

King Saud University  
College of Science  
Statistics and Operations Research Department

# **PROBABILITY (I)**

## **215 STAT**

**By  
Weaam Alhadlaq**

## Topics to be covered

1. **General Review:** Basic concepts of Probability and Random variables (discrete and continuous cases), moments and moment generating function and its applications.
2. **Discrete Probability Distributions:** Uniform – Bernoulli – Binomial - Negative Binomial – Geometric – Poisson - Hyper Geometric.
3. **Continuous Probability Distributions:** Uniform – Exponential – Normal – Beta – Gamma - Chi-squared.
4. **Bivariate Distributions (discrete and continuous cases):** Joint probability function - Joint distribution function - Marginal distributions - Expectations of jointly distributed random variables - Joint moment generating function.
5. **Conditional Distribution:** Conditional distribution function - Conditional expectation - Moment generating function.
6. **Stochastic Independence of random variables:** some consequent results of independence on Joint probability functions – Covariance – Correlation.
7. **Distributions of functions of random variables (Transformations):** The case of one variable and two variables (discrete and continuous).

## References

- Probabilities Theory and Its Applications: Dr.M.Ibraheem Akeel and A.M.Abuammoh.1421 KSU.
- Introduction to the Statistical Theory: Dr.Ahmed Aodah,1411 KSU.
- Probabilities Theory: Dr. Jalal Mostafa Assayyad, 1427 A.H. haves Publications House Jeddah.
- A First Course in Probability: S. Ross – Maxwell Macmillan – 1989.
- An Introduction to Probability and Its Applications: Larson & Marx – prentice Hall – 1985.
- Modern Probability Theory and Its Applications: Parzen – John Wiley 1992.
- Probability; An Introduction: G. Grimmett and D. Welsh – Oxford – 1991.
- Probability and Mathematical Statistics; An Introduction: L. Lukacs – Academic Press – New York – 1972.

**Credits:** (3+1)

### Assessment Tasks for Students During the Semester:

	Assessment task	Date	Proportion of Total Assessment
1	First Term Exam		25%
2	Second Term Exam		25%
3	Home works and Quizzes (tutorial)		10%
4	Final Exam		40%

## Contents

Topics to be covered .....	II
References.....	III
Chapter One: Basic concepts of Probability and Random variables.....	1
1.1 Basic concepts of Probabilities .....	1
1.2 Random Variable .....	3
1.2.1 Discrete Random Variable.....	4
1.2.2 Continuous random variable.....	4
1.3 Probability Function .....	5
1.3.1 Discrete Case (Probability Mass function).....	5
1.3.2 Continuous Case (Probability Density function) .....	7
1.4 Cumulative distribution function (CDF) .....	8
1.4.1 Discrete Case .....	9
1.4.2 Continuous Case .....	10
Chapter Two: Mathematical Expectation, Moments and Moment Generating Function .....	11
2.1 Mathematical Expectation .....	11
2.1.1 Expected value of a random variable.....	11
2.1.2 Variance of a Random Variable .....	13
2.1.3 Standard deviation .....	14
2.1.4 Mean and Variance for Linear Combination .....	15
2.1.5 Chebyshev's inequality .....	15
2.2 Central Moments and Raw Moments .....	16
2.2.1 Central Moments.....	16
2.2.2 Raw Moments.....	16
2.3 Moment Generating Function .....	17
2.3.1 Finding Moments from MGF.....	18
2.3.2 Moment Generating Function for Linear Combination.....	20
Chapter Three: Frequently Used Discrete Probability Distributions .....	22

3.1 Discrete Uniform Distribution .....	22
3.2 Bernoulli Distribution .....	23
3.3 Binomial Distribution.....	23
3.4 Geometric Distribution .....	27
3.5 Negative Binomial Distribution.....	29
3.6 Hypergeometric Distribution .....	31
3.6.1 Binomial Approximation to Hypergeometric Distribution.....	34
3.7 Poisson Distribution .....	35
3.7.1 Poisson Approximation to Binomial Distribution.....	38
Chapter Four: Frequently Used Continuous Probability Distributions .....	40
4.1 Uniform Distribution .....	40
4.2 Exponential Distribution.....	42
4.2.1 Lack of Memory Property (Memoryless) .....	45
4.2.2 Relation Between Exponential and Poisson Distributions.....	46
4.3 Gamma Distribution.....	47
4.3.1 Gamma function .....	47
4.3.2 Some useful Properties for Gamma distribution.....	47
4.3.3 Definition of Gamma Distribution .....	48
4.3.4 Special Cases.....	49
4.3.5 Incomplete Gamma Function .....	50
4.4 Chi-squared Distribution .....	52
4.5 Beta Distribution .....	53
4.6 Normal Distribution.....	54
4.6.1 Some properties of the normal curve $f(x)$ of $N(\mu, \sigma)$ .....	55
4.6.2 Standard Normal Distribution .....	55
4.6.3 Calculating Probabilities of Standard Normal Distribution .....	56
4.6.4 Calculating Probabilities of Normal Distribution .....	57
4.6.3 Normal Approximation to Binomial.....	58

4.7 Student's t Distribution.....	60
4.7.1 Some properties of the t distribution .....	60
Chapter Five: Joint, Marginal, and Conditional Distributions.....	61
5.1 Joint Distributions.....	61
5.1.1 Joint Probability function .....	61
5.1.2 Joint distribution function.....	62
5.1.3 Marginal probability Distributions .....	63
5.1.4 Joint Mathematical Expectation .....	63
5.1.5 Joint Moment Generating Function .....	66
5.2 Conditional Distributions.....	67
5.2.1 Conditional Probability function .....	67
5.2.2 Conditional Distribution Function .....	67
5.2.3 Conditional Expectation .....	68
5.2.4 Moment Generating Function for Conditional Distributions.....	70
Chapter Six: Covariance, Correlation, Independence of Variables (Stochastic Independence) .....	71
6.1 Covariance of random variables.....	71
6.2 Correlation Coefficient.....	72
6.3 Independence of random variables .....	73
Chapter Seven: Distributions of Functions of Random Variables (Transformations) .....	77
7.1 Discrete Case.....	77
7.1.1 The Case of One Variable .....	77
7.1.2 The Case of Two Variables .....	78
7.2 Continuous Case .....	79
7.2.1 Distribution Function Method (CDF).....	80
7.2.2 Change-of-Variable Method.....	81
7.2.3 Moment-Generating Function Method.....	84
Appendix A.....	87

## Chapter One

### Basic concepts of Probability and Random variables

#### 1.1 Basic concepts of Probabilities

##### An Experiment

An experiment is some procedure (or process) that we do.

##### Sample Space

The sample space of an experiment is the set of all possible outcomes of an experiment. Also, it is called the universal set, and is denoted by  $\Omega$ .

##### An Event

Any subset of the sample space  $A \subset \Omega$  is called an event.

##### Note

$\varphi \subset \Omega$ : Is the impossible event.

$\Omega \subset \Omega$ : Is the sure event.

##### Complement of an Event

The complement of the event  $A$  is denoted by  $A^c$  or  $\bar{A}$ . The event  $\bar{A}$  consists of all outcomes of  $\Omega$  but are not in  $A$ .

##### Probability

Probability is a measure (or number) used to measure the chance of the occurrence of some event. This number is between 0 and 1.

##### Equally Likely Outcomes

The outcomes of an experiment are equally likely if the outcomes have the same chance of occurrence.

##### Probability of an Event

If the experiment has  $n(\Omega)$  equally likely outcomes, then the probability of the event

$E$  is denoted by  $P(E)$  and is defined by:  $P(E) = \frac{n(E)}{n(\Omega)}$ .

Thus, probabilities have the following properties

1.  $0 \leq P(A) \leq 1$  for each event  $A$ .
2.  $P(\Omega) = 1$ .
3.  $P(\varphi) = 0$ .
4.  $P(A^c) = 1 - P(A)$ .

### Some Operations on Events

Union: The event  $A \cup B$  consists of all outcomes in  $A$  **or** in  $B$  **or** in **both**  $A$  and  $B$ .

Intersection: The event  $A \cap B$  consists of all outcomes in both  $A$  **and**  $B$ .

Sub event: The event  $A$  is called a sub event of  $B$  " $A \subset B$ " if event  $B$  occurs whenever the event  $A$  occurs.

### Addition Rule

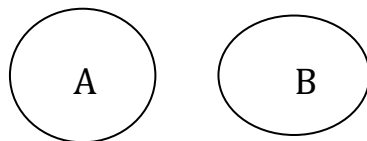
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

### Mutually exclusive (disjoint) Events

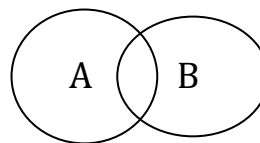
Two events are mutually exclusive if they have no sample points in common, or equivalently, if they have **empty intersection**. Events  $A_1, A_2, \dots, A_n$  are mutually exclusive if  $A_i \cap A_j = \varphi$  for all  $i \neq j$ , where  $\varphi$  denotes the empty set with no sample points.

For this case:

1.  $P(A \cap B) = 0$ .
2.  $A$  and  $A^c$  are disjoint.
3.  $P(A \cup B) = P(A) + P(B)$ . "special case of addition rule"
4.  $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ .



Disjoint Events



Joint (Not Disjoint) Events

### Exhaustive Events

The events  $A_1, A_2, \dots, A_n$  are exhaustive events if:  $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$ .

For this case:  $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(\Omega) = 1$ .



## A Partition

A collection of events  $A_1, A_2, \dots, A_n$  of a sample space  $\Omega$  is called a partition of  $\Omega$  if  $A_1, A_2, \dots, A_n$  are mutually exclusive and exhaustive events.

### Example 1.1

$\Omega = \{1,2,3,4,5,6\}$ ,  $A = \{1,2\}$ ,  $B = \{2,3\}$  and  $C = \{1,5,6\}$ . Find:  $P(A)$ ,  $P(B)$ ,  $P(C)$ ,  $P(A \cap B)$ ,  $P(B \cap C)$  and  $P(A \cup C)$ .

### Solution

$$P(A) = P(B) = \frac{1}{3}, P(C) = \frac{1}{2}, P(A \cap B) = \frac{1}{6}, P(B \cap C) = 0,$$

$$P(A \cup C) = P(A) + P(C) - P(A \cap C) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2}{3}.$$

### What can we say about B&C?

Since  $P(B \cap C) = 0$ , Then B&C are disjoint events.

## Conditional Probability

The conditional probability of A, given B, written as  $P(A|B)$ , is defined to be

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

## Independence

Two events A and B are said to be independent if B provides no information about whether A has occurred and vice versa. In symbols:

- $P(A \cap B) = P(A)P(B)$ , Or
- $P(A|B) = P(A)$ , Or
- $P(B|A) = P(B)$ .

### Return to Example 1.1. Compute $P(A|C)$ .

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

### Are A&C independent?

Since  $P(A|C) = \frac{1}{3} = P(A)$ , then A&C are independent.

## 1.2 Random Variable

The outcome of an experiment need not be a number, for example, the outcome when a coin is tossed can be 'heads' or 'tails'. However, we often want to represent

outcomes as numbers. A random variable is a function that associates a unique numerical value with every outcome of an experiment  $X: \Omega \rightarrow \mathcal{R}$ . The value of the random variable will vary from trial to trial as the experiment is repeated.

There are two types of random variable **discrete** and **continuous**.

### 1.2.1 Discrete Random Variable

The random variable  $X$  is discrete and is said to have a discrete distribution if it can take on values only from a finite  $X \in \{x_1, x_2, \dots, x_n\}$  or countable infinite sequence  $X \in \{x_1, x_2, \dots\}$ . Discrete random variables are usually represents count data, such as, the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery and the number of defective light bulbs in a box of ten.

#### Example 1.2

Consider the experiment of successive tosses of a coin. Define a variable  $X$  as  $X = 1$  if the first head occurs on an even-numbered toss,  $X = 0$  if the first head occurs on an odd-numbered toss; and define the variable  $Y$  denote the number of the toss on which the first head occurs. The sample space for this experiment is  $\Omega = \{H, TH, TTH, TTTH, \dots\}$ . Therefore,

w	$X(w)$	$Y(w)$
H	0	1
TH	1	2
TTH	0	3
TTTH	1	4
$\vdots$	$\vdots$	$\vdots$

Both  $X$  and  $Y$  are discrete random variables, where the set of all possible values of  $X$  is  $\{0,1\}$  (finite), and the set of all possible values of  $Y$  is  $\{1,2,3,4, \dots\}$  (infinite but countable).

### 1.2.2 Continuous random variable

A continuous random variable usually can assume numerical values from an interval of real numbers, perhaps the whole set of real numbers  $\mathcal{R}$ ;  $X \in \{x: a < x < b; a, b \in \mathcal{R}\}$ . Continuous random variables are usually measurements, for example, height, weight, the amount of sugar in an orange, the time required to run a mile.

## 1.3 Probability Function

### 1.3.1 Discrete Case (Probability Mass function)

The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values. It is called the probability mass function (pmf) which is usually denoted by  $f_X(x)$ ,  $f(x)$ ,  $p(x)$  or  $p_x$  and is equal to  $P(X = x)$ .

The probability mass function must satisfy

1.  $0 \leq f(x) \leq 1$  for all  $x$ ,
2.  $\sum_x f(x) = 1$ .

#### Example 1.3

Consider the following game. A fair 4-sided die, with the numbers 1, 2, 3, 4 is rolled twice. If the score on the second roll is strictly greater than the score on the first the player wins the difference in euro. If the score on the second roll is strictly less than the score on the first roll, the player loses the difference in euro. If the scores are equal, the player neither wins nor loses. If we let  $X$  denote the (possibly negative) winnings of the player, what is the probability mass function of  $X$ ?

#### Solution

The total number of outcomes of the experiment is  $4 \times 4 = 16$ . The sample space of this experiment is

$\Omega = \{(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), \dots, (4,3), (4,4)\}$ . Thus,  $X$  can take any of the values  $-3, -2, -1, 0, 1, 2, 3$ .

The pmf can be found as follow

$$\begin{aligned} f(-3) &= P(X = -3) = P\{(4,1)\} = \frac{1}{16}, \\ f(-2) &= P(X = -2) = P\{(4,2)\} + P\{(3,1)\} = \frac{2}{16}, \\ &\vdots \\ f(3) &= P(X = 3) = P\{(1,4)\} = \frac{1}{16}. \end{aligned}$$

Hence, the distribution of  $X$  can be written as

X	-3	-2	-1	0	1	2	3	Total
$f(x)$	1/16	2/16	3/16	4/16	3/16	2/16	1/16	1

#### Example 1.4

Consider an experiment of tossing an unfair coin twice. The probability is 0.6 that a coin will turn up heads on any given toss, and let  $X$  be defined as the number of

heads observed. Find the range (possible values) of  $X$ , as well as its probability mass function. Then find  $P(X = 1)$ ,  $P(X = 1.5)$ ,  $P(1 \leq x \leq 3)$ ,  $P(1 < x \leq 3)$ ,  $P(X > 4)$  &  $P(X > -2)$ .

### Solution

- I. The sample space of this experiment is  $\Omega = \{HH, HT, TH, TT\}$ , therefore, the number of heads will be 0, 1 or 2. Thus, the possible values of  $X$  are  $\{0, 1, 2\}$ .

Since this is a finite countable set, the random variable  $X$  is discrete. Next, we need to find the pmf

w	X
HH	2
HT	1
TH	1
TT	0

$P(H) = 0.6$  &  $P(T) = 0.4$ , Therefore we have

$$f(0) = P(X = 0) = P(TT) = (0.4)(0.4) = 0.16,$$

$$f(1) = P(X = 1) = P(HT) + P(TH) = (0.6)(0.4) + (0.4)(0.6) = 0.48,$$

$$f(2) = P(X = 2) = P(HH) = (0.6)(0.6) = 0.36.$$

Hence, the distribution of  $X$  can be written as

X	0	1	2	Total
$f(x)$	0.16	0.48	0.36	1

- II. Now we can calculate the probabilities as follow

1.  $P(X = 1) = 0.48$
2.  $P(X = 1.5) = 0$
3.  $P(1 \leq x \leq 3) = P(X = 1) + P(X = 2) = 0.48 + 0.36 = 0.84,$
4.  $P(1 < x \leq 3) = P(X = 2) = 0.36,$
5.  $P(X > 4) = 0.$
6.  $P(X > -2) = P(X = 0) + P(X = 1) + P(X = 2) = 1$

### Example 1.5

Suppose the range of a discrete random variable is  $\{1, 2, 3, 4\}$ . If the probability mass function is  $f(x) = cx$  for  $x = 1, 2, 3, 4$ . Find is the value of  $c$ , then calculate  $P(X = 3.25)$ ,  $P(X > 2)$ ,  $P(1 < X \leq 5)$

### Solution

I. Since  $f(x)$  is a pmf, it should satisfy two conditions

1. First,  $f(x) \geq 0 \rightarrow c \geq 0$ .
2. Second,  $\sum_x f(x) = 1 \Rightarrow f(1) + f(2) + f(3) + f(4) = 1$   
 $\Rightarrow c + 2c + 3c + 4c = 1 \Rightarrow 10c = 1 \Rightarrow c = 0.1$

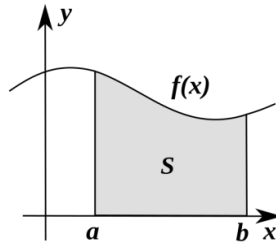
$$\Rightarrow f(x) = \frac{x}{10}; x = 1, 2, 3, 4.$$

II.

