta}$ will underestimate $\theta$. If we write

$$\hat{\theta} = \theta + \text{error of estimation}$$

then an accurate estimator would be one resulting in small estimation errors, so that estimated values will be near the true value.

A popular way to quantify the idea of $\hat{\theta}$ being close to $\theta$ is to consider the squared error $(\hat{\theta} - \theta)^2$. An omnibus measure of accuracy is the mean squared error (expected squared error), which entails averaging the squared error over all possible samples and resulting estimates.

***Definition*:** The **mean square error** of a point estimator $\hat{\theta}$ is

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

The mean square error of an estimator $\hat{\theta}$, $MSE(\hat{\theta})$, is a function of both its variance and its bias. If $B(\hat{\theta})$ denotes the bias of the estimator $\hat{\theta}$, it can be shown that, since

$$Var(X) = E[X^2] - (E[X])^2 \;\Rightarrow\; E[X^2] = Var(X) + (E[X])^2,$$

then,

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [B(\hat{\theta})]^2.$$

**EXAMPLE 7.2**

Consider once again estimating a population proportion of "successes" $p$. The natural estimator of $p$ is the sample proportion of successes $\hat{p} = X/n$. The number of successes $X$ in the sample has a binomial distribution with parameters $n$ and $p$, so $E[X] = np$ and $Var(X) = np(1-p)$. The expected value of the estimator is

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{1}{n}E[X] = \frac{1}{n}np = p.$$

Thus the bias of $\hat{p}$ is $p - p = 0$, and therefore $\hat{p}$ is an unbiased estimator. Giving the mean squared error as

$$E[(\hat{p}-p)^2] = Var(\hat{p}) + 0^2 = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(X) = \frac{p(1-p)}{n}.$$

Now consider the alternative estimator $\hat{p} = (X+2)/(n+4)$. That is, add two successes and two failures to the sample and then calculate the sample proportion of successes. The bias of the alternative estimator is

$$E\left[\frac{X+2}{n+4}\right] - p = \frac{1}{n+4}E[X+2] - p = \frac{np+2}{n+4} - p = \frac{2/n - 4p/n}{1+4/n}.$$

This bias is not zero unless $p = 0.5$. The variance of the estimator is

$$Var\left(\frac{X+2}{n+4}\right) = \frac{1}{(n+4)^2}Var(X+2) = \frac{Var(X)}{(n+4)^2} = \frac{np(1-p)}{(n+4)^2} = \frac{p(1-p)}{n+8+16/n}.$$

This variance approaches zero as the sample size increases. The mean squared error of the alternative estimator is

$$MSE(\hat{p}) = \frac{p(1-p)}{n+8+16/n} + \left(\frac{2/n - 4p/n}{1+4/n}\right)^2.$$

### 7.2.2 Some Common Unbiased Point Estimators

In this section, we focus on some estimators that merit consideration on the basis of intuition. For example, it seems natural to use the sample mean $\bar{X}$ to estimate the population mean $\mu$ and $S^2$ as the sample variance unbiased estimator of $\sigma^2$. Also to use the sample proportion $\hat{p} = x/n$ to estimate a binomial parameter $p$.

***Propositions*:**

1. When $X$ is a binomial r.v with parameters $n$ and $p$, the sample proportion $\hat{p} = X/n$ is an unbiased estimator of $p$.
2. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean $\mu$, then $\bar{X}$ is an unbiased estimator of $\mu$.

**EXAMPLE 7.3**

Let $X_1, X_2, \ldots, X_n$ be a random sample with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$. Show that

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

is an unbiased estimator for $\sigma^2$.

**Solution.** It can be shown that

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

Hence

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

Notice that $E[X_i^2]$ is the same for $i=1,2,\ldots,n$. We use this and the fact that the variance of a r.v is given by $Var(X)=E[X^2]-(E[X])^2$ to conclude that $E[X_i^2]=Var(X_i)+(E[X_i])^2=\sigma^2+\mu^2$, $E[\bar{X}^2]=Var(\bar{X})+(E[\bar{X}])^2=\sigma^2/n+\mu^2$, and that

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

It follows that

$$E[S^2]=\frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\frac{1}{n-1}(n-1)\sigma^2=\sigma^2,$$

so we see that $S^2$ is an unbiased estimator for $\sigma^2$. ◄

***Proposition*:** Among all estimators of $\theta$ that are unbiased, choose the one that has minimum variance. The resulting $\hat{\theta}$ Is called the **minimum variance unbiased estimator** (MVUE) of $\theta$. Since $MSE=variance+(bias)^2$, seeking an unbiased estimator with minimum variance is the same as seeking an unbiased estimator that has minimum mean squared error.

***Definition*:** The standard error of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}}=\sqrt{Var(\hat{\theta})}$. If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into $s_{\hat{\theta}}$ yields the estimated standard error (estimated standard deviation) of the estimator. The estimated standard error can be denoted either by $\hat{\sigma}_{\hat{\theta}}$ (the ˆ over $\sigma$ emphasises that $\sigma_{\hat{\theta}}$ is being estimated) or by $S_{\hat{\theta}}$.

**EXAMPLE 7.4**

Assuming that breakdown voltage is normally distributed, $\hat{\mu}=\bar{X}$ is the best estimator of $\mu$. If the value of $\sigma$ is known to be 1.5, the standard error of $\bar{X}$ is $\sigma_{\bar{X}}=\sigma/\sqrt{n}=1.5/\sqrt{20}=0.335$. If, as is usually the case, the value of $\sigma$ is unknown, the estimate $\hat{\sigma}=s=1.462$ is substituted into $\sigma_{\bar{X}}$ to obtain the estimated standard error $\hat{\sigma}_{\bar{X}}=s_{\bar{X}}=s/\sqrt{n}=1.462/\sqrt{20}=0.327$.

**EXAMPLE 7.5**

Back to Example **??**, the standard error of $\hat{p}=X/n$ is

$$\sigma_{\hat{p}}=\sqrt{Var(X/n)}=\sqrt{\frac{Var(X)}{n^2}}=\sqrt{\frac{npq}{n^2}}=\sqrt{\frac{pq}{n}}.$$

Since $p$ and $q=1-p$ are unknown (else why estimate?), we substitute $\hat{p}=x/n$ and $\hat{q}=1-x/n$ into $\sigma_{\hat{p}}$, yielding the estimated standard error $\hat{\sigma}_{\hat{p}}=\sqrt{\hat{p}\hat{q}/n}=\sqrt{(0.6)(0.4)/25}=0.098$. Alternatively, since the largest value of $pq$ is attained when $p=q=0.5$, an upper bound on the standard error is $\sqrt{1/(4n)}=0.1$.

## 7.3 Methods of Point Estimation

So far the point estimators we have introduced were obtained via intuition and/or educated guesswork. We now discuss two "constructive" methods for obtaining point estimators: the method of moments and the method of maximum likelihood. By constructive we mean that the general definition of each type of estimator suggests explicitly how to obtain the estimator in any specific problem. Although maximum likelihood estimators are generally preferable to moment estimators because of certain efficiency properties, they often require significantly more computation than do moment estimators. It is sometimes the case that these methods yield unbiased estimators.

### 7.3.1 The method of Moment

The basic idea of this method is to equate certain sample characteristics, such as the mean, to the corresponding population expected values. Then solving these equations for unknown parameter values yields the estimators.

***Definition*:** Let $X_1, X_2, \ldots, X_n$ be a r.s from pdf $f(x)$. For $k = 1, 2, 3, \ldots$, the k$^{th}$ population moment, or k$^{th}$ moment of the distribution $f(x)$, is $\mu_r = E[X^k]$. The k$^{th}$ sample moment is $M_r = (1/n)\sum_{i=1}^{n} X_i^k$.

Thus the first population moment is $E[X] = \mu$ and the first sample moment is $\sum X_i/n = \bar{X}$. The second population and sample moments are $E[X^2]$ and $\sum X_i^2/n$, respectively. The population moments will be functions of any unknown parameters $\theta_1, \theta_2, \ldots$.