1.  $P(X = 3.25) = 0$
2.  $P(X > 2) = f(3) + f(4) = \frac{3}{10} + \frac{4}{10} = \frac{7}{10} = 0.7$
3.  $P(1 < X \leq 5) = f(2) + f(3) + f(4) = \frac{2}{10} + \frac{3}{10} + \frac{4}{10} = \frac{9}{10} = 0.9$

### 1.3.2 Continuous Case (Probability Density function)

The probability density function of a continuous random variable is a function which can be integrated to obtain the probability that the random variable takes a value in a given interval. It is called the probability density function (pdf) which is usually denoted by  $f_X(x)$  or  $f(x)$ . Symbolically,  $P(a < X < b) = \int_a^b f(x)dx$  = the area under the curve of  $f(x)$  and over the interval  $(a,b)$ .



The probability density function must satisfy

1.  $f(x) \geq 0$  or all  $x$ ,
2.  $\int_{-\infty}^{\infty} f(x)dx = 1$ .

**Note:** In the continuous case for any  $x \in \mathcal{R}$ .

1.  $f(x) \neq P(X = x)$ ,
2.  $P(X = x) = 0$ ,
3.  $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$

#### Example 1.6

$$\text{Let } f(x) = \begin{cases} 2x; & 0 < x < 1 \\ 0; & \text{otherwise} \end{cases}$$

I. Check if  $f(x)$  is pdf.

**Solution**

Since,

$$1. x > 0 \Rightarrow f(x) = 2x > 0,$$

$$2. \int_0^1 f(x)dx = \int_0^1 2x dx = 1,$$

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

II. Calculate  $P\left(\frac{1}{4} < X < \frac{1}{2}\right)$ ,  $P(X = 0.5)$  &  $P(-2 < X < 0.75)$

**Solution**

$$1. P\left(\frac{1}{4} < X < \frac{1}{2}\right) = \int_{\frac{1}{4}}^{\frac{1}{2}} f(x)dx = \int_{\frac{1}{4}}^{\frac{1}{2}} 2x dx = 0.1875.$$

$$2. P(X = 0.5) = 0$$

$$3. P(-2 < X < 0.75) = \int_{-2}^{0.75} f(x)dx = \int_0^{0.75} 2x dx = 0.5625.$$

**Example 1.7**

Let  $X$  be a continuous random variable with the following pdf  $f(x) =$

$$\begin{cases} ke^{-x}; & x \geq 0 \\ 0; & \text{otherwise} \end{cases}$$

I. Find  $k$ .

**Solution**

The probability density function must satisfy two conditions

$$1. f(x) \geq 0 \Rightarrow ke^{-x} \geq 0 \Rightarrow k \geq 0$$

$$2. \int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_0^{\infty} ke^{-x}dx = 1 \Rightarrow -k[e^{-\infty} - e^0] = 1 \Rightarrow -k[0 - 1] = 1 \\ \Rightarrow k = 1 \Rightarrow f(x) = e^{-x}; x \geq 0.$$

II. Find  $P(1 < X < 3)$ ,  $P(X > 4)$  &  $P(X \geq 4)$

**Solution**

$$1. P(1 < X < 3) = \int_1^3 e^{-x}dx = -[e^{-3} - e^{-1}] = 0.318.$$

$$2. P(X > 4) = \int_4^{\infty} e^{-x}dx = -[e^{-\infty} - e^{-4}] = -[0 - 0.0183] = 0.0183.$$

$$3. P(X \geq 4) = P(X > 4) = 0.0183.$$

### 1.4 Cumulative distribution function (CDF)

All random variables (discrete and continuous) have a cumulative distribution function (CDF) denoted by  $F(x)$ . It is a function giving the probability that the random variable  $X$  is less than or equal to  $x$ , for every value  $x$ . Formally, the cumulative distribution function  $F(x)$  is defined to be:

$$F(x) = P(X \leq x) \quad \text{for} \quad -\infty < x < \infty.$$

### 1.4.1 Discrete Case

For a discrete random variable, the cumulative distribution function is found by summing up the probabilities

$$F(x) = \sum_{t \leq x} f(t).$$

#### Example 1.8

Return to Examples 1.4, 1.5. Find the distribution function CDF for the r.v. X.

#### Solution

For example 1.4 we have

X	0	1	2	Total
$f(x)$	0.16	0.48	0.36	1
$F(x)$	0.16	0.64	1	

In a formal way,

$$F(x) = \begin{cases} 0; & x < 0 \\ 0.16; & 0 \leq x < 1 \\ 0.64; & 1 \leq x < 2 \\ 1; & x \geq 2 \end{cases}$$

Thus, we can immediately calculate:

$$P(X \leq 1) = F(1) = 0.64.$$

$$P(X > 0) = 1 - P(X \leq 0) = 1 - F(0) = 1 - 0.16 = 0.84, \text{ and so on.}$$

For example 1.5 we have

$$F(x) = P(X \leq x) = \sum_{i=1}^x \frac{i}{10};$$

$$\text{Let say we want to calculate } P(X \leq 2) = F(2) = \sum_{i=1}^2 \frac{i}{10} = \frac{1}{10} + \frac{2}{10} = \frac{3}{10} = 0.3.$$

Generally,

X	1	2	3	4	Total
$f(x)$	0.1	0.2	0.3	0.4	1
$F(x)$	0.1	0.3	0.6	1	

Formally,

$$F(x) = \begin{cases} 0; & x < 1 \\ 0.1; & 1 \leq x < 2 \\ 0.3; & 2 \leq x < 3 \\ 0.6; & 3 \leq x < 4 \\ 1; & x \geq 4 \end{cases}$$

### 1.4.2 Continuous Case

For a continuous random variable, the cumulative distribution function is the integral of its probability density function on the interval  $(-\infty, x)$ .

$$F(x) = \int_{-\infty}^x f(x)dx$$

#### Result

The different Inequalities probabilities can be written by using the CDF as follow

- $P(X \leq a) = F(a),$
- $P(X > a) = 1 - P(X \leq a) = 1 - F(a),$
- $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$

#### Example 1.9

Return to Examples 1.6,1.7. Find the distribution function CDF for the r.v. X.

#### Solution

**For example 1.6 we have**

$$F(x) = \int_0^x 2x \, dx = x^2; \quad 0 < x < 1,$$

Formally,

$$F(x) = \begin{cases} 0; & x < 0 \\ x^2; & 0 \leq x < 1. \\ 1; & x \geq 1 \end{cases}$$

Now, we can immediately calculate the probabilities on the form  $P(X < x)$  or  $P(X \leq x)$ , such as

$$P(X < 0.5) = F(0.5) = 0.5^2 = 0.25,$$

$$P(X < 3) = 1,$$

$$P(X < -2) = 0.$$

**For example 1.7 we have**

$$F(x) = \int_0^x e^{-x} \, dx = -[e^{-x} - e^0] = 1 - e^{-x}; \quad x > 0,$$

Formally,

$$F(x) = \begin{cases} 0; & x < 0 \\ 1 - e^{-x}; & x \geq 0 \end{cases}$$

Now, let find  $P(X < 4) = F(4) = 1 - e^{-4} = 0.9817,$

$$P(0.25 < X < 0.5) = F(0.5) - F(0.25) = 1 - e^{-0.5} - (1 - e^{-0.25}) = 0.1723.$$



## Chapter Two

### Mathematical Expectation, Moments and Moment Generating Function

#### 2.1 Mathematical Expectation

In this section, we learn a general definition of mathematical expectation, as well as some specific mathematical expectations, such as the mean and variance.

##### 2.1.1 Expected value of a random variable

For a random variable  $X$ , the expected value (mean, average, a predicted value of a variable) is denoted  $E(X)$ ,  $\mu_X$  or  $\mu$ .

###### Discrete case

If  $f(x)$  is the pmf of the discrete random variable  $X$ , then the expected value of  $X$  is

$$\mu = E(X) = \sum_x x f(x) = \sum_x x P(X = x).$$

###### Continuous case

If  $f(x)$  is the pdf of the continuous random variable  $X$ , then the expected value of  $X$  is  $E(X) = \int_{-\infty}^{\infty} x f(x) dx$ .

**Note:** although the integral is written with lower limit  $-\infty$  and upper limit  $\infty$ , the interval of integration is the interval of non-zero-density for  $X$ .

##### Example 2.1

Compute the expected values of the r.v.'s which presented in Examples 1.4 & 1.6.

###### Solution

For Example 1.4, the expected value is calculated by  $E(X) = \sum_x x f(x)$ . Thus,

X	-3	-2	-1	0	1	2	3	Total
$f(x)$	1/16	2/16	3/16	4/16	3/16	2/16	1/16	1
$xf(x)$	-3/16	-4/16	-3/16	0	3/16	4/16	3/16	$E(X) = 0$

For Example 1.6, the pmf is  $f(x) = \frac{x}{10}$ ;  $x = 1, 2, 3, 4$ . Then

$$\begin{aligned} E(X) &= \sum_x x f(x) = 1.f(1) + 2.f(2) + 3.f(3) + 4.f(4) \\ &= 0.1 + 0.4 + 0.9 + 1.6 = 3. \end{aligned}$$

**Example 2.2**

Compute the expected values of the r.v.'s which presented in Examples 1.6 & 1.7.

**Solution**

For Example 1.7, the pdf is  $f(x) = 2x$ ;  $0 < x < 1$ . Thus,

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x (2x) dx = \frac{2}{3}.$$

For Example 1.8, the pdf is  $f(x) = e^{-x}$ ;  $x > 0$ . Hence,

$$E(X) = \int_0^{\infty} x e^{-x} dx$$

Use integration by parts:

$$\begin{aligned} u &= x & dv &= e^{-x} dx \\ du &= dx & v &= -e^{-x} \end{aligned}$$

Hence,

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

**Expectation of a function**

If  $g$  is a function, then  $E(g(X))$  is equal to  $\sum_x g(x)f(x)$  if  $X$  is a discrete random variable, and it is equal to  $\int_{-\infty}^{\infty} g(x)f(x)dx$  if  $X$  is a continuous random variable.

**Corollary**

Let  $a, b \in \mathcal{R}$  are two constants and  $g_1, g_2$  are two functions of a random variable  $X$ .

Then

- $E(aX \pm b) = aE(X) \pm b.$
- $E(ag_1(X) + bg_2(X)) = aE(g_1(X)) + bE(g_2(X)).$

**Corollary**

If  $X_1, X_2, \dots, X_n$  are  $n$  independent r.v.'s and  $g_1, g_2, \dots, g_n$  are any functions then

$$\begin{aligned} E[g_1(X_1) \cdot g_2(X_2) \cdot \dots \cdot g_n(X_n)] &= E[g_1(X_1)] \cdot E[g_2(X_2)] \cdot \dots \cdot E[g_n(X_n)] \\ &= \prod_{i=1}^n E[g_i(X_i)]. \end{aligned}$$

**Example 2.3**

Compute the expected values of  $g(X) = X^2 - 1$  and  $h(X) = 3X + 2$  in Examples 1.3 & 1.5.

**Solution**

For Example 1.3, the expected value is calculated by

$E(g(X)) = E(X^2 - 1) = E(X^2) - 1$ . Thus,

X	-3	-2	-1	0	1	2	3	Total
$f(x)$	1/16	2/16	3/16	4/16	3/16	2/16	1/16	1
$x^2 f(x)$	9/16	8/16	3/16	0	3/16	8/16	9/16	$E(X) = 2.5$

Hence,  $E(X^2 - 1) = 2.5 - 1 = 1.5$ .

Also,  $E(3X + 2) = 3E(X) + 2 = 3(0) + 2 = 2$ .

For Example 1.5, the expected value is calculated by

$$\begin{aligned} E(g(X)) &= E(X^2 - 1) = E(X^2) - 1 = (\sum_x x^2 f(x)) - 1 \\ &= 0 + 1(0.1) + 4(0.2) + 9(0.3) + 16(0.4) - 1 = 10 - 1 = 9. \end{aligned}$$

Also,  $E(3X + 2) = 3E(X) + 2 = 3(3) + 2 = 11$ .

### Example 2.4

Compute the expected value of  $g_1(X) = X^{\frac{2}{3}}$  in Example 1.6, and the expected value of  $g_2(X) = X^2 + X$  in Example 1.7.

### Solution

For Example 1.6, the pdf is  $f(x) = 2x$ ;  $0 < x < 1$ . Thus,

$$E(g_1(X)) = E\left(X^{\frac{2}{3}}\right) = \int_{-\infty}^{\infty} x^{\frac{2}{3}} f(x) dx = \int_0^1 x^{\frac{2}{3}} (2x) dx = 2 \int_0^1 x^{\frac{5}{3}} dx = \frac{6}{8} x^{\frac{8}{3}} \Big|_0^1 = \frac{6}{8}.$$

For Example 1.7, the pdf is  $f(x) = e^{-x}$ ;  $x > 0$ . Hence,

$$E(g_2(X)) = E(X^2 + X) = E(X^2) + E(X) = \int_0^{\infty} x^2 e^{-x} dx + 1$$

Use integration by parts:

$$\begin{aligned} u &= x^2 & dv &= e^{-x} dx \\ du &= 2x dx & v &= -e^{-x} \end{aligned}$$

Hence,

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 e^{-x} dx = -x^2 e^{-x} \Big|_0^{\infty} - \int_0^{\infty} -2x e^{-x} dx = 0 + \int_0^{\infty} 2x e^{-x} dx = \\ &= 2 \int_0^{\infty} x e^{-x} dx = 2E(X) = 2(1) = 2. \end{aligned}$$

Therefore,  $E(g_2(X)) = 2 + 1 = 3$ .

## 2.1.2 Variance of a Random Variable

The variance (which denoted by  $V(X)$ ,  $\sigma_X^2$  or  $\sigma^2$ ) is a measure of the "dispersion" of  $X$  about the mean. A large variance indicates significant levels of probability or

density for points far from  $E(X)$ . The variance is always positive ( $\sigma^2 \geq 0$ ). It is defined as the average of the squared differences from the Mean. Symbolically,

$$V(X) = E[(X - \mu_X)^2],$$

This is equivalent to

$$V(X) = E(X^2) - \mu_X^2 = E(X^2) - [E(X)]^2.$$

$$\text{Where } E(X^2) = \begin{cases} \sum_x x^2 f(x) & \text{if } X \text{ is discrete r.v.} \\ \int_{-\infty}^{\infty} x^2 f(x) & \text{if } X \text{ is continuous r.v.} \end{cases}$$

The second formula is commonly used in calculations.

### 2.1.3 Standard deviation

The standard deviation of the random variable  $X$  is the square root of the variance, and is denoted  $\sigma_X = \sqrt{\sigma_X^2} = \sqrt{V(X)}$ .

#### Corollary

Let  $a, b \in \mathcal{R}$  are two constants. If  $X$  is a random variable with variance  $V(x)$ , then

$$\blacksquare \quad V(aX \pm b) = a^2 V(X).$$

#### Example 2.5

Compute the variance and standard deviation of the r.v.'s which presented in Examples 1.3. Then, calculate  $V(X - 6)$ .

#### Solution

$$V(X) = E(X^2) - [E(X)]^2$$

From Example 2.1, we found that  $E(X) = 0$

Now, let calculate  $E(X^2)$

X	-3	-2	-1	0	1	2	3	Total
$f(x)$	1/16	2/16	3/16	4/16	3/16	2/16	1/16	1
$x^2 f(x)$	9/16	8/16	3/16	0	3/16	8/16	9/16	$E(X^2) = 2.5$

$$\text{Then, } V(X) = E(X^2) - [E(X)]^2 = 2.5 - 0^2 = 2.5, \sigma_X = \sqrt{2.5} = 1.58,$$

$$\text{and } V(X - 6) = V(X) = 2.5.$$

#### Example 2.6

Compute the variance and standard deviation of the r.v.'s which presented in Examples 1.6. Then, calculate  $V(3X - 6)$ .

#### Solution

From Example 2.2, we found that  $E(X) = \frac{2}{3}$ .

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) = \int_0^1 x^2 (2x) dx = \frac{1}{2} = 0.5.$$

$$\Rightarrow V(X) = E(X^2) - [E(X)]^2 = 0.5 - \left(\frac{2}{3}\right)^2 = 0.0556, \sigma_X = \sqrt{0.0556} = 0.2357, \text{ and}$$

$$V(3X - 6) = 3^2 V(X) = 9(0.0556) = 0.5.$$

### 2.1.4 Mean and Variance for Linear Combination

Suppose  $X_1, X_2, \dots, X_n$  are  $n$  independent random variables with means  $\mu_1, \mu_2, \dots, \mu_n$  and variances  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ . Then, the mean and variance of the linear combination  $Y = \sum_{i=1}^n a_i X_i$ , where  $a_1, a_2, \dots, a_n$  are real constants are:

$$\mu_Y = E(Y) = \sum_{i=1}^n a_i \mu_i = a_1 E(X_1) + a_2 E(X_2) + \dots + a_n E(X_n)$$

and

$$\sigma_Y^2 = V(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2 = a_1^2 V(X_1) + a_2^2 V(X_2) + \dots + a_n^2 V(X_n)$$

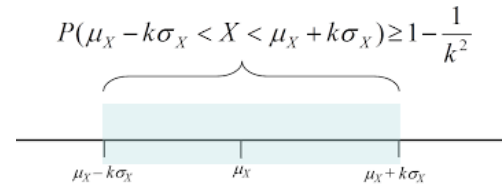
respectively.

### 2.1.5 Chebyshev's inequality

If  $X$  is a random variable with mean  $\mu$  and standard deviation  $\sigma$ , then for any real number  $k > 0$ ,

$$P[|X - \mu| > k\sigma] \leq \frac{1}{k^2}$$

$$\equiv P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$



#### Example 2.7

Let  $Y$  be the outcome when a single and fair die is rolled. If  $E(Y) = 3.5$  and  $V(Y) = 2.9$ . Evaluate  $P(|Y - 3.5| \geq 2.5)$ .

#### Solution

Since the distribution is unknown, we cannot compute the exact value of the probability. To estimate the probability we will use Chebyshev's inequality.

$$k\sigma = 2.5 \Rightarrow 1.7 k = 2.5 \Rightarrow k = 1.47. \text{ Thus,}$$

$$P(|Y - 3.5| \geq 2.5) \leq \frac{1}{(1.47)^2} = 0.463.$$

**Example 2.8**

Toss 100 coins and let  $X$  count the number of heads where  $E(X) = 50, V(X) = 25$ .  
Estimate the probability that  $40 \leq X \leq 60$ .

**Solution**

To estimate the probability we will use Chebyshev's inequality.

$$60 = \mu + \sigma k \Rightarrow 60 = 50 + 5k \Rightarrow 10 = 5k \Rightarrow k = 2.$$

$$P(40 < X < 60) \geq 1 - \frac{1}{2^2} = 0.75.$$

**2.2 Central Moments and Raw Moments****2.2.1 Central Moments**

The  $r^{\text{th}}$  central moment of a random variable  $X$  (moment about the mean  $\mu$ ) denoted by  $\mu_r$  is the expected value of  $(X - \mu)^r$ ; symbolically,

$$\mu_r = E[(X - \mu)^r] \quad \text{for} \quad r = 0, 1, 2, \dots$$

Therefore,

- $\mu_0 = E[(X - \mu)^0] = E(1) = 1.$
- The first central moment  $\mu_1 = E[(X - \mu)] = E(X) - \mu = 0.$
- The second central moment  $\mu_2 = E[(X - \mu)^2] = V(X) = \sigma^2$  (The Variance).
- $\vdots$

**2.2.2 Raw Moments**

The  $r^{\text{th}}$  moment about the origin of a random variable  $X$ , denoted by  $\mu'_r$ , is the expected value of  $X^r$ ; symbolically,

$$\mu'_r = E(X^r) \quad \text{for} \quad r = 0, 1, 2, \dots$$

Therefore,

- $\mu'_0 = E(X^0) = E(1) = 1.$
- The first raw moment  $\mu'_1 = E(X) = \mu$  (The expected value of  $X$ ).
- The second raw moment  $\mu'_2 = E(X^2).$
- $\vdots$

**Notes**

- It is known that  $V(X) = E(X^2) - [E(X)]^2$ , thus  $\mu_2 = \mu'_2 - \mu_1'^2$ .

- Henceforth, the term "moments" will be indicate to "raw moments".

## 2.3 Moment Generating Function

If  $X$  is a random variable, then its moment generating function (MGF) denoted by  $M_X(t)$  or  $M(t)$  is defined as

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f(x) & \text{if } X \text{ is a discrete } r.v \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{if } X \text{ is a continuous } r.v \end{cases}$$

We say that MGF of  $X$  exists, if there exists a positive constant  $h$  such that  $M_X(t)$  is finite for all  $t \in [-h, h]$ .

### Notes

- We always have  $M_X(0) = E[e^{0X}] = 1$ .
- There are some cases where MGF does not exist.

### Example 2.9

For each of the following random variables, find the MGF.

I.  $X$  is a discrete random variable, with pmf

$$f(x) = \begin{cases} \frac{1}{3} & , \quad x = 1 \\ \frac{2}{3} & , \quad x = 2 \end{cases}$$

### Solution

$$M_X(t) = E(e^{tX}) = \sum_{x=1}^2 e^{tx} f(x) = e^{t(1)} f(1) + e^{t(2)} f(2) = \frac{1}{3} e^t + \frac{2}{3} e^{2t}.$$

II.  $Y$  is a random variable, with pdf

$$f(y) = 1, \quad 0 < y < 1$$

### Solution

$$M_Y(t) = E(e^{tY}) = \int_0^1 e^{ty} f(y) dy = \int_0^1 e^{ty} dy = \frac{e^{ty}}{t} \Big|_0^1 = \frac{e^t - e^0}{t} = \frac{e^t - 1}{t}.$$

### Why is the MGF useful?

There are basically two reasons for this:

- First, the MGF of  $X$  gives us all moments of  $X$ . That is why it is called the moment generating function.

- Second, the MGF (if it exists) uniquely determines the distribution. That is, if two random variables have the same MGF, then they must have the same distribution. Thus, if you find the MGF of a random variable, you have indeed determined its distribution.

### 2.3.1 Finding Moments from MGF

Remember the Taylor series for  $e^x$ : for all  $x \in R$ , we have

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Now, we can write

$$e^{tX} = \sum_{k=0}^{\infty} \frac{(tX)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k x^k}{k!} = 1 + tx + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + \cdots.$$

Thus, we have

$$M_X(t) = E[e^{tX}] = \sum_{k=0}^{\infty} E(X^k) \frac{t^k}{k!} = 1 + E(X)t + E(X^2) \frac{t^2}{2!} + E(X^3) \frac{t^3}{3!} + \cdots.$$

#### Proposition

The  $r^{th}$  moment about the origin can be found by evaluating the  $r^{th}$  derivative of the moment generating function at  $t = 0$ . That is

$$\frac{d^r}{dt^r} M(t)|_{t=0} = M^r(0) = E(X^r) = \mu'_r$$

#### Example 2.10

Let  $X$  be a r.v. with MGF  $M_X(t) = \left(\frac{1}{3}e^t + \frac{2}{3}\right)^{10}$ . Drive the first and the second moments of  $X$ .

#### Solution

By using the previous proposition we can find the moments as follow

- The first moment is

$$\begin{aligned} E(X) = \mu &= \frac{d}{dt} M(t)|_{t=0} = \frac{d}{dt} \left(\frac{1}{3}e^t + \frac{2}{3}\right)^{10} \Big|_{t=0} = 10 \left(\frac{1}{3}e^t + \frac{2}{3}\right)^9 \left(\frac{1}{3}e^t\right) \Big|_{t=0} \\ &= \frac{10}{3} \left(\frac{1}{3} + \frac{2}{3}\right)^9 = \frac{10}{3} = 3.33. \end{aligned}$$

- The second moment is

$$\begin{aligned} E(X^2) = \mu'_2 &= \frac{d^2}{dt^2} M(t)|_{t=0} = \frac{d^2}{dt^2} \left(\frac{1}{3}e^t + \frac{2}{3}\right)^{10} \Big|_{t=0} = \frac{d}{dt} 10 \left(\frac{1}{3}e^t + \frac{2}{3}\right)^9 \left(\frac{1}{3}e^t\right) \Big|_{t=0} \\ &= \left[ 90 \left(\frac{1}{3}e^t + \frac{2}{3}\right)^8 \left(\frac{1}{3}e^t\right)^2 + 10 \left(\frac{1}{3}e^t + \frac{2}{3}\right)^9 \left(\frac{1}{3}e^t\right) \right] \Big|_{t=0} = \frac{90}{9} + \frac{10}{3} = 13.33. \end{aligned}$$



**Example 2.11**

Let  $X$  be a r.v. with MGF  $M_X(t) = (1 - 2t)^{-\frac{1}{2}}$ . Drive the Mean (expected value) and the standard deviation of  $X$ .

**Solution**

- The mean is

$$E(X) = \mu = \frac{d}{dt} M(t)|_{t=0} = \frac{d}{dt} (1 - 2t)^{-\frac{1}{2}}|_{t=0} = -\frac{1}{2} (1 - 2t)^{-\frac{3}{2}} (-2)|_{t=0} = 1.$$

- The variance is

$$V(X) = E(X^2) - \mu^2$$

$$E(X^2) = \frac{d^2}{dt^2} M(t)|_{t=0} = \frac{d}{dt} (1 - 2t)^{-\frac{3}{2}}|_{t=0} = -\frac{3}{2} (1 - 2t)^{-\frac{5}{2}} (-2)|_{t=0} =$$

$$3(1 - 2t)^{-\frac{5}{2}}|_{t=0} = 3.$$

$$V(X) = E(X^2) - \mu^2 \Rightarrow V(X) = 3 - 1^2 = 2.$$

Hence, the standard deviation is  $\sigma_X = \sqrt{2} = 1.41$ .

**Example 2.12**

Find the MGF of the r.v.  $X$ , then use it to find the first four moments. Where

$$f(x) = \frac{x}{2}; \quad 0 < x < 2$$

**Solution**

$$M_X(t) = E(e^{tX}) = \int_0^2 \frac{xe^{tx}}{2} dx.$$

Use integration by parts:

$$u = \frac{x}{2} \quad dv = e^{tx} dx$$

$$du = \frac{1}{2} dx \quad v = \frac{e^{tx}}{t}$$

Hence,

$$M_X(t) = \frac{xe^{tx}}{2t} \Big|_0^2 - \int_0^2 \frac{e^{tx}}{2t} dx = \frac{e^{2t}}{t} - \frac{e^{tx}}{2t^2} \Big|_0^2 = \frac{e^{2t}}{t} - \frac{e^{2t}}{2t^2} + \frac{1}{2t^2}.$$

Since the derivative of  $M_X(t)$  does not exist at  $t = 0$ , we will use the Taylor series form. Thus, we have to put the MGF on the form

$$M_X(t) = 1 + E(X)t + E(X^2)\frac{t^2}{2!} + E(X^3)\frac{t^3}{3!} + \dots$$

We have

$$\begin{aligned} M_X(t) &= \frac{1}{2t^2} + \frac{e^{2t}}{t} - \frac{e^{2t}}{2t^2} = \frac{1}{2t^2} + \frac{(2t-1)}{2t^2} e^{2t} \\ &= \frac{1}{2t^2} + \frac{(2t-1)}{2t^2} \left[ 1 + \frac{2}{1!}t + \frac{2^2}{2!}t^2 + \frac{2^3}{3!}t^3 + \frac{2^4}{4!}t^4 + \frac{2^5}{5!}t^5 + \frac{2^6}{6!}t^6 + \dots \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2t^2} + \frac{(2t-1)}{2t^2} + \frac{(2t-1)}{t} + \frac{2(2t-1)}{2!} + \frac{2^2 t(2t-1)}{3!} + \frac{2^3 t^2(2t-1)}{4!} + \frac{2^4 t^3(2t-1)}{5!} + \frac{2^5 t^4(2t-1)}{6!} + \dots \\
&= \frac{1}{2t^2} + \frac{1}{t} - \frac{1}{2t^2} + 2 - \frac{1}{t} + 2t - 1 + \frac{2^3 t^2}{3!} - \frac{2^2 t}{3!} + \frac{2^4 t^3}{4!} - \frac{2^3 t^2}{4!} + \frac{2^5 t^4}{5!} - \frac{2^4 t^3}{5!} + \frac{2^6 t^5}{6!} - \frac{2^5 t^4}{6!} + \dots \\
&= 1 + \left(2 - \frac{2^2}{3!}\right)t + \left(\frac{2^3}{3!} - \frac{2^3}{4!}\right)t^2 + \left(\frac{2^4}{4!} - \frac{2^4}{5!}\right)t^3 + \left(\frac{2^5}{5!} - \frac{2^5}{6!}\right)t^4 + \dots \\
&= 1 + \frac{8}{3}t + \left(\frac{2^3}{3} - \frac{2^3}{12}\right)\frac{t^2}{2!} + \left(\frac{2^4}{4} - \frac{2^4}{20}\right)\frac{t^3}{3!} + \left(\frac{2^5}{5} - \frac{2^5}{30}\right)\frac{t^4}{4!} + \dots \\
&= 1 + \frac{8}{3}t + 2\frac{t^2}{2!} + \left(\frac{2^4}{5}\right)\frac{t^3}{3!} + \left(\frac{2^4}{3}\right)\frac{t^4}{4!} + \dots
\end{aligned}$$

Therefore, by comparing with the Taylor form the first four moments are

$$E(X) = \frac{8}{3}; \quad E(X^2) = 2; \quad E(X^3) = \frac{2^4}{5} = \frac{16}{5}; \quad E(X^4) = \frac{2^4}{3} = \frac{16}{3}.$$

### 2.3.2 Moment Generating Function for Linear Combination

#### Theorem

Suppose  $X_1, X_2, \dots, X_n$  are  $n$  independent random variables, and the random variable  $Y$  is defined as  $Y = X_1 + X_2 + \dots + X_n$ ; Then

$$M_Y(t) = M_{X_1+X_2+\dots+X_n}(t) = M_{X_1}(t)M_{X_2}(t) \dots M_{X_n}(t)$$

#### Proof

$$\begin{aligned}
M_Y(t) &= E[e^{tY}] = E[e^{t(X_1+X_2+\dots+X_n)}] = E[e^{tX_1}e^{tX_2} \dots e^{tX_n}] \\
&= E[e^{tX_1}]E[e^{tX_2}] \dots E[e^{tX_n}] \quad (\text{since } X_i\text{'s are independent}) \\
&= M_{X_1}(t)M_{X_2}(t) \dots M_{X_n}(t).
\end{aligned}$$

#### Special cases

- If  $X$  and  $Y$  are two independent r.v.'s ( $n=2$ ), then  $M_{X+Y}(t) = M_X(t)M_Y(t)$ .
- If  $X$  and  $Y$  are i.i.d. r.v.'s (independent identically distributed), then  $M_{X+Y}(t) = [M(t)]^2$ ; where  $M(t)$  is the common MGF.

#### Proposition

If  $X$  is any random variable and  $Y = a + bX$ , then

$$M_Y(t) = e^{at}M_X(bt)$$

- In particular, if  $Z = \frac{X-\mu}{\sigma}$ ; then  $M_Z(t) = e^{-\frac{\mu}{\sigma}}M_X\left(\frac{t}{\sigma}\right)$ .

**Proof**

$$\begin{aligned}
 M_Y(t) &= E(e^{tY}) = E(e^{t(a+bX)}) = E(e^{at+btX}) = E(e^{at} \cdot e^{btX}) \\
 &= e^{at} E(e^{btX}) \quad (e^{at} \text{ is a constant}) \\
 &= e^{at} M_X(bt) \quad (\text{from MGF definition}).
 \end{aligned}$$

**Example 2.13**

Let  $X$  be a discrete random variable with values in  $\{0, 1, 2, \dots\}$  and moment generating function  $M(t) = \frac{3}{3-t}$ . Find, in terms of  $M(t)$ , the generating functions for

I.  $Y = 3X + 7$

**Solution**

$$M_Y(t) = e^{7t} M_X(3t) = e^{7t} \left( \frac{3}{3-3t} \right) = e^{7t} (1-t)^{-1}.$$

II.  $W = -X$

**Solution**

$$M_W(t) = M_X(-t) = \frac{3}{3+t}.$$

**Example 2.14**

Let  $X$  and  $Y$  are two independent random variables. Find the MGF for  $Z = X + Y$ , if

I.  $M_X(t) = \frac{e^t - 1}{t}$ ,  $M_Y(t) = \frac{1}{1-4t}$ .

**Solution**

$$M_Z(t) = M_X(t) M_Y(t) = \left( \frac{e^t - 1}{t} \right) \left( \frac{1}{1-4t} \right) = \frac{e^t - 1}{t(1-4t)}.$$

II.  $M_X(t) = M_Y(t) = M(t) = \frac{1}{4} + \frac{3}{4} e^t.$

**Solution**

$$M_Z(t) = [M(t)]^2 = \left( \frac{1}{4} + \frac{3}{4} e^t \right)^2.$$

## Chapter Three

### Frequently Used Discrete Probability Distributions

#### Distributions to be Covered

- Discrete uniform distribution.
- Bernoulli distribution.
- Binomial distribution.
- Geometric distribution.
- Negative binomial distribution.
- Hypergeometric distribution.
- Poisson distribution.

#### 3.1 Discrete Uniform Distribution

The discrete uniform distribution is also known as the "equally likely outcomes" distribution.

A random variable  $X$  has a discrete uniform distribution if each of the  $k$  values in its range, say  $x_1, x_2, \dots, x_k$ , has equal probability. Then,

$$f(x) = f(x; k) = \begin{cases} \frac{1}{k}; & x = x_1, x_2, \dots, x_k, \\ 0; & \text{otherwise} \end{cases}$$

where  $k$  is a constant.

**Parameter of the Distribution:**  $k \in \mathbb{N}^+$  (number of outcomes of the experiment).

#### Mean and Variance

Suppose  $X$  is a discrete uniform random variable on the consecutive integers  $a, a + 1, a + 2, \dots, b$  for  $a \leq b$ . The mean and the variance of  $X$  are

$$E(X) = \mu = \frac{b+a}{2},$$

$$V(X) = \sigma^2 = \frac{(b-a+1)^2-1}{12}$$

#### Note

If you compute the mean and variance by their definitions, you will get the same answer.

**Example 3.1**

Let  $X$  represent a random variable taking on the possible values of  $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ , and each possible value has equal probability. Find the distribution of  $X$ . Then, calculate the mean and the variance.

**Solution**

$X$  has a discrete uniform distribution, thus

$$f(x) = \frac{1}{10}; \quad x = 0, 1, \dots, 9.$$

Therefore,  $E(X) = \frac{9+0}{2} = 4.5$  and  $V(X) = \frac{(9-0+1)^2-1}{12} = 8.25$ .

**3.2 Bernoulli Distribution**

**Bernoulli trial** It is a trial has only two outcomes, denoted by  $S$  for success and  $F$  for failure with  $P(S) = p$  and  $P(F) = q = 1 - p$ .