***Definition*:** Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x; \theta_1, \ldots, \theta_m)$, where $\theta_1, \ldots, \theta_m$ are parameters whose values are unknown. Then **the moment estimators** $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are obtained by equating the first $m$ sample moments to the corresponding first $m$ population moments and solving for $\theta_1, \ldots, \theta_m$.

**EXAMPLE 7.6**

Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n$ from $P(\lambda)$. Estimate $\lambda$ by the method of moment.

**Solution.** Since $X_i \sim P(\lambda)$, then $f(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, \ldots$. In this case, we have one unknown parameter, $\lambda$, so we set

$$\mu_r = M_r \text{ at } \lambda = \hat{\lambda}, r = 1$$

therefore,

$$\mu_1 = E[X] = \lambda \text{ and } M_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

So, $\hat{\lambda} = \bar{x}$ is the moment estimate of $\lambda$. ◀

**EXAMPLE 7.7**

Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n$ from $Exp(1/\theta)$. Estimate $\lambda$ by the method of moment.

**Solution.** Since $X \sim Exp(1/\theta)$, then $f(x) = \theta e^{-\theta x}$, $0 < x < \infty$. In this case, we have one unknown parameter, $\lambda$, so we set

$$\bar{X} = \frac{1}{\theta} \Rightarrow \hat{\theta} = \frac{1}{\bar{x}}.$$

◀

**EXAMPLE 7.8**

Let $X_1, X_2, \dots, X_n$ be a random sample from a gamma distribution with parameters $\alpha$ and $\beta$. Find the Moment estimators for $\alpha$ and $\beta$.

**Solution.** Since $X_i \sim G(\alpha, \beta)$, then $\mu = E[X] = \alpha\beta$ and $E[X^2] = \beta^2\Gamma(\alpha+2)/\Gamma(\alpha) = \beta^2(\alpha+1)\alpha$. Then The moment estimators of $\alpha$ and $\beta$ are obtained by solving

$$\bar{X} = \alpha\beta, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \alpha(\alpha+1)\beta^2.$$

Since $\alpha(\alpha+1)\beta^2 = \alpha^2\beta^2 + \alpha\beta^2$ and the first equation implies $\alpha^2\beta^2 = (\bar{X})^2$, the second equation becomes

$$\frac{1}{n}\sum X_i^2 = (\bar{X})^2 + \alpha\beta^2.$$

Now dividing each side of this second equation by the corresponding side of the first equation and substituting back gives the estimators

$$\hat{\alpha} = \frac{(\bar{X})^2}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2} = \frac{n(\bar{X})^2}{(n-1)S^2}, \qquad \hat{\beta} = \frac{\frac{1}{n}\sum X_i^2 - (\bar{X})^2}{\bar{X}} = \frac{(n-1)S^2}{n(\bar{X})^2}$$

◀

**EXAMPLE 7.9**

Let $X_1, \dots, X_n$ be a r.s from a generalized negative binomial distribution with parameters $r$ and $p$. Estimate the value of $r$ and $p$.

**Solution.** Since $E[X] = r(1-p)/p$ and $Var(X) = r(1-p)/p^2$, $E[X^2] = Var(X) + (E[X])^2 = r(1-p)(r - rp + 1)/p^2$. Equating

$$E[X] = \bar{X} \quad \text{and} \quad E[X^2] = (1/n)\sum X_i^2$$

eventually gives,

$$\hat{p} = \frac{\bar{X}}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2} = \frac{n\bar{X}}{(n-1)S^2}, \qquad \hat{r} = \frac{(\bar{X})^2}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2 - \bar{X}} = \frac{n(\bar{X})^2}{(n-1)S^2 - \bar{X}}.$$

◀

### 7.3.2 Maximum Likelihood Method

The method of maximum likelihood was first introduced by R. A. Fisher, a geneticist and statistician, in the 1920s. Most statisticians recommend this method, at least when the sample size is large, since the resulting estimators have certain desirable efficiency properties