Suppose that  $X$  is a r.v. representing the number of successes ( $x = 0$  or  $1$ ). Therefore,  $X$  has a Bernoulli distribution ( $X \sim \text{Ber}(p)$ ) and its pmf is given by

$$f(x) = f(x; p) = \begin{cases} p^x q^{1-x}; & x = 0, 1. \\ 0; & \text{otherwise} \end{cases}$$

**Parameter of the Distribution:**  $0 \leq p \leq 1$  (probability of success).

**Mean and Variance**

If  $X$  is a discrete random variable has Bernoulli distribution with parameter  $p$  then,

$$E(X) = \mu = p \quad \text{and} \quad V(x) = \sigma^2 = pq.$$

**Example 3.2**

Let  $X \sim \text{Ber}(0.6)$ . Find the mean and the standard deviation

**Solution**

$$p = 0.6 \Rightarrow q = 1 - 0.6 = 0.4.$$

$$E(X) = p = 0.6 \quad \text{and} \quad V(X) = pq = (0.6)(0.4) = 0.24 \Rightarrow \sigma = 0.49.$$

**3.3 Binomial Distribution**

If we perform a random experiment by repeating  $n$  independent Bernoulli trials where the probability of successes is  $p$ , then the random variable  $X$  representing the number of successes in the  $n$  trials has a binomial distribution ( $X \sim \text{Bin}(n, p)$ ). The

possible values for binomial random variable  $X$  depends on the number of Bernoulli trials independently repeated, and is  $\{0, 1, 2, \dots, n\}$ . Thus, the pmf of  $X$  is given by

$$f(x) = f(x; n, p) = \begin{cases} \binom{n}{x} p^x q^{n-x}; & x = 0, 1, \dots, n, \\ 0; & \text{otherwise} \end{cases},$$

where  $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ .

**Parameters of the Distribution:**  $n \in \mathbb{N}^+$  (number of trials or sample size) and  $0 \leq p \leq 1$  (probability of success).

### Characteristics of Binomial Distribution

1. There is a fixed number,  $n$ , of identical trials.
2. The trials are independent of each other.
3. For each trial, there are only two possible outcomes (success/failure).
4. The probability of success,  $p$ , remains the same for each trial (constant).

### Mean and Variance

If  $X$  is a discrete random variable has binomial distribution with parameters  $n, p$  then,

$$E(X) = \mu = np \text{ and } V(x) = \sigma^2 = npq.$$

### Proof

**I.  $E(X) = np$ .**

$$\begin{aligned} E(X) &= \sum_{x=0}^n x f(x) = \sum_{x=0}^n x \binom{n}{x} p^x q^{n-x} \\ &= \sum_{x=1}^n x \binom{n}{x} p^x q^{n-x} \text{ (Set summation from 1, since when } x = 0 \text{ the expression} = 0) \\ &= \sum_{x=1}^n x \cdot \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=1}^n x \cdot \frac{n(n-1)!}{x(x-1)!(n-x)!} p^x q^{n-x} \\ &= n \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^x q^{n-x} \\ &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \quad (\text{Assume } Y = X - 1 \Rightarrow X = Y + 1) \\ &= np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-(y+1))!} p^y q^{n-(y+1)} = np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-y-1)!} p^y q^{n-y-1} \\ &= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y q^{(n-1)-y} \quad (\text{Assume } m = n - 1) \\ &= np \sum_{y=0}^m \binom{m}{y} p^y q^{m-y} = np (1) = np. \end{aligned}$$

(The part  $\binom{m}{y} p^y q^{m-y}$  is a binomial pmf for y successes in m trials. Hence, the summation= 1).

## II. $V(X) = npq$ .

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

$$= E(X^2 - X + X) - n^2 p^2 \quad (\text{Add and subtract } X)$$

$$= E(X^2 - X) + E(X) - n^2 p^2 = E(X^2 - X) + np - n^2 p^2.$$

Now, let's simplify the part  $E(X^2 - X)$ :

$$E(X^2 - X) = E(X(X - 1)) = \sum_{x=0}^n x(x - 1)f(x) = \sum_{x=0}^n x(x - 1)\binom{n}{x}p^x q^{n-x}$$

$$= \sum_{x=2}^n x(x - 1)\binom{n}{x}p^x q^{n-x} \quad (\text{Set summation from 2, since when } x = 0, 1 \text{ the expression} = 0)$$

$$= \sum_{x=2}^n x(x - 1) \cdot \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=2}^n x(x - 1) \cdot \frac{n(n-1)(n-2)!}{x(x-1)(x-2)!(n-x)!} p^x q^{n-x}$$

$$= n(n - 1) \sum_{x=2}^n \frac{(n-2)!}{(x-2)!(n-x)!} p^x q^{n-x}$$

$$= n(n - 1)p^2 \sum_{x=2}^n \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} \quad (\text{Assume } Z = X - 2 \Rightarrow X = Z + 2)$$

$$= n(n - 1)p^2 \sum_{Z=0}^{n-2} \frac{(n-2)!}{Z!(n-(Z+2))!} p^Z q^{n-(Z+2)}$$

$$= n(n - 1)p^2 \sum_{Z=0}^{n-2} \frac{(n-2)!}{Z!(n-Z-2)!} p^Z q^{n-Z-2}$$

$$= n(n - 1)p^2 \sum_{Z=0}^{n-2} \binom{n-2}{Z} p^Z q^{(n-2)-Z} \quad (\text{Assume } k = n - 2)$$

$$= n(n - 1)p^2 \sum_{Z=0}^k \binom{k}{Z} p^Z q^{k-Z} = n(n - 1)p^2 (1) = n(n - 1)p^2.$$

(The part  $\binom{k}{z} p^z q^{k-z}$  is a binomial pmf for z successes in k trials. Hence, the summation = 1).

Therefore,

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

$$= n^2 p^2 - np^2 + np - n^2 p^2 = np - np^2 = np(1 - p) = npq.$$

## Moment Generating Function

If X is a discrete random variable has binomial distribution with parameters n, p then, the MGF of X is

$$M_X(t) = (pe^t + q)^n$$

## Proof

**Hint:** The binomial formula gives  $\sum_{x=0}^n \binom{n}{x} u^x v^{n-x} = (u + v)^n$ .

$$M_X(t) = E(e^{tx}) = \sum_{x=0}^n e^{tx} f(x) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^n \binom{n}{x} (pe^t)^x q^{n-x} \\ = (pe^t + q)^n.$$

**Note**

The Bernoulli distribution is a special case of the binomial distribution when  $n = 1$ .

**Example 3.3**

Suppose 40% of a large population of registered voters favor the candidate Obama. A random sample of  $n = 5$  voters will be selected, and  $X$ , the number favoring Obama out of 5, is to be observed. What is the probability of getting no one who favors Obama (i.e.  $P(X = 0)$ )?. Then compute the mean and the variance.

**Solution**

$$n = 5, p = \frac{40}{100} = 0.4 \Rightarrow q = 0.6.$$

$$X \sim \text{Bin}(5, 0.4)$$

$$f(x) = \binom{5}{x} (0.4)^x (0.6)^{5-x}; x = 0, 1, 2, 3, 4, 5.$$

Hence,

$$P(X = 0) = f(0) = \binom{5}{0} (0.4)^0 (0.6)^5 = 0.0778.$$

$$E(X) = np = 5(0.4) = 2.$$

$$V(X) = npq = 5(0.4)(0.6) = 1.2.$$

**Example 3.4**

If the MGF of the r.v.  $X$  is  $M_X(t) = (0.8 + 0.2e^t)^4$ . Find  $P(X \leq 3)$ ,  $P(-1 < X < 2)$  and the mean.

**Solution**

The MGF is on the form  $(pe^t + q)^n$ , thus,  $X \sim \text{Bin}(4, 0.2)$ . So, the possible values of  $X$  are 0, 1, 2, 3, 4.

Therefore,

- $P(X \leq 3) = f(0) + f(1) + f(2) + f(3) = 1 - f(4) = 1 - \binom{4}{4} (0.2)^4 (0.8)^0 \\ = 1 - 0.0016 = 0.9984.$
- $P(-1 < X < 2) = f(0) + f(1) = \binom{4}{0} (0.2)^0 (0.8)^4 + \binom{4}{1} (0.2)^1 (0.8)^3 \\ = 0.4096 + 0.4096 = 0.8192.$
- $E(X) = 4(0.2) = 0.8.$



### 3.4 Geometric Distribution

A single trial of an experiment results in either success with probability  $p$ , or failure with probability  $q = 1 - p$ . The experiment is performed with successive independent trials until the first success occurs. If  $X$  represents the number of trials until the first success, then  $X$  is a discrete random variable that can be 1,2,3, ....  $X$  is said to have a geometric distribution with parameter  $p$  ( $X \sim \text{Geom}(0.01)$ ) and its pmf is given by

$$f(x) = f(x; p) = \begin{cases} pq^{x-1}; & x = 1, 2, \dots \\ 0; & \text{otherwise} \end{cases}$$

**Parameter of the Distribution:**  $0 \leq p \leq 1$  (probability of success).

#### Characteristics of Geometric Distribution

1. The outcome of each trial is Bernoulli, that is either a success(S) or failure(F).
2. The Probability of success is constant  $P(S) = p$ .
3. The trials are repeated until 'one' successes occur.

#### For example

- A coin is tossed until a head is obtained.
- From now we count the no. of days until we get a rainy day.

#### Mean and Variance

If  $X$  is a discrete random variable has geometric distribution with parameter  $p$  then,

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

#### Proof

#### Hint

Let  $i \in \mathcal{R}: |i| < 1$ .

- $\sum_{n=1}^{\infty} ni^{n-1} = 1 + 2i + 3i^2 + \dots = \frac{1}{(1-i)^2}.$
- $\sum_{n=1}^{\infty} n(n+1)i^{n-1} = \frac{2}{(1-i)^3}.$

I.  $E(X) = \frac{1}{p}.$

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} xf(x) = \sum_{x=1}^{\infty} x \cdot pq^{x-1} = p \sum_{x=1}^{\infty} xq^{x-1} = p(1 + 2q + 3q^2 + \dots) \\ &= \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}. \end{aligned}$$

$$\text{II. } V(X) = \frac{q}{p^2}.$$

$$V(X) = E(X^2) - [E(X)]^2 = E(X^2) - \left(\frac{1}{p}\right)^2 = E(X^2) - \frac{1}{p^2}.$$

Now,

$$\begin{aligned} E(X^2) &= \sum_{x=1}^{\infty} x^2 f(x) = \sum_{x=1}^{\infty} x^2 \cdot pq^{x-1} \\ &= \sum_{x=1}^{\infty} (x^2 + x - x) pq^{x-1} \text{ (Add and subtract } x) \\ &= \sum_{x=1}^{\infty} [x(x+1) - x] pq^{x-1} \\ &= \sum_{x=1}^{\infty} x(x+1) pq^{x-1} - \sum_{x=1}^{\infty} x pq^{x-1} = p \sum_{x=1}^{\infty} x(x+1) q^{x-1} - E(X) \\ &= p \cdot \frac{2}{(1-q)^3} - \frac{1}{p} = \frac{2p}{p^3} - \frac{1}{p} = \frac{2}{p^2} - \frac{1}{p}. \end{aligned}$$

Therefore,

$$V(X) = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2} = \frac{q}{p^2}.$$

### Moment Generating Function

If  $X$  is a discrete random variable has geometric distribution with parameter  $p$  then, the MGF of  $X$  is

$$M_X(t) = \frac{pe^t}{1-qe^t}.$$

### Proof

**Hint:** For any  $i \in \mathcal{R}$ :  $|i| < 1$ ;  $\sum_{n=1}^{\infty} i^{n-1} = \sum_{n=0}^{\infty} i^n = 1 + i + i^2 + \dots = \frac{1}{1-i}$ .

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} f(x) = \sum_{x=1}^{\infty} e^{tx} pq^{x-1} = pe^t \sum_{x=1}^{\infty} e^{t(x-1)} q^{x-1} \\ &= pe^t \sum_{x=1}^{\infty} (qe^t)^{x-1} = pe^t \cdot \frac{1}{1-qe^t} = \frac{pe^t}{1-qe^t}. \end{aligned}$$

### Example 3.5

In a certain manufacturing process it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found? Find the mean and the variance.

### Solution

Let  $X$  represents the no. of items until the first defective item is found. The probability of successes (defective item) is  $p = \frac{1}{100} = 0.01$ . Thus,  $X \sim \text{Geom}(0.01)$ . So, we want to find

$$P(X = 5) = f(5) = (0.01)(0.99)^4 = 0.0096.$$

$$\mu = \frac{1}{0.01} = 100.$$

$$\sigma^2 = \frac{0.99}{(0.01)^2} = 9900.$$

### 3.5 Negative Binomial Distribution

If  $r$  is an integer, then the negative binomial random variable  $X$  can be interpreted as being the number of trials until the  $r^{\text{th}}$  success occurs when successive independent trials of an experiment are performed for which the probability of success in a single particular trial is  $p$ . The pmf of  $X \sim \text{NBin}(r, p)$  is given by

$$f(x) = f(x; r, p) = \begin{cases} \binom{x-1}{r-1} p^r q^{x-r}; & x = r, r+1, r+2, \dots \\ 0; & \text{otherwise} \end{cases}$$

**Parameters of the Distribution:**  $r \in N^+$  (number of successes),  $0 \leq p \leq 1$  (probability of success).

#### Characteristics of Negative Binomial Distribution

1. The outcome of each trial is Bernoulli, that is either a success(S) or failure(F).
2. The Probability of success is constant  $P(S) = p$ .
3. The trials are repeated until 'r' successes occur.

#### Note

The geometric distribution is a special case of the negative binomial distribution when  $r = 1$ .

#### Moment Generating Function

If  $X$  is a discrete random variable has negative binomial distribution with parameters  $k, p$  then, the MGF of  $X$  is

$$M_X(t) = \left( \frac{pe^t}{1-qe^t} \right)^r.$$

#### Proof

**Hint:** The sum of a negative binomial series  $\sum_{k=0}^{\infty} \binom{k+r-1}{r-1} u^k = (1-u)^{-r}$ .

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \sum_{x=r}^{\infty} e^{tx} f(x) = \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r q^{x-r} \\ &= p^r e^{rt} \sum_{x=r}^{\infty} \binom{x-1}{r-1} e^{t(x-r)} q^{x-r} \end{aligned}$$

$$\begin{aligned}
&= p^r e^{rt} \sum_{x=r}^{\infty} \binom{x-1}{r-1} (qe^t)^{x-r} \quad (\text{Let } Y = X - r \Rightarrow X = Y + r) \\
&= p^r e^{rt} \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} (qe^t)^y = p^r e^{rt} (1 - qe^t)^{-r} = \left( \frac{pe^t}{1-qe^t} \right)^r.
\end{aligned}$$

### Mean and Variance

If  $X$  is a discrete random variable has negative binomial distribution with parameters  $r, p$  then,

$$E(X) = \mu = \frac{r}{p} \text{ and } V(x) = \sigma^2 = \frac{r(1-p)}{p^2} = \frac{rq}{p^2}.$$

### Proof

$$\text{III. } E(X) = \frac{r}{p}.$$

$$\begin{aligned}
E(X) &= \frac{d}{dt} M_X(t) \big|_{t=0} = \frac{d}{dt} \left( \frac{pe^t}{1-qe^t} \right)^r \big|_{t=0} \\
&= r \left( \frac{pe^t}{1-qe^t} \right)^{r-1} \left[ \frac{(1-qe^t)pe^t + pe^t qe^t}{(1-qe^t)^2} \right] \big|_{t=0} = r \frac{(pe^t)^r}{(1-qe^t)^{r+1}} \big|_{t=0} = r \frac{p^r}{(1-q)^{r+1}} = r \frac{p^r}{p^{r+1}} = \frac{r}{p}.
\end{aligned}$$

$$\text{IV. } V(X) = \frac{q}{p^2}.$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$\begin{aligned}
E(X^2) &= \frac{d^2}{dt^2} M_X(t) \big|_{t=0} = \frac{d}{dt} r \frac{(pe^t)^r}{(1-qe^t)^{r+1}} \big|_{t=0} = \frac{d}{dt} r (pe^t)^r (1-qe^t)^{-(r+1)} \big|_{t=0} \\
&= [r(pe^t)^r (-(r+1))(1-qe^t)^{-(r+2)}(-qe^t) + r^2(pe^t)^{r-1}(pe^t)(1-qe^t)^{-(r+1)}] \big|_{t=0} \\
&= [rp^r (-(r+1))(1-q)^{-(r+2)}(-q) + r^2 p^r (1-q)^{-(r+1)}] \\
&= [r(r+1)p^{-2}q + r^2 p^{-1}] = \frac{r^2}{p} \left[ \frac{q}{p} + 1 \right] + \frac{rq}{p^2} = \frac{r^2}{p^2} + \frac{rq}{p^2} = \frac{r^2+rq}{p^2} \quad (p+q=1)
\end{aligned}$$

Hence,

$$V(X) = \frac{r^2+rq}{p^2} - \left( \frac{r}{p} \right)^2 = \frac{rq}{p^2}.$$

### Example 3.6

Bob is a high school basketball player. He is a 70% free throw shooter. That means his probability of making a free throw is 0.70. During the season, what is the probability that Bob makes his third free throw on his fifth shot?. Find the MGF.

### Solution

Let  $X$  represents the no. of throws until the third free throw is done. The probability of successes (free throw) is  $p = 0.7$ . Thus,  $X \sim \text{NBin}(3, 0.7)$ . So, we want to find

$$P(X = 5) = f(5) = \binom{5-1}{3-1}(0.7)^3(0.3)^{5-3} = 0.1852.$$

$$M_X(t) = \left( \frac{0.7e^t}{1 - 0.3e^t} \right)^3$$

### Example 3.7

Sara flips a coin repeatedly and counts the number of heads (successes).. Find the probability that Sara gets

- I. The fourth head before the seventh flip.
- II. The first head on the fourth flip.

### Solution

The probability of successes (getting head) is  $p = 0.5$ . Let  $X$  represents the no. of throws until the fourth head is shown. Thus,  $X \sim NBin(4, 0.5)$ . So, we want to find

$$\begin{aligned} \text{I. } P(X < 6) &= f_X(4) + f_X(5) + f_X(6) = \binom{4-1}{4-1}(0.5)^4 + \binom{5-1}{4-1}(0.5)^5 + \binom{6-1}{4-1}(0.5)^6 \\ &= 0.0625 + 0.125 + 0.1563 = 0.3438. \end{aligned}$$

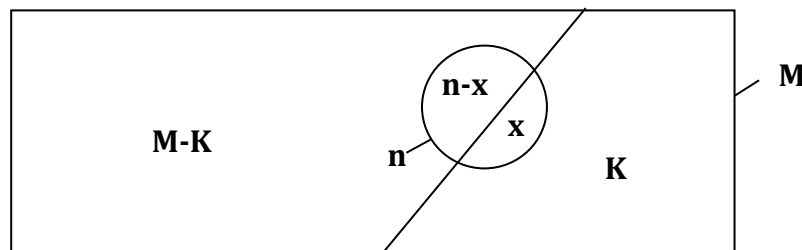
Now, let  $Y$  represents the no. of throws until the first head is shown. Thus,  $Y \sim Geom(0.5)$ .

$$\text{II. } P(Y = 4) = f_Y(4) = (0.5)^4 = 0.0625.$$

### Comparison

- For Bernoulli and binomial distributions the number of trials is fixed (1 for Ber. And  $n > 1$  for Bin.) while the number of successes is variable.
- For geometric and negative binomial distributions the number of trials is variable and the number of successes is fixed (1 for Geom.  $r > 1$  for NBin.).

## 3.6 Hypergeometric Distribution



In a group of  $M$  objects,  $K$  are of Type I and  $M - K$  are of Type II. If  $n$  objects are randomly chosen without replacement from the group of  $M$ , let  $X$  denote the

number that are of Type I in the group of  $n$ . Thus,  $X$  has a hypergeometric distribution  $X \sim H(M, n, K)$ . The pmf for  $X$  is

$$f(x) = f(x; M, n, K) = \begin{cases} \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}}; & x = \text{Max}[0, n - (M - K)], \dots, \text{Min}[n, K] \\ 0; & \text{otherwise} \end{cases}$$

**Parameters of the Distribution:**  $M \in N^+$  (population size),  $n \in N^+$  (sample size),  $K \in N^+$  (population elements with a certain characteristic).

### Characteristics of Hypergeometric Distribution

1. 'n' trials in a sample taken from a finite population of size M.
2. The population (outcome of trials) has two outcomes Success (S) and Failure (F).
3. Sample taken without replacement.
4. Trials are dependent.
5. The probability of success changes from trial to trial.

### Mean and Variance

If  $X$  is a discrete random variable has hypergeometric distribution with parameters  $M, n, K$  then,

$$E(X) = \mu = \frac{nK}{M} \text{ and } V(x) = \sigma^2 = \frac{nk(M-K)(M-n)}{M^2(M-1)}.$$

### Proof

We will assume that the bounds of  $X$  are 0 and  $n$ .

$$\text{I. } E(X) = \frac{nK}{M}.$$

$$\begin{aligned} E(X) &= \sum_{x=0}^n x f(x) = \sum_{x=0}^n x \cdot \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} \\ &= \sum_{x=1}^n x \cdot \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} \quad (\text{Set summation from 1, since when } x = 0 \text{ the expression} = 0) \\ &= \sum_{x=1}^n x \cdot \frac{\frac{K!}{x!(K-x)!} \binom{M-K}{n-x}}{\frac{M!}{n!(M-n)!}} = \frac{nK}{M} \sum_{x=1}^n \frac{\frac{(K-1)!}{(x-1)!(K-x)!} \binom{M-K}{n-x}}{\frac{(M-1)!}{(n-1)!(M-n)!}} \\ &= \frac{nK}{M} \sum_{x=1}^n \frac{\binom{K-1}{x-1} \binom{M-K}{n-x}}{\binom{M-1}{n-1}} \quad (\text{Let } Y = X - 1, L = M - 1, S = K - 1 \text{ and } r = n - 1) \\ &= \frac{nK}{M} \sum_{y=0}^r \frac{\binom{S}{y} \binom{L-S}{r-y}}{\binom{L}{r}} = \frac{nK}{M} (1) = \frac{nK}{M} \quad (Y \sim H(L, r, S) \Rightarrow \text{Sum of pmf} = 1) \end{aligned}$$

$$\text{II. } V(X) = \frac{nk(M-K)(M-n)}{M^2(M-1)}.$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$\begin{aligned} E(X^2) &= \sum_{x=0}^n x^2 f(x) = \sum_{x=0}^n x^2 \cdot \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} \\ &= \sum_{x=1}^n x^2 \cdot \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} \quad (\text{Set summation from 1, since when } x = 0 \text{ the expression} = 0) \\ &= \sum_{x=1}^n x^2 \cdot \frac{\frac{K!}{x!(K-x)!} \binom{M-K}{n-x}}{\frac{M!}{n!(M-n)!}} = \frac{nK}{M} \sum_{x=1}^n x \cdot \frac{\frac{(K-1)!}{(x-1)!(K-x)!} \binom{M-K}{n-x}}{\frac{(M-1)!}{(n-1)!(M-n)!}} = \frac{nK}{M} \sum_{x=1}^n x \cdot \frac{\binom{K-1}{x-1} \binom{M-K}{n-x}}{\binom{M-1}{n-1}} \end{aligned}$$

We use the same variable substitution as when deriving the mean.

$$E(X^2) = \frac{nK}{M} \sum_{y=0}^r (y+1) \cdot \frac{\binom{S}{y} \binom{L-S}{r-y}}{\binom{L}{r}} = \frac{nK}{M} \left[ \sum_{y=0}^r y \cdot \frac{\binom{S}{y} \binom{L-S}{r-y}}{\binom{L}{r}} + \sum_{y=0}^r \frac{\binom{S}{y} \binom{L-S}{r-y}}{\binom{L}{r}} \right]$$

The first sum is the expected value of a hypergeometric random variable with parameters  $(L, r, S)$ . The second sum is the total sum that random variable's pmf.

$$E(X^2) = \frac{nK}{M} [E(Y) + 1] = \frac{nK}{M} \left[ \frac{rS}{L} + 1 \right] = \frac{nK}{M} \left[ \frac{(n-1)(K-1)}{(M-1)} + 1 \right]; \text{ Thus,}$$

$$\begin{aligned} V(X) &= \frac{nK}{M} \left[ \frac{(n-1)(K-1)}{(M-1)} + 1 \right] - \left( \frac{nK}{M} \right)^2 = \frac{nK}{M} \left[ \frac{(n-1)(K-1)}{(M-1)} + 1 - \frac{nK}{M} \right] \\ &= \frac{nK}{M} \left[ \frac{M(n-1)(K-1) + M(M-1) - nk(M-1)}{M(M-1)} \right] \\ &= \frac{nK}{M^2(M-1)} [MnK - Mn - MK + M + M^2 - M - nkM + nk] \\ &= \frac{nK}{M^2(M-1)} [-Mn - MK + M^2 + nk] = \frac{nK}{M^2(M-1)} [M(M-n) - K(M-n)] \\ &= \frac{nK(M-n)(M-K)}{M^2(M-1)}. \end{aligned}$$

### Example 3.8

Lots of 40 components each are called acceptable if they contain no more than 3 defectives. The procedure for sampling the lot is to select 5 components at random (without replacement) and to reject the lot if a defective is found. What is the probability that exactly one defective is found in the sample if there are 3 defectives in the entire lot.

### Solution

$$M = 40, n = 5, K = 3.$$

Let  $X$  represents the no. of defective items in the sample.  $X \sim H(40, 5, 3)$ . We want to find

$$P(X = 1) = f(1) = \frac{{}^3_1({}^{37}_4)}{{}^{40}_5} = 0.3011.$$

Does the procedure being used is good?

### 3.6.1 Binomial Approximation to Hypergeometric Distribution

Suppose we still have the population of size  $M$  with  $K$  units labeled as 'success' and  $M - K$  labeled as 'failure,' but now we take a sample of size  $n$  is drawn *with replacement*. Then, with each draw, the units remaining to be drawn look the same: still  $K$  'successes' and  $M - K$  'failures.' Thus, the probability of drawing a 'success' on each single draw is  $p = \frac{K}{M}$ , and this doesn't change. When we were drawing without replacement, the proportions of successes would change, depending on the result of previous draws. For example, if we were to obtain a 'success' on the first draw, then the proportion of 'successes' for the second draw would be  $\frac{K-1}{M-1}$ , whereas if we were to obtain a 'failure' on the first draw the proportion of successes for the second draw would be  $\frac{K}{M-1}$ .

#### Proposition

If the population size  $M \rightarrow \infty$  in such a way that the proportion of successes  $\frac{K}{M} \rightarrow p$ , and  $n$  is held constant, then the hypergeometric probability mass function approaches the binomial probability mass function i.e.  $H(M, n, K) \rightarrow \text{Bin}\left(n, p = \frac{K}{M}\right)$ .

As a rule of thumb, if the population size is more than 20 times the sample size ( $M > 20n$ ), then we may use binomial probabilities in place of hypergeometric probabilities.

#### Example 3.9

A box contains 6 blue and 4 red balls. An experiment is performed a ball is chosen and its color observed. Find the probability, that after 5 trials, 3 blue balls will have been chosen when

- I. The balls are replaced.
- II. The balls not replaced.



**Solution**

I. Let  $X$  represents the no. of blue balls in the sample.  $X \sim \text{Bin}(5, 0.6)$ . So, we want to find

$$P(X = 3) = \binom{5}{3}(0.6)^3(0.4)^2 = 0.3456.$$

II. Let  $Y$  represents the no. of blue balls in the sample.  $Y \sim H(10, 5, 6)$ . So, we want to find

$$P(Y = 3) = \frac{\binom{6}{3}\binom{4}{2}}{\binom{10}{5}} = 0.4762.$$

**Example 3.10**

It is estimated that 4000 of the 10,000 voting residents of a town are against a new sales tax. If 15 eligible voters are selected at random and asked their opinion, what is the probability that at most 3 favor the new tax? Use binomial approximation.

**Solution**

$M = 10000, n = 15, K = 4000$ . To use the binomial approximation we have to check if  $M > 20n$ ?

$$M = 10000 > 20 \cdot 15 = 300.$$

Thus,  $X$  the no. of voting that favor the new sales tax in the sample has binomial distribution with parameters  $n = 15, p = \frac{K}{M} = 0.4$ .

$$\begin{aligned} P(X \leq 3) &= f(0) + f(1) + f(2) + f(3) \\ &= \binom{15}{0}(0.4)^0(0.6)^{15} + \binom{15}{1}(0.4)(0.6)^{14} + \binom{15}{2}(0.4)^2(0.6)^{13} + \binom{15}{3}(0.4)^3(0.6)^{12} \\ &= 0.0019. \end{aligned}$$

**3.7 Poisson Distribution**

The Poisson distribution is often used as a model for counting the number of events of a certain type that occur in a certain period of time (or space). If the r.v.  $X$  has Poisson distribution  $X \sim \text{Poisson}(\lambda)$  then its pmf is given by

$$f(x) = f(x; \lambda) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!}; & x = 0, 1, 2, \dots \\ 0; & \text{otherwise} \end{cases}$$

**Parameter of the Distribution:**  $\lambda > 0$  (The average)

**For example**

- The number of births per hour during a given day.
- The number of failures of a machine in one month.
- The number of typing errors on a page.
- The number of postponed baseball games due to rain.

**Note**

Suppose that  $X$  represents the number of customers arriving for service at bank in a one hour period, and that a model for  $X$  is the Poisson distribution with parameter  $\lambda$ . Under some reasonable assumptions (such as independence of the numbers arriving in different time intervals) it is possible to show that the number arriving in any time period also has a Poisson distribution with the appropriate parameter that is "scaled" from  $\lambda$ . Suppose that  $\lambda = 40$  'meaning that  $X$ , the number of bank customers arriving in one hour, has a mean of 40'. If  $Y$  represents the number of customers arriving in 2 hours, then  $Y$  has a Poisson distribution with a parameter of 80. In general, for any time interval of length  $t$ , the number of customers arriving in that time interval has a Poisson distribution with parameter  $\lambda t = 40t$ . So, the number of customers arriving during a 15-minute period ( $t = \frac{1}{4}$  hour) will have a Poisson distribution with parameter  $40 \cdot \frac{1}{4} = 10$ . In general, If  $W$  represents the number of customers arriving in  $t$  hours  $W \sim \text{Poisson}(\lambda t)$  therefore,

$$f(w) = \frac{e^{-\lambda t} (\lambda t)^w}{w!}; \quad w = 0, 1, 2, \dots$$

**Mean and Variance**

If  $X$  is a discrete random variable has Poisson distribution with parameter  $\lambda$  then,

$$E(X) = V(x) = \lambda.$$

**Proof**

**Hint:**  $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$

I.  $E(X) = \lambda.$

$$E(X) = \sum_{x=0}^{\infty} x f(x)$$

$$= \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!} \quad (\text{Set summation from 1, since when } x = 0 \text{ the expression} = 0)$$

$$\begin{aligned}
 &= \sum_{x=1}^{\infty} \frac{e^{-\lambda} \lambda^x}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \quad (\text{Let } Y = X - 1) \\
 &= \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.
 \end{aligned}$$

II.  $V(X) = \lambda$ .

$$V(X) = E(X^2) - [E(X)]^2$$

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

$$= E[X(X-1)] + \lambda = \sum_{x=0}^{\infty} x(x-1)f(x) + \lambda$$

$$= \sum_{x=2}^{\infty} x(x-1) \frac{e^{-\lambda} \lambda^x}{x!} + \lambda \quad (\text{Set summation from 2, since when } x = 0, 1 \text{ the expression} = 0)$$

$$= \sum_{x=2}^{\infty} \frac{e^{-\lambda} \lambda^x}{(x-2)!} + \lambda = \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda \quad (\text{Let } Z = X - 2)$$

$$= \lambda^2 e^{-\lambda} \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} + \lambda = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda = \lambda^2 + \lambda.$$

Hence,

$$V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

### Moment Generating Function

If  $X$  is a discrete random variable has Poisson distribution with parameter  $\lambda$  then, the MGF of  $X$  is

$$M_X(t) = e^{\lambda(e^t - 1)}.$$

### Proof

**Hint:**  $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$

$$\begin{aligned}
 M_X(t) &= E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} f(x) = \sum_{x=0}^{\infty} e^{tx} \cdot \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} = e^{-\lambda} e^{e^t \lambda} \\
 &= e^{\lambda(e^t - 1)}.
 \end{aligned}$$

### Example 3.11

Suppose that the number of typing errors per page has a Poisson distribution with average 6 typing errors. What is the probability that

- I. the number of typing errors in a page will be 7.
- II. the number of typing errors in a page will be at least 2.
- III. in 2 pages there will be 10 typing errors.
- IV. in a half page there will be no typing errors.

**Solution**

I. Let  $X$  represents the no. of typing errors per page.

Therefore,  $\lambda_X = 6 \Rightarrow X \sim \text{Poisson}(6)$ .

$$P(X = 7) = \frac{e^{-6}6^7}{7!} = 0.1377.$$

$$\begin{aligned} \text{II. } P(X \geq 2) &= f(2) + f(3) + \dots = 1 - P(X < 2) = 1 - f(0) - f(1) \\ &= 1 - \frac{e^{-6}6^0}{0!} - \frac{e^{-6}6}{1!} = 0.9826. \end{aligned}$$

III. Let  $Y$  represents the no. of typing errors in 2 pages.

Therefore,  $\lambda_Y = \lambda_X t = 6 \cdot 2 = 12 \Rightarrow Y \sim \text{Poisson}(12)$ .

$$P(Y = 10) = \frac{e^{-12}(12)^{10}}{10!} = 0.1048.$$

IV. Let  $Z$  represents the no. of typing errors in a half pages.

Therefore,  $\lambda_Z = \lambda_X t = 6 \cdot \frac{1}{2} = 3 \Rightarrow Z \sim \text{Poisson}(3)$ .

$$P(Z = 0) = \frac{e^{-3}3^0}{0!} = 0.0498.$$

**3.7.1 Poisson Approximation to Binomial Distribution**

For those situations in which  $n$  is large ( $\geq 100$ ) and  $p$  is very small ( $\leq 0.1$ ), the Poisson distribution can be used to approximate the binomial distribution. The larger the  $n$  and the smaller the  $p$ , the better is the approximation. The following mathematical expression for the Poisson model is used to approximate the true (binomial) result:

$$f(x) = \frac{e^{-(np)}(np)^x}{x!}$$

Where  $n$  is the sample size and  $p$  is the true probability of success (i.e.  $\lambda = np$ ).

**Example 3.12**

Given that 5% of a population are left-handed, use the Poisson distribution to estimate the probability that a random sample of 100 people contains 2 or more left-handed people, then compare the result with the true probability using the binomial distribution.

**Solution**

Let  $X$  represents the no. of left-handed on the sample.

To use Poisson approximation we should check if  $n \geq 100$  and  $p \leq 0.1$ .