***Definition*:** The **Likelihood Function** of a r.s $X_1, X_2, \dots, X_n$ of size $n$ from a distribution with a pdf $f(x)$ with parameters $\theta_1, \theta_2, \dots, \theta_m$, is defined to be the joint pdf of the n r.v's $X_1, X_2, \dots, X_n$ which is considered as a function of $\theta$'s and denoted by $L(\tilde{\theta}; \tilde{x})$, i.e

$$L(\tilde{\theta}; \tilde{x}) = L(\theta_1, \theta_2, \dots, \theta_m; x_1, x_2, \dots, x_n) = f(\tilde{x}; \tilde{\theta}) = \prod_{i=1}^{n} f(x_i).$$

The maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ are those values of the $\theta_i$'s that maximize the likelihood function, so that

$$L(x_1, x_2, \ldots, x_n; \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m) \geq L(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_m), \quad \forall \theta_1, \theta_2, \ldots, \theta_m$$

When the $X_i$'s are substituted in place of the $x_i$'s, the maximum likelihood estimators (mle's) result.

***Notes*:**

1. Many likelihood functions satisfy the condition that mle is the solution of the likelihood equations

$$\frac{\partial L(\tilde{\theta}; \tilde{x})}{\partial \theta_r} = 0, \text{ at } \tilde{\theta} = \hat{\tilde{\theta}}, \quad r = 1, 2, \ldots, m.$$

2. Since $L(\tilde{\theta}; \tilde{x})$ and $\ln L(\tilde{\theta}; \tilde{x})$ have their maxima at the same value of $\tilde{\theta}$, so it is some times easier to find the maximum of the logarithm of the likelihood function. In this case, the mle $\tilde{\theta}$ of $\theta$ which maximises $L(\tilde{\theta}; \tilde{x})$ may be given by the solution of

$$\frac{\partial \ln L(\tilde{\theta}; \tilde{x})}{\partial \theta_r} = 0, \text{ at } \tilde{\theta} = \hat{\tilde{\theta}}, \quad r = 1, 2, \ldots, m.$$

**EXAMPLE 7.10**

Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n$ from Bernoulli distribution with a parameter $p$ ($X \sim Ber(p)$). Estimate $p$ using the maximum likelihood method.

**Solution.** Since $X \sim Ber(p)$, then the pdf of $X$ is: $f(x) = p^x(1-p)^{1-x}$, $x = 0, 1$. The likelihood function is

$$L(p; x_1, \ldots, x_n) = f(x_1, \ldots, x_n; p) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

It is easier to maximise $\ln L(p, x_1, \ldots, x_n)$, and

$$\ln l(p; \tilde{x}) = (\sum_{i=1}^{n} x_i) \ln p + (n - \sum_{i=1}^{n}) \ln(1-p),$$

Hence,

$$\frac{\partial \ln L}{\partial p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p}$$
$$\left.\frac{\partial \ln L}{\partial p}\right|_{p=\hat{p}} = 0 \Rightarrow \frac{\sum_{i=1}^n x_i}{\hat{p}} - \frac{n - \sum_{i=1}^n x_i}{1-\hat{p}} = 0 \Rightarrow \frac{n\bar{x}}{\hat{p}} = \frac{n - n\bar{x}}{1-\hat{p}}$$

◄

Therefore,

$$n\bar{x} - n\bar{x}\hat{p} = n\hat{p} - n\bar{x}\hat{p} \Rightarrow \hat{p} = \bar{x}$$

Then, $\bar{X}$ is the mle of $p$.

**EXAMPLE 7.11**

Let $X_1, X_2, \ldots, X_n$ be a r.s from a normal distribution with mean $\mu$ and variance $\sigma^2$. Find the MLEs of $\mu$ and $\sigma^2$.