Since  $n = 100 \geq 100$ ,  $p = 0.05 \leq 0.1$  we can use Poisson approximation.

$$\lambda = np = 100 \cdot 0.05 = 5 \Rightarrow X \sim \text{Poisson}(5).$$

Thus,

$$P(X \geq 2) = 1 - P(X < 2) = 1 - f(0) - f(1) = 1 - \frac{e^{-5}(5)^0}{0!} - \frac{e^{-5}(5)}{1!} = 0.9596 \cong 0.96.$$

Now, let us use binomial distribution.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - f(0) - f(1) \\ &= 1 - \binom{100}{0}(0.05)^0(0.95)^{100} - \binom{100}{1}(0.05)^1(0.95)^{99} = 0.9629 \cong 0.96. \end{aligned}$$

## Chapter Four

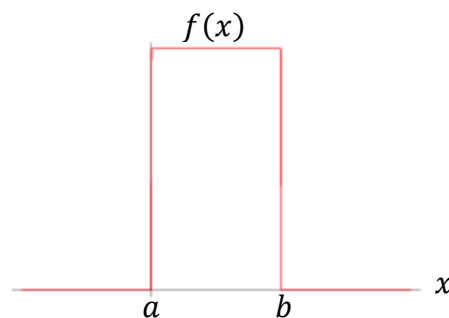
### Frequently Used Continuous Probability Distributions

#### Distributions to be Covered

- Uniform distribution.
- Exponential distribution.
- Gamma distribution.
- Chi-squared distribution.
- Beta distribution.
- Normal distribution.
- Standard normal distribution.
- T distribution.

#### 4.1 Uniform Distribution

A uniform distribution, sometimes also known as a rectangular distribution, is a distribution that has constant probability.



The probability density function for a continuous uniform distribution on the interval  $[a, b]$  is

$$f(x) = f(x; a, b) = \begin{cases} \frac{1}{b-a}; & a \leq x \leq b \\ 0; & \text{otherwise} \end{cases}$$

We write  $X \sim U(a, b)$

**Parameters of the Distribution:**  $a, b \in \mathcal{R}$  (The limits of the interval)

### Mean and Variance

If  $X$  is a continuous random variable has uniform distribution with parameters  $a$  and  $b$  then,

$$E(X) = \frac{a+b}{2} \quad \text{and} \quad V(X) = \frac{(b-a)^2}{12}.$$

#### Proof

##### Hint

$$b^2 - a^2 = (b-a)(b+a),$$

$$b^3 - a^3 = (b-a)(b^2 + ab + a^2) \text{ and}$$

$$(b-a)^2 = b^2 - 2ab + a^2.$$

$$\text{I. } E(X) = \frac{a+b}{2}.$$

$$E(X) = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[ \frac{x^2}{2} \right] \Big|_a^b = \frac{1}{b-a} \left[ \frac{b^2-a^2}{2} \right] = \frac{1}{b-a} \left[ \frac{(b-a)(b+a)}{2} \right] = \frac{a+b}{2}.$$

$$\text{II. } V(X) = \frac{(b-a)^2}{12}.$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$\begin{aligned} E(X^2) &= \int_a^b x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \left[ \frac{x^3}{3} \right] \Big|_a^b = \frac{1}{b-a} \left[ \frac{b^3-a^3}{3} \right] \\ &= \frac{1}{b-a} \left[ \frac{(b-a)(b^2+ab+a^2)}{3} \right] = \frac{b^2+ab+a^2}{3}. \end{aligned}$$

Hence,

$$\begin{aligned} V(X) &= \frac{b^2+ab+a^2}{3} - \left( \frac{a+b}{2} \right)^2 = \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4} = \frac{4b^2+4ab+4a^2-3b^2-6ab-3a^2}{12} \\ &= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

### Moment Generating Function

If  $X$  is a continuous random variable has uniform distribution with parameters  $a$  and  $b$  then, the MGF of  $X$  is

$$M_X(t) = \frac{e^{bt}-e^{at}}{t(b-a)}.$$

#### Proof

$$M_X(t) = E(e^{tX}) = \int_a^b e^{tx} f(x) dx = \int_a^b \frac{e^{tx}}{b-a} dx = \frac{e^{tx}}{t(b-a)} \Big|_a^b = \frac{e^{bt}-e^{at}}{t(b-a)}.$$

Note that the above derivation is valid only when  $t \neq 0$ . However, remember that it always when  $t = 0$ ,  $M_X(t) = E(e^{x(0)}) = E(1) = 1$ .

**Example 4.1**

The daily amount of coffee, in liters, dispensed by a machine located in an airport lobby is a random variable  $X$  having a continuous uniform distribution with  $a = 7$  and  $b = 10$ . Find the probability that on a given day the amount of coffee dispensed by this machine will be

- I. at most 8.8 liters.
- II. more than 7.4 liters but less than 9.5 liters.
- III. at least 12.5 liters.

**Solution**

$$f(x) = \frac{1}{10-7} = \frac{1}{3}.$$

$$\text{I. } P(X \leq 8.8) = \int_7^{8.8} \frac{1}{3} dx = \frac{x}{3} \Big|_7^{8.8} = \frac{8.8-7}{3} = 0.6.$$

$$\text{II. } P(7.4 \leq X \leq 9.5) = \frac{9.5-7.4}{3} = 0.7.$$

$$\text{III. } P(X \geq 12.5) = 0.$$

**Example 4.2**

A bus arrives every 10 minutes at a bus stop. It is assumed that the waiting time for a particular individual is a r.v. with a continuous uniform distribution. What is the probability that the individual waits more than 7 minutes. Find the mean and the standard deviation.

**Solution**

Let  $X$  is the waiting time for the individual. Thus,  $X \sim U(0,10)$ .

$$P(X > 7) = \frac{10-7}{10} = 0.3.$$

$$\mu = \frac{0+10}{2} = 5 \quad \text{and} \quad \sigma = \sqrt{\frac{(10-0)^2}{12}} = 2.89.$$

**4.2 Exponential Distribution**

A continuous random variable  $X$  is said to have an exponential distribution  $X \sim \text{Exp}(\theta)$  if it has probability density function

$$f(x) = f(x; \theta) = \begin{cases} \theta e^{-\theta x}; & x \geq 0 \\ 0; & \text{otherwise} \end{cases},$$

where  $\theta > 0$ .



The exponential distribution is usually used to model the time until something happens in the process.

**For example**

- The exponential random variable is used to measure the waiting time for elevator to come.
- The time it takes to load a truck.
- The waiting time at a car wash.

Another form of exponential distribution is  $f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$ ;  $x \geq 0$ . However, for the rest of this course, we will use the first form.

**Parameter of the Distribution:**  $\theta > 0$ .

**Cumulative Distribution Function**

$$F(x) = P(X \leq x) = \int_0^x \theta e^{-\theta x} dx = -e^{-\theta x} \Big|_0^x = 1 - e^{-\theta x}.$$

**Direct way to find probabilities**

- I.  $P(X \leq a) = F(a) = 1 - e^{-\theta a}$ .
- II.  $P(a \leq X \leq b) = F(b) - F(a) = e^{-\theta a} - e^{-\theta b}$ .
- III.  $P(X \geq b) = e^{-\theta b}$ .

**Mean and Variance**

If  $X$  is a continuous random variable has exponential distribution with parameter  $\theta$  then,

$$E(X) = \frac{1}{\theta} \quad \text{and} \quad V(x) = \frac{1}{\theta^2}.$$

**Proof**

$$\text{I. } E(X) = \frac{1}{\theta}.$$

$$E(X) = \int_0^{\infty} x f(x) dx = \int_0^{\infty} \theta x e^{-\theta x} dx = \theta \int_0^{\infty} x e^{-\theta x} dx.$$

Use integration by parts:

$$\begin{aligned} u &= x & dv &= e^{-\theta x} dx \\ du &= dx & v &= -\frac{e^{-\theta x}}{\theta} \end{aligned}$$

$$\begin{aligned} E(X) &= \theta \int_0^{\infty} x e^{-\theta x} dx = \theta \left[ -\frac{x e^{-\theta x}}{\theta} \Big|_0^{\infty} + \int_0^{\infty} \frac{e^{-\theta x}}{\theta} dx \right] = 0 + \int_0^{\infty} e^{-\theta x} dx = -\frac{e^{-\theta x}}{\theta} \Big|_0^{\infty} \\ &= \frac{1}{\theta}. \end{aligned}$$

$$\text{II. } V(X) = \frac{1}{\theta^2}.$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_0^\infty x^2 f(x) dx = \int_0^\infty \theta x^2 e^{-\theta x} dx = \theta \int_0^\infty x^2 e^{-\theta x} dx.$$

Use integration by parts:

$$\begin{aligned} u &= x^2 & dv &= e^{-\theta x} dx \\ du &= 2x dx & v &= -\frac{e^{-\theta x}}{\theta} \end{aligned}$$

$$\begin{aligned} E(X^2) &= \theta \int_0^\infty x^2 e^{-\theta x} dx = \theta \left[ -\frac{x^2 e^{-\theta x}}{\theta} \Big|_0^\infty + 2 \int_0^\infty \frac{x e^{-\theta x}}{\theta} dx \right] = 0 + 2 \int_0^\infty x e^{-\theta x} dx \\ &= 2 \left[ -\frac{x e^{-\theta x}}{\theta} \Big|_0^\infty + \int_0^\infty \frac{e^{-\theta x}}{\theta} dx \right] = 2 \int_0^\infty \frac{e^{-\theta x}}{\theta} dx = -2 \frac{e^{-\theta x}}{\theta^2} \Big|_0^\infty = \frac{2}{\theta^2}. \end{aligned}$$

Hence,

$$V(X) = \frac{2}{\theta^2} - \left(\frac{1}{\theta}\right)^2 = \frac{1}{\theta^2}.$$

### Moment Generating Function

If  $X$  is a continuous random variable has exponential distribution with parameter  $\theta$  then, the MGF of  $X$  is

$$M_X(t) = \frac{\theta}{\theta - t}.$$

#### Proof

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \int_0^\infty e^{tx} f(x) dx = \int_0^\infty e^{tx} \cdot \theta e^{-\theta x} dx = \theta \int_0^\infty e^{-(\theta-t)x} dx \\ &= -\frac{\theta}{\theta-t} e^{-(\theta-t)x} \Big|_0^\infty = \frac{\theta}{\theta-t}. \end{aligned}$$

Note that the above derivation is valid only when  $t < \theta$ .

### Example 4.3

The time between arrivals of cars at Al's full-service gas pump follows an exponential distribution with a mean time between arrivals of 3 minutes. Al would like to know the probability that the time between two successive arrivals will be 2 minutes or less. Then find the variance.

#### Solution

Let  $X$  represents the time between two successive arrivals.

$$\theta = \frac{1}{3} \Rightarrow X \sim \text{Exp}\left(\frac{1}{3}\right).$$

$$P(X \leq 2) = F(2) = 1 - e^{-\frac{2}{3}} = 0.4866.$$

$$V(X) = \frac{1}{\theta^2} = 3^2 = 9.$$

#### 4.2.1 Lack of Memory Property (Memoryless)

Let  $X$  be exponentially distributed with parameter  $\theta$ . Suppose we know  $X > t$ . What is the probability that  $X$  is also greater than some value  $s + t$ ? That is, we want to know  $P(X > s + t | X > t)$ . This type of problem shows up frequently when we are interested in the time between events; (Such as, queuing system). For example, suppose that costumers in a particular system have exponentially distributed service times. If we have a costumer that's been waiting for one minute, what's the probability that it will continue to wait for more than two minutes?

Using the definition of conditional probability, we have

$$P(X > s + t | X > t) = \frac{P(X > s + t \cap X > t)}{P(X > t)}$$

If  $X > s + t$ , then  $X > t$  is redundant, so we can simplify the numerator as

$$P(X > s + t | X > t) = \frac{P(X > s + t)}{P(X > t)}$$

Using the CDF of the exponential distribution,

$$P(X > s + t | X > t) = \frac{P(X > s + t)}{P(X > t)} = \frac{e^{-\theta(s+t)}}{e^{-\theta t}} = e^{-\theta s}$$

It turns out that the conditional probability does not depend on  $t$ . Thus, In our queue example, the probability that a costumer waits for one additional minute is the same as the probability that it wait for one minute originally, regardless of how long it's been waiting.

**This is called the lack of memory property.**

$$P(X > s + t | X > t) = P(X > s).$$

#### Example 4.4

On average, it takes about 5 minutes to get an elevator at stat building. Let  $X$  be the waiting time until the elevator arrives. Find the pdf of  $X$  then calculate the probability that

- I. you will wait less than 3 minutes?
- II. you will wait for more than 10 minutes?

III. you will wait for more than 7 minutes?

IV. you will wait for more than 10 minutes given that you already wait for more than 3 minutes?

**Solution**

$$X \sim \text{Exp}\left(\frac{1}{5}\right)$$

$$f(x) = \frac{1}{5} e^{-\frac{x}{5}}; \quad x \geq 0.$$

$$\text{I. } P(X < 3) = 1 - e^{-\frac{3}{5}} = 0.4512.$$

$$\text{II. } P(X > 10) = e^{-2} = 0.1353.$$

$$\text{III. } P(X > 7) = e^{-\frac{7}{5}} = 0.2466.$$

$$\text{IV. } P(X > 10 | X > 3) = P(X > 7 + 3 | X > 3) = P(X > 7) = 0.2466.$$

#### 4.2.2 Relation Between Exponential and Poisson Distributions

An interesting feature of these two distributions is that, if the Poisson provides an appropriate description of the number of occurrences per interval of time, then the exponential will provide a description of the length of time between occurrences.

Consider the probability function for the Poisson distribution,

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

where,  $\lambda$  is the mean rate of arrivals and  $t$  is a period of time.

Defining the r.v.  $Y$  as the time of an event, we have (by definition),

$$F(t) = P(T \leq t) = 1 - P(T > t).$$

Now, The 1<sup>st</sup> arrival occurs after time  $t$  iff there are no arrivals in the interval  $[0, t]$ .

Therefore,

$$P(T > t) = P(\text{zero events occur in time } 0 \text{ to } t) = P(X = 0) = \frac{e^{-(\lambda t)} (\lambda t)^0}{0!} = e^{-\lambda t}.$$

Hence,

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

which is the distribution function for the exponential distribution with parameter  $\lambda$ .

Briefly,

$X$  (no. of occurrences per interval of time  $t$ )  $\sim \text{Poisson}(\lambda t)$ ,

$Y$  (time between occurrences)  $\sim \text{Exp}(\lambda)$ .

**Example 4.5**

If we know that there are on average 10 customers visiting a store within 2 hours (120 minutes) interval, then the r.v. that represents the number of costumers is  $X \sim \text{Poisson}(\lambda t = 10)$ . From another point of view, the r.v. that represents the time between costumers arrivals  $Y \sim \text{Exp}\left(\lambda = \frac{10}{120}\right)$ , where the average time between customers' arrival is  $\frac{1}{\lambda} = \frac{120}{10} = 12$  minutes.

**4.3 Gamma Distribution**

The gamma distribution is another widely used distribution. Its importance is largely due to its relation to exponential and normal distributions. Before introducing the gamma random variable, we need to introduce the gamma function.

**4.3.1 Gamma function**

The gamma function denoted by  $\Gamma(\alpha)$ , is an extension of the factorial function to real (and complex) numbers. Specifically, if  $n \in \{1, 2, 3, \dots\}$ , then

$$\Gamma(n) = (n - 1)!$$

More generally, for any positive real number  $\alpha$ ,  $\Gamma(\alpha)$  is defined as

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx; \quad \alpha > 0.$$

**4.3.2 Some useful Properties for Gamma distribution**

$$\text{I. } \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

$$\text{II. } \Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \alpha > 0.$$

**Proof**

$$\Gamma(\alpha + 1) = \int_0^{\infty} x^{\alpha} e^{-x} dx \quad (\text{from gamma function definition}).$$

Use integration by parts:

$$u = x^{\alpha} \quad dv = e^{-x} dx$$

$$du = \alpha x^{\alpha-1} dx \quad v = -e^{-x}$$

$$\Gamma(\alpha + 1) = -x^{\alpha} e^{-x} \Big|_0^{\infty} + \int_0^{\infty} \alpha x^{\alpha-1} e^{-x} dx$$

$$= \alpha \int_0^{\infty} x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha). \quad (\text{from gamma function definition})$$

$$\text{III. } \int_0^{\infty} x^{\alpha} e^{-\beta x} dx = \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}}; \quad \alpha, \beta > 0.$$

**Proof**

Let  $y = \beta x \Rightarrow dy = \beta dx \Rightarrow dx = \frac{dy}{\beta}$ , thus,  $x: 0 \rightarrow \infty \Rightarrow y: 0 \rightarrow \infty$  then,

$$\begin{aligned} \int_0^{\infty} x^{\alpha} e^{-\beta x} dx &= \int_0^{\infty} \left(\frac{y}{\beta}\right)^{\alpha} e^{-y} \frac{dy}{\beta} = \frac{1}{\beta^{\alpha+1}} \int_0^{\infty} y^{\alpha} e^{-y} dy \\ &= \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \quad (\text{from gamma function definition}) \end{aligned}$$

**Example 4.6**

I. Find  $\Gamma\left(\frac{7}{2}\right)$ .

II. Find the value of the following integral  $I = \int_0^{\infty} x^6 e^{-5x} dx$ .

**Solution**

I.  $\Gamma\left(\frac{7}{2}\right) = \frac{5}{2} \cdot \Gamma\left(\frac{5}{2}\right) = \frac{5}{2} \cdot \frac{3}{2} \Gamma\left(\frac{3}{2}\right) = \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi} = \frac{15}{8} \sqrt{\pi}.$

II.  $I = \int_0^{\infty} x^6 e^{-5x} dx = \frac{\Gamma(6+1)}{5^{6+1}} = \frac{6!}{5^7} = 0.0092.$

**4.3.3 Definition of Gamma Distribution**

We now define the gamma distribution by providing its pdf.