**Solution.** Since, $X \sim N(\mu, \sigma^2)$, then $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$. The likelihood function is

$$L(\nu, \sigma^2; \tilde{x}) = f(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right].$$

Further,

$$\ln L(\mu, \sigma^2; \tilde{x}) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

The MLEs of $\mu$ and $\sigma^2$ are the values that make $\ln L(\mu, \sigma^2; \tilde{x})$ a maximum. Taking derivatives with respect to $\mu$ and $\sigma^2$, we obtain

$$\frac{\partial \ln L(\mu, \sigma^2; \tilde{x})}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu),$$

and

$$\frac{\partial \ln L(\mu, \sigma^2; \tilde{x})}{\partial \sigma} = -\left(\frac{n}{2}\right)\left(\frac{1}{\sigma^2}\right) + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Setting these derivatives equal to zero and solving simultaneously, we obtain from the first equation

$$\frac{1}{\hat{\sigma}}^2\sum_{i=1}^{n}(x_i-\hat{\mu}) = 0, \quad \text{or} \quad \sum_{i=1}^{n} x_i - n\hat{\mu} = 0, \quad \text{and} \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

Substituting $\bar{x}$ for $\hat{\mu}$ in the second equation and solving for $\hat{\sigma}^2$, we have

$$-\left(\frac{n}{\hat{\sigma}^2}\right) + \frac{1}{\hat{\sigma}^4}\sum_{i=1}^{n}(x_i-\bar{x})^2 = 0, \quad \text{or} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

◀

**EXAMPLE 7.12**

Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n$ from Gamma distribution with parameters $\alpha$ and $\beta$. Find the MLEs $\alpha$ and $\beta$.

**Solution.** The pdf of a r.v $X \sim G(\alpha, \beta)$ is: $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, $0 < x < \infty$. The likelihood function is

$$L(\aleph, \beta; \tilde{x}) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-x_i/\beta} = [\Gamma(\alpha)]^{-n}\beta^{-n\alpha}\left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} e^{-\frac{1}{\beta}\sum_{i=1}^{n} x_i},$$

Also,

$$\ln L(\alpha, \beta; \tilde{x}) = -n\ln\Gamma(\alpha) - n\alpha\ln\beta + (\alpha-1)\sum_{i=1}^{n}\ln x_i - \frac{1}{\beta}\sum_{i=1}^{n} x_i.$$

Now, to maximise $\ln L(\alpha, \beta; \tilde{x})$,

$$\frac{\partial \ln L}{\partial \alpha} = -n\psi(\alpha) - n\ln\beta + \sum_{i=1}^{n} \ln x_i,$$

where, $\psi(\alpha) = \frac{d}{d\alpha}\ln\Gamma(\alpha)$. Also

$$\frac{\partial \ln L(\alpha, \beta; \tilde{x})}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} x_i.$$

Set $\frac{\partial \ln L}{\partial \alpha} = 0$ and $\frac{\partial \ln L}{\partial \beta} = 0$ at $\alpha = \hat{\alpha}$ and $\beta = \hat{\beta}$, we have

$$-n\psi(\hat{\alpha}) - n\ln\hat{\beta} + \sum_{i=1}^{n} \ln x_i = 0 \tag{7.1}$$

and

$$-\frac{n\hat{\alpha}}{\hat{\beta}} + \frac{1}{\hat{\beta}}\sum_{i=1}^{n} x_i = 0 \Rightarrow n\hat{\alpha} = \frac{n\bar{x}}{\hat{\beta}} \Rightarrow \hat{\alpha}\hat{\beta} = \bar{x}. \tag{7.2}$$

From equations (**??**) and (**??**), we get

$$-n\psi(\hat{\alpha}) - \ln\frac{\bar{x}}{\hat{\alpha}} + \sum_{i=1}^{n} \ln x_i = 0 \Rightarrow -n\psi(\hat{\alpha}) - n\ln\bar{x} + n\ln\hat{\alpha} + \sum_{i=1}^{n} \ln x_i,$$

hence

$$n[\ln\hat{\alpha} - \psi(\hat{\alpha})] = n\ln\bar{x} - \sum_{i=1}^{n} \ln x_i \tag{7.3}$$

There is no analytical solution for equation (**??**) because of $\psi(\hat{\alpha})$. In this case, we need to approximate the value of $\hat{\alpha}$ and $\hat{\beta}$ by the use of numerical methods. ◀