A continuous random variable  $X$  is said to have a gamma distribution with parameters  $\alpha > 0$  and  $\beta > 0$ , denoted by  $X \sim \text{Gamma}(\alpha, \beta)$ , if its pdf is given by

$$f_X(x) = f(x; \alpha, \beta) = \begin{cases} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}; & x \geq 0 \\ 0; & \text{otherwise} \end{cases}$$

**Parameters of the Distribution:**  $\alpha > 0, \beta > 0$ .

**Mean and Variance**

If  $X$  is a continuous random variable has gamma distribution with parameters  $\alpha, \beta$  then,

$$E(X) = \frac{\alpha}{\beta} \quad \text{and} \quad V(x) = \frac{\alpha}{\beta^2}.$$

**Proof**

III.  $E(X) = \frac{\alpha}{\beta}.$

$$\begin{aligned} E(X) &= \int_0^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha} e^{-\beta x} dx \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \quad (\text{Using property III}) \end{aligned}$$

$$= \frac{1}{\Gamma(\alpha)} \cdot \frac{\alpha \Gamma(\alpha)}{\beta} = \frac{\alpha}{\beta}. \quad (\text{Using property II})$$

$$\text{IV. } V(X) = \frac{\alpha}{\beta^2}.$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_0^\infty x^2 f(x) dx = \int_0^\infty x^2 \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-\beta x} dx$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+2)}{\beta^{\alpha+2}} \quad (\text{Using property III})$$

$$= \frac{1}{\Gamma(\alpha)} \cdot \frac{(\alpha+1)\alpha \Gamma(\alpha)}{\beta^2} = \frac{\alpha(\alpha+1)}{\beta^2}. \quad (\text{Using property II})$$

Hence,

$$V(X) = \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2}.$$

### Moment Generating Function

If  $X$  is a continuous random variable has gamma distribution with parameters  $\alpha, \beta$  then, the MGF of  $X$  is

$$M_X(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha = \left(1 - \frac{t}{\beta}\right)^{-\alpha}.$$

#### Proof

$$\begin{aligned} M_X(t) &= E(e^{tx}) = \int_0^\infty e^{tx} f(x) dx = \int_0^\infty e^{tx} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-t)x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{(\beta-t)^\alpha} \quad (\text{Using property III}) \\ &= \frac{\beta^\alpha}{(\beta-t)^\alpha} = \left(\frac{\beta}{\beta-t}\right)^\alpha. \end{aligned}$$

Note that the above derivation is valid only when  $t < \beta$ .

### 4.3.4 Special Cases

#### First Case

If we let  $\alpha = 1$ , we obtain  $f(x; 1, \beta) = \beta e^{-\beta x}$ ;  $x \geq 0$ . Thus, we conclude that

$$\text{Gamma}(1, \beta) = \text{Exp}(\beta).$$

**Second Case**

If we let  $\alpha = \frac{v}{2}, \beta = \frac{1}{2}$ , we obtain  $f\left(x; \frac{v}{2}, \frac{1}{2}\right) = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} x^{\frac{v}{2}-1} e^{-\frac{x}{2}}; x \geq 0$ , which is the pdf of the Chi-squared distribution with  $v$  degrees of freedom. (This distribution will be discussed in section 4.4).

**Third Case**

If we let  $\beta = 1$ , we obtain  $f(x; \alpha, 1) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}; x \geq 0$ , which is called the standard gamma distribution with parameter  $\alpha$ .

**4.3.5 Incomplete Gamma Function**

When  $X$  follows the standard Gamma distribution then its cdf is

$$F^*(x; \alpha) = \int_0^x \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} dx; \quad x \geq 0.$$

This is also called the incomplete gamma function.

**Proposition**

If  $X \sim \text{Gamma}(\alpha, \beta)$ , then

$$F(x; \alpha, \beta) = P(X \leq x) = F^*(\beta x; \alpha),$$

where  $F^*$  is the incomplete gamma function, and  $F$  is the cdf of the gamma distribution.

**Note**

Table 4.1 in appendix (A) provides some values of  $F^*(x; \alpha)$  for  $\alpha = 1, 2, \dots, 10$  and  $x = 1, 2, \dots, 15$ .

**Example 4.7**

Let  $X$  represents the survival time in weeks, where  $X \sim \text{Gamma}(6, 0.05)$ . Find the mean and the variance, then calculate the probabilities  $P(60 < X < 120)$ ,  $P(X < 30)$ .

**Solution**

$$\mu = \frac{\alpha}{\beta} = \frac{6}{0.05} = 120 \text{ weeks, and } \sigma^2 = \frac{\alpha}{\beta^2} = \frac{6}{(0.05)^2} = 2400.$$

$$P(60 < X < 120) = P(X < 120) - P(X < 60) = F(120; 6, 0.05) - F(60; 6, 0.05)$$

$$F^*(120 \cdot 0.05; 6) - F^*(60 \cdot 0.05; 6) = F^*(6; 6) - F^*(3; 6) = 0.554 - 0.084 = 0.47.$$



**Example 4.8**

Suppose that the time, in hours, taken to repair a heat pump is a r.v.  $X$  having a gamma distribution with parameters  $\alpha = \beta = 2$ . What is the probability that the next service call will required

- I. at least 2 hours to repair the heat pump.
- II. at most 1.25 hours to repair the heat pump.

**Solution**

$$f(x) = \frac{2^2}{\Gamma(2)} x^{2-1} e^{-2x} = 4x e^{-2x}.$$

$$\text{I. } P(X \geq 2) = 1 - P(X < 2) = 1 - F(2; 2, 2) = 1 - F^*(4; 2) = 1 - 0.908 = 0.092.$$

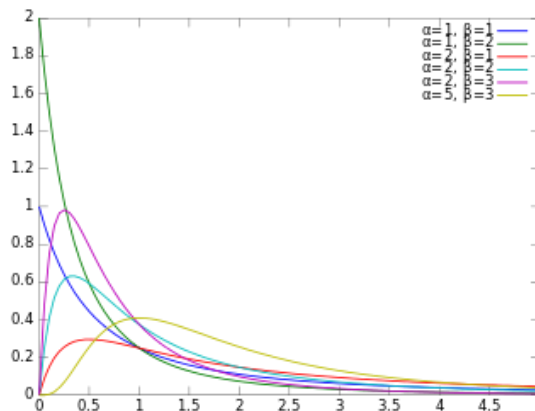
$$\text{II. } P(X \leq 1.25) = \int_0^{1.25} 4x e^{-2x} dx$$

Use integration by parts:

$$u = 4x \quad dv = e^{-2x} dx$$

$$du = 4 dx \quad v = -\frac{e^{-2x}}{2}$$

$$\begin{aligned} P(X \leq 1.25) &= -2x e^{-2x} \Big|_0^{1.25} + \int_0^{1.25} 2e^{-2x} dx = -2(1.25)e^{-2(1.25)} + 0 - e^{-2x} \Big|_0^{1.25} \\ &= 1 - 3.5e^{-2.5} = 0.7127. \end{aligned}$$

**Why do we need gamma distribution?**

Any normal distribution is bell-shaped and symmetric. There are many practical situations that do not fit to symmetrical distribution. The Gamma family pdfs can yield a wide variety of skewed distributions.  $\beta$  is called the scale parameter because values other than 1 either stretch or compress the pdf in the  $x$ -direction.

## 4.4 Chi-squared Distribution

The r.v.  $X$  is said to have a Chi-Squared distribution with parameter  $\nu$  ( $X \sim \chi_\nu^2$ ) if its pdf is given by

$$f(x) = f(x; \nu) = \begin{cases} \frac{1}{2^{(\frac{\nu}{2})} \Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}; & x \geq 0 \\ 0; & \text{otherwise} \end{cases}$$

**Parameter of the Distribution:**  $\nu > 0$  (The degrees of freedom 'df').

### Mean and Variance

If  $X$  is a continuous random variable has chi-squared distribution with parameter  $\nu$  then,

$$E(X) = \nu \quad \text{and} \quad V(x) = 2\nu.$$

### Moment Generating Function

If  $X$  is a continuous random variable has Chi-squared distribution with parameter  $\nu$  then, the MGF of  $X$  is

$$M_X(t) = \left( \frac{1}{1-2t} \right)^{\frac{\nu}{2}} = (1-2t)^{-\frac{\nu}{2}}.$$

### Note

Table 4.2 in appendix (A) provides some Lower Critical values for Chi-square distribution.

### Example 4.9

If  $X \sim \chi_7^2$  find  $a, b$  if

I.  $P(X < a) = 0.1.$

### Solution

$a = 2.83.$  (From the table)

II.  $P(X \geq b) = 0.99.$

### Solution

$$P(X \geq b) = 1 - P(X < b) = 0.99 \Rightarrow P(X < b) = 1 - 0.99 = 0.01.$$

Therefore,  $b = 1.24.$

## 4.5 Beta Distribution

A continuous random variable  $X$  is said to have a beta distribution  $X \sim \text{Beta}(\alpha, \beta)$  if it has probability density function

$$f(x) = f(x; \alpha, \beta) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}; & 0 < x < 1. \\ 0; & \text{otherwise} \end{cases}$$

Where the beta function is defined as  $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ .

Thus,  $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$ .

**Parameters of the Distribution:**  $\alpha, \beta > 0$ .

### Mean and Variance

If  $X$  is a continuous random variable has beta distribution with parameters  $\alpha$  and  $\beta$  then,

$$E(X) = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad V(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

### Moment Generating Function

The moment-generating function for the beta distribution is complicated. Therefore, we will not mention it.

#### Example 4.10

If  $X \sim \text{Beta}(3, 2)$ , Find  $P(X < 3)$ ,  $P(X = 7)$  and  $P(X < 0.5)$ .

**Solution**

$$f(x) = \frac{\Gamma(3+2)}{\Gamma(3)\Gamma(2)} x^{3-1} (1-x)^{2-1} = \frac{4!}{2!} x^2 (1-x) = 12 x^2 (1-x); 0 < x < 1.$$

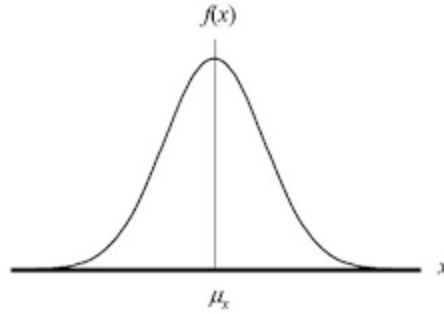
$$P(X < 3) = \int_0^1 12 x^2 (1-x) dx = 1.$$

$$P(X = 7) = 0.$$

$$\begin{aligned} P(X < 0.5) &= \int_0^{0.5} 12 x^2 (1-x) dx = 12 \int_0^{0.5} (x^2 - x^3) dx = 12 \left( \frac{x^3}{3} - \frac{x^4}{4} \right) \Big|_0^{0.5} \\ &= (4x^3 - 3x^4) \Big|_0^{0.5} = 0.3125. \end{aligned}$$

## 4.6 Normal Distribution

The normal distribution is one of the most important continuous distributions. Many measurable characteristics are normally or approximately normally distributed, such as, height and weight. The graph of the probability density function pdf of a normal distribution, called the normal curve, is a bell-shaped curve.



A continuous random variable  $X$  is said to have a normal distribution  $X \sim N(\mu, \sigma)$  if it has probability density function

$$f(x) = f(x; \mu, \sigma) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; & -\infty < x < \infty. \\ 0; & \text{otherwise} \end{cases}$$

**Parameters of the Distribution:**  $-\infty < \mu < \infty$  (The mean),  $\sigma > 0$  (The standard deviation).

### Mean and Variance

If  $X$  is a continuous random variable has normal distribution with parameters  $\mu$  and  $\sigma$  then,

$$E(X) = \mu \text{ and } V(x) = \sigma^2.$$

### Moment Generating Function

If  $X$  is a continuous random variable has normal distribution with parameters  $\mu$  and  $\sigma$  then, the MGF of  $X$  is

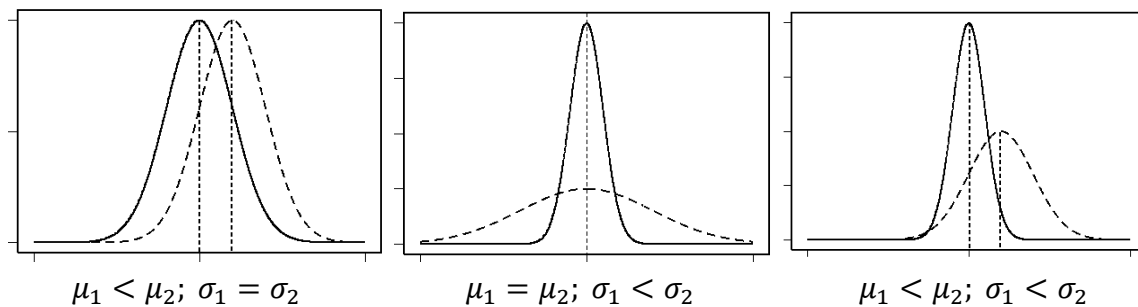
$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

### Note

The proof for the normal distribution MGF will be reviewed later in this chapter.

### 4.6.1 Some properties of the normal curve $f(x)$ of $N(\mu, \sigma)$

- I.  $f(x)$  is symmetric about the mean  $\mu$ .
- II. The total area under the curve of  $f(x) = 1$ .
- III. The highest point of the curve of  $f(x)$  at the mean  $\mu$ .
- IV. The mode, which is the point on the horizontal axis where the curve is a maximum, occurs at  $= \mu$ , (Mode = Median = Mean).
- V. The curve has its points of inflection at  $X = \mu \pm \sigma$  is concave downward if  $\mu - \sigma < X < \mu + \sigma$  and is concave upward otherwise.
- VI. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
- VII. The location of the normal distribution depends on  $\mu$  and its shape depends on  $\sigma$ .



Where the solid line represents  $N(\mu_1, \sigma_1)$ , and the dashed line represents  $N(\mu_2, \sigma_2)$ .

### 4.6.2 Standard Normal Distribution

The special case of the normal distribution where the mean  $\mu = 0$  and the variance  $\sigma^2 = 1$  called the standard normal distribution denote  $N(0,1)$ . Thus, the pdf is reduced to

$$f(z) = f(z; 0,1) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}; & -\infty < z < \infty. \\ 0; & \text{otherwise} \end{cases}$$

#### Notations

- The random variable which has a standard normal distribution is usually denoted by  $Z$ .
- If  $0 < \alpha < 1$  the notation  $z_\alpha$  refers to the point in the standard normal distribution  $Z$  such that  $P(Z < z_\alpha) = \alpha$ .

### Moment Generating Function

If  $Z$  is a continuous random variable has standard normal distribution then, the MGF of  $Z$  is

$$M_Z(t) = e^{\frac{t^2}{2}}.$$

#### Proof

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} f(z) dz = \int_{-\infty}^{\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2-2tz)} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2-2tz+t^2-t^2)} dz \quad (\text{add and subtract } t^2) \\ &= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2-2tz+t^2)} dz = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz = e^{\frac{t^2}{2}} \cdot 1 = e^{\frac{t^2}{2}}. \\ &(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz = 1 \text{ because it is a pdf of a } N(t, 1)). \end{aligned}$$

### Deriving the MGF of a Normal Distribution

Recall the MGF of the normal distribution  $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ .

#### Proof

We know that  $Z = \frac{X-\mu}{\sigma} \Rightarrow X = \sigma Z + \mu$ . Where  $Z \sim N(0,1)$  and  $X \sim N(\mu, \sigma)$ .

Using the theorem: If  $Y = a + bX \Rightarrow M_Y(t) = e^{at} M_X(bt)$ , we get

$$M_X(t) = M_{\sigma Z + \mu}(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} \cdot e^{\frac{1}{2}\sigma^2 t^2} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

#### Note

Table 4.3 in appendix (A) provides the area to the left of  $Z$  for standard normal distribution.

### 4.6.3 Calculating Probabilities of Standard Normal Distribution

The standard normal distribution is very important because probabilities of any normal distribution can be calculated from the probabilities of the standard normal distribution.

- I.  $P(Z \leq a)$  from the table.
- II.  $P(Z \geq b) = 1 - P(Z \leq b)$  where  $P(Z \leq b)$  from the table.
- III.  $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$ , where  $P(Z \leq a)$  and  $P(Z \leq b)$  from the table.

**Proposition**

If  $X \sim N(\mu, \sigma)$ , then  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ .

**4.6.4 Calculating Probabilities of Normal Distribution**

- I.  $P(X \leq a) = P\left(Z \leq \frac{a - \mu}{\sigma}\right)$  from the table.
- II.  $P(X \geq b) = P\left(Z \geq \frac{b - \mu}{\sigma}\right) = 1 - P\left(Z \leq \frac{b - \mu}{\sigma}\right)$ , where  $P\left(Z \leq \frac{b - \mu}{\sigma}\right)$  from the table.
- III.  $P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = P\left(Z \leq \frac{b - \mu}{\sigma}\right) - P\left(Z \leq \frac{a - \mu}{\sigma}\right)$ ,  
where  $P\left(Z \leq \frac{b - \mu}{\sigma}\right)$  and  $P\left(Z \leq \frac{a - \mu}{\sigma}\right)$  from the table.

**Example 4.11**

If  $Z \sim N(0, 1)$ . Find  $P(Z < 1.5)$ ,  $P(Z > 0.98)$ ,  $P(Z < 0)$  and  $P(-1.33 < Z < 2.42)$ .

**Solution**

$$P(Z < 1.5) = 0.9332.$$

$$P(Z > 0.98) = 1 - 0.8365 = 0.1635.$$

$$P(Z < 0) = 0.5.$$

$$P(-1.33 < Z < 2.42) = P(Z < 2.42) - P(Z < -1.33) = 0.9922 - 0.0918 = 0.9004.$$

**Example 4.12**

Suppose that the birth weight of Saudi babies  $X$  has a normal distribution with mean  $\mu = 3.4$  and standard deviation  $\sigma = 0.35$ .

- I. Find the probability that a randomly chosen Saudi baby has a birth weight between 3.0 and 4.0 kg.
- II. What is the percentage of Saudi babies who have a birth weight between 3.0 and 4.0 kg.

**Solution**

$$\begin{aligned} \text{I. } P(3 < X < 4) &= P\left(\frac{3 - 3.4}{0.35} < Z < \frac{4 - 3.4}{0.35}\right) = P(-1.14 < Z < 1.71) \\ &= P(Z < 1.711) - P(Z < -1.14) = 0.9564 - 0.1271 = 0.8293. \end{aligned}$$

$$\text{II. } P(3 < X < 4) \cdot 100\% = 0.8293 \cdot 100\% = 82.93\%.$$

### 4.6.3 Normal Approximation to Binomial

The binomial distribution is symmetrical (like the normal distribution) whenever  $p = 0.5$ . When  $p \neq 0.5$ , the binomial distribution is not symmetrical. However, the closer  $p$  is to 0.5 and the larger the sample size  $n$ , the more symmetric the distribution becomes. Fortunately, whenever the sample size is large, you can use the normal distribution to approximate the exact probabilities of the items of interest.

#### Proposition

If  $X$  is a binomial r.v. with mean  $\mu = np$  and variance  $\sigma^2 = npq$ , then the limiting form of the distribution of

$$Z = \frac{X - np}{\sqrt{npq}},$$

As  $n \rightarrow \infty$ , is the standard normal distribution  $N(0,1)$ .

#### Note

As a general rule, you can use the normal distribution to approximate the binomial distribution whenever  $np, nq \geq 5$ .

#### Steps to solving a Normal approximation to the Binomial distribution

**Step 1.** Check if appropriate to use the approximation.

Rule of Thumb:  $np \geq 5$  and  $nq \geq 5$ .

**Step 2.** Calculate  $\mu = np$  and  $\sigma = \sqrt{npq}$  where  $p + q = 1$ .

**Step 3.** Approximate the r.v.  $X$  with a normal r.v.  $N(\mu = np, \sigma = \sqrt{npq})$ .

**Step 4.** Use the continuity correction factor.

Change all the ' $<$ ' to ' $\leq$ ' and ' $>$ ' to ' $\geq$ ', remember to correct the endpoints to new value (i.e.  $X > 6$  would change to  $X \geq 7$  and  $X < 9$  would change to  $X \leq 8$ ). 2. For the ' $\geq$ ', subtract 0.5, and for the ' $\leq$ ', add 0.5.

$$P(X \geq a) = P(Y \geq a - 0.5),$$

$$P(X \leq b) = P(Y \leq b + 0.5),$$

$$P(X = c) = P(c - 0.5 \leq Y \leq c + 0.5),$$

where  $Y \sim N(\mu, \sigma)$ .



**Step 5.** Solve the normal problem.

$$P(X \geq a) = P\left(Z \geq \frac{(a-0.5)-np}{\sqrt{npq}}\right),$$

$$P(X \leq b) = P\left(Z \leq \frac{(b+0.5)-np}{\sqrt{npq}}\right),$$

$$P(X = c) = P\left(\frac{(c-0.5)-np}{\sqrt{npq}} \leq Z \leq \frac{(c+0.5)-np}{\sqrt{npq}}\right).$$

### Example 4.13

Let  $X$  be the number of times that a fair coin when flipped 40 times lands on head.

Use normal distribution to find the probability that

- I. it will be equal to 20.
- II. it will be less than 17.

### Solution

$$p(H) = p = 0.5 = q = p(T),$$

$$n = 40, x = 20.$$

**Step 1.**  $np = nq = 40 \times 0.5 = 20 \geq 5$ .

**Step 2.**  $\mu = 20, \sigma = \sqrt{npq} = \sqrt{10} = 3.16$ . thus,  $X \approx N(20, 3.16)$ .

**Step 3.**

$$\text{I. } P(X = 20) = P\left(\frac{(20-0.5)-20}{3.16} \leq Z \leq \frac{(20+0.5)-20}{3.16}\right)$$

$$= P(-0.16 \leq Z \leq 0.16)$$

$$= P(Z \leq 0.16) - P(Z \leq -0.16)$$

$$= 0.5636 - 0.4364 = 0.1272.$$

$$\text{II. } P(X < 17) = P(X \leq 16) = P\left(Z \leq \frac{(16+0.5)-20}{3.16}\right)$$

$$= P(Z \leq -1.11) = 0.1335.$$

## 4.7 Student's t Distribution

A continuous random variable  $T$  is said to have a  $t$  distribution with parameter  $\nu$  if it has probability density function

$$f(t) = f(t; \nu) = \begin{cases} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}; & -\infty < t < \infty. \\ 0; & \text{otherwise} \end{cases}$$

**Parameters of the Distribution:**  $\nu > 0$ .

### 4.7.1 Some properties of the $t$ distribution

- I. It has mean of zero.
- II. It is symmetric about zero (mean=median=mode=0).
- III. Compared to the standard normal distribution, the  $t$  distribution is less peaked in the center and has higher tails.
- IV. It depends on the degrees of freedom.
- V.  $t$  distribution approaches the standard normal distribution as  $\nu$  approaches  $\infty$ .

#### Notation

$$P(T < t_{\alpha}) = \alpha.$$

#### Note

Table 4.4 in appendix (A) provides the Lower Critical Values for  $t$  Distribution.

#### Example 4.14

If  $\nu = 15$ , find  $P(T < 1.753)$ ,  $t_{0.99}$ .

#### Solution

$$P(T < 1.753) = 0.95.$$

$$t_{0.99} = 2.602.$$

## Chapter Five

### Joint, Marginal, and Conditional Distributions

In the study of probability, given at least two random variables  $X, Y, \dots$ , that are defined on a probability space, the joint probability distribution for  $X, Y, \dots$  is a probability distribution that gives the probability that each of  $X, Y, \dots$  falls in any particular range or discrete set of values specified for that variable. In the case of only two random variables, this is called a bivariate (joint) distribution, but the concept generalizes to any number of random variables, giving a multivariate distribution.

**For both discrete and continuous random variables we will discuss the following**

- Joint Distributions.
- Cumulative distribution.
- Marginal Distributions (computed from a joint distribution).
- Joint Mathematical Expectation
- Conditional Distributions (e.g.  $P(Y = y|X = x)$ ).
- Joint Moment Generating Function.

### 5.1 Joint Distributions

#### 5.1.1 Joint Probability function

Joint distribution of two random variables  $X$  and  $Y$  has a probability function or probability density function  $f(x, y)$  that is a function of two variables (sometimes denoted  $f_{X,Y}(x, y)$ ).

#### Discrete Case

If  $X = x_1, x_2, \dots, x_n$  and  $Y = y_1, y_2, \dots, y_m$  are two discrete random variables, then the values of the joint probability function of  $X$  and  $Y$  ' $f(x, y)$ ' is

$f(x, y)$	$y_1$	$y_2$	...	$y_j$	...	$y_m$	$\sum_y f(x, y)$
$x_1$	$f(x_1, y_1)$	$f(x_1, y_2)$	...	$f(x_1, y_j)$	...	$f(x_1, y_m)$	$f_X(x_1)$
$x_2$	$f(x_2, y_1)$	$f(x_2, y_2)$	...	$f(x_2, y_j)$	...	$f(x_2, y_m)$	$f_X(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_i$	$f(x_i, y_1)$	$f(x_i, y_2)$	...	$f(x_i, y_j)$	...	$f(x_i, y_m)$	$f_X(x_i)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_n$	$f(x_n, y_1)$	$f(x_n, y_2)$	...	$f(x_n, y_j)$	...	$f(x_n, y_m)$	$f_X(x_n)$
$\sum_x f(x, y)$	$f_Y(y_1)$	$f_Y(y_2)$	...	$f_Y(y_j)$	...	$f_Y(y_m)$	1

$f(x, y) = P(X = x, Y = y)$  must satisfy

- I.  $0 \leq f(x, y) \leq 1$ .
- II.  $\sum_x \sum_y f(x, y) = 1$ .

### Continuous Case

If  $X$  and  $Y$  are continuous random variables, then  $f(x, y)$  must satisfy

- I.  $f(x, y) \geq 0$ .
- II.  $\int_x \int_y f(x, y) dy dx = 1$ .

### 5.1.2 Joint distribution function

If random variables  $X$  and  $Y$  have a joint distribution, then the cumulative distribution function is

$$F(x, y) = P(X \leq x, Y \leq y) = \begin{cases} \sum_{t=-\infty}^y \sum_{s=-\infty}^x f(s, t); & \text{If } X, Y \text{ are discrete r. v.'s} \\ \int_{-\infty}^y \int_{-\infty}^x f(s, t) ds dt; & \text{If } X, Y \text{ are continuous r. v.'s} \end{cases}$$

#### Note

In the continuous case,  $\frac{\partial^2}{\partial x \partial y} F(x, y) = f(x, y)$ .

#### Some Properties of the Joint CDF

- $F(x, y)$  is non-decreasing in both  $x$  and  $y$ .
- $F(x, \infty) = F(x)$ .
- $F(\infty, y) = F(y)$ .
- $F(\infty, \infty) = 1$ .
- $F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0$ .

### 5.1.3 Marginal probability Distributions

If  $X$  and  $Y$  have a joint distribution with joint density or probability function  $f(x, y)$  then

- The marginal distribution of  $X$  has a probability function or density function denoted  $f_X(x)$  is equal to

$$f_X(x) = \begin{cases} \sum_y f(x, y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} f(x, y) dy; & \text{In the continuous case} \end{cases}$$

- The marginal distribution of  $Y$  has a probability function or density function denoted  $f_Y(y)$  is equal to

$$f_Y(y) = \begin{cases} \sum_x f(x, y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} f(x, y) dx; & \text{In the continuous case} \end{cases}$$

### 5.1.4 Joint Mathematical Expectation

If  $g(x, y)$  is a function of two variables, and  $X$  and  $Y$  are jointly distributed random variables with joint probability function  $f(x, y)$ , then the expected value of  $g(x, y)$  is defined to be

$$E[g(X, Y)] = \begin{cases} \sum_y \sum_x g(x, y) f(x, y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy; & \text{In the continuous case} \end{cases}$$

#### Special cases

- I. If  $g(X, Y) = X$ , we get

$$E[g(X, Y)] = \begin{cases} \sum_y \sum_x x f(x, y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy; & \text{In the continuous case} \end{cases}$$

$$= \begin{cases} \sum_x x \sum_y f(x, y) = \sum_x x f(x) = E(X); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x, y) dy dx = \int_{-\infty}^{\infty} x f(x) dx = E(X); & \text{In the continuous case} \end{cases}$$

Similarly for  $g(X, Y) = Y$ .

- II. If  $g(X, Y) = (X - \mu)^2$ , we get

$$E[g(X, Y)] = \begin{cases} \sum_y \sum_x (x - \mu)^2 f(x, y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu)^2 f(x, y) dx dy; & \text{In the continuous case} \end{cases}$$

$$= \begin{cases} \sum_x (x - \mu)^2 \sum_y f(x, y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} (x - \mu)^2 \int_{-\infty}^{\infty} f(x, y) dy dx; & \text{In the continuous case} \end{cases}$$

$$= \begin{cases} \sum_x (x - \mu)^2 f(x) = V(X); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = V(X); & \text{In the continuous case} \end{cases}$$

Similarly for  $g(X, Y) = (Y - \mu)^2$ .

### Example 5.1

A company that services air conditioner units in residences and office blocks is interested in how to schedule its technicians in the most efficient manner. The random variable  $X$ , taking the values 1, 2, 3 and 4, is the service time in hours. The random variable  $Y$ , taking the values 1, 2 and 3, is the number of air conditioner units. The joint probability function for  $X$  and  $Y$  is given in the table below

		X: Service Time			
		1	2	3	4
Y: Number of Air Conditions Units	1	0.12	0.08	0.07	0.05
	2	0.08	0.15	0.21	0.13
	3	0.01	0.01	0.02	0.07

- I. Proof that  $f(x, y)$  is a joint probability function.
- II. Find:  $f(2, 1), f_X(3), F(2, 2), P(X < 3, Y > 2), P(X + Y \leq 4)$ .
- III. Find the marginal function  $f_Y(y)$ .
- IV. Find  $E(Y), V(Y), E(XY)$ .

### Solution

- I. First, it is clear that  $0 \leq f(x, y) \leq 1, \forall x, y$ .

Second,  $\sum_x \sum_y f(x, y) = 0.12 + 0.08 + \dots + 0.07 = 1$ .

- II.  $f(2, 1) = 0.08$ .

$$f_X(3) = 0.07 + 0.21 + 0.02 = 0.3.$$

$$F(2, 2) = f(1, 1) + f(1, 2) + f(2, 1) + f(2, 2) = 0.12 + 0.08 + 0.08 + 0.15 = 0.43.$$

$$P(X < 3, Y > 2) = f(1, 3) + f(2, 3) = 0.01 + 0.01 = 0.02.$$

$$P(X + Y \leq 4) = f(1, 1) + f(1, 2) + f(1, 3) + f(2, 1) + f(2, 2) + f(3, 1) = 0.51.$$

III.

Y	1	2	3	Sum
$f_Y(y)$	0.32	0.57	0.11	1

IV.

Y	1	2	3	Sum
$f_Y(y)$	0.32	0.57	0.11	1
$yf(y)$	0.32	1.14	0.33	$E(Y) = 1.79$
$y^2f(y)$	0.32	2.28	0.99	$E(Y^2) = 3.59$

Thus,  $E(Y) = 1.79$ , and  $V(Y) = E(Y^2) - [E(Y)]^2 = 3.59 - (1.79)^2 = 0.3859$ .

Now,

$$E(XY) = \sum_{y=1}^3 \sum_{x=1}^4 xyf(x, y) = (1)(1)(0.12) + (1)(2)(0.08) + (1)(3)(0.01) + \dots + (4)(3)(0.07) = 4.86.$$

**Example 5.2**

Consider the joint probability function

$$f(x, y) = c(x + y); \quad 0 < x, y < 2.$$

Find  $c, f_X(x), f_Y(y), F(1,1), V(X), E[X(X + 6)]$ .**Solution**

- To find  $c$  we know that

$$\begin{aligned} \int_x \int_y f(x, y) dy dx &= 1 \Rightarrow 1 = c \int_0^2 \int_0^2 (x + y) dy dx = c \int_0^2 \left[ xy + \frac{y^2}{2} \right] \Big|_0^2 dx \\ &= c \int_0^2 (2x + 2) dx = c [x^2 + 2x] \Big|_0^2 = 8c \Rightarrow c = \frac{1}{8}. \end{aligned}$$

Thus,  $f(x, y) = \frac{x+y}{8}$ .

- $f_X(x) = \frac{1}{8} \int_0^2 (x + y) dy = \frac{1}{8} \left[ xy + \frac{y^2}{2} \right] \Big|_0^2 = \frac{1}{8} (2x + 2) = \frac{x+1}{4}; \quad 0 < x < 2.$
- $f_Y(y) = \frac{1}{8} \int_0^2 (x + y) dx = \frac{1}{8} \left[ \frac{x^2}{2} + xy \right] \Big|_0^2 = \frac{1}{8} (2y + 2) = \frac{y+1}{4}; \quad 0 < y < 2.$
- $F(1,1) = P(X \leq 1, Y \leq 1) = \frac{1}{8} \int_0^1 \int_0^1 (x + y) dy dx = \frac{1}{8} \int_0^1 \left[ xy + \frac{y^2}{2} \right] \Big|_0^1 dx = \frac{1}{8} \int_0^1 \left( x + \frac{1}{2} \right) dx = \frac{1}{8} \left[ \frac{x^2}{2} + \frac{x}{2} \right] \Big|_0^1 = \frac{1}{8}.$
- $V(X) = E(X^2) - [E(X)]^2$

$$E(X) = \frac{1}{4} \int_0^2 x(x + 1) dx = \frac{1}{4} \int_0^2 (x^2 + x) dx = \frac{1}{4} \left[ \frac{x^3}{3} + \frac{x^2}{2} \right] \Big|_0^2 = \frac{1}{4} \left[ \frac{8}{3} + 2 \right] = \frac{7}{6}.$$

$$E(X^2) = \frac{1}{4} \int_0^2 x^2(x+1) dx = \frac{1}{4} \int_0^2 (x^3 + x^2) dx = \frac{1}{4} \left[ \frac{x^4}{4} + \frac{x^3}{3} \right] \Big|_0^2 = \frac{1}{4} \left[ \frac{16}{4} + \frac{8}{3} \right] = \frac{5}{3}.$$

$$\Rightarrow V(X) = \frac{5}{3} - \left[ \frac{7}{6} \right]^2 = \frac{11}{36}.$$

$$\blacksquare E[X(X+6)] = E(X^2 + 6X) = E(X^2) + 6E(X) = \frac{5}{3} + 6 \left( \frac{7}{6} \right) = \frac{26}{3}.$$

### 5.1.5 Joint Moment Generating Function

Given jointly distributed random variables X and Y, the moment generating function of the joint distribution is

$$M_{X,Y}(t_1, t_2) = E(e^{Xt_1 + Yt_2}) = \begin{cases} \sum_x \sum_y e^{xt_1 + yt_2} f(x, y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{xt_1 + yt_2} f(x, y) dx dy; & \text{In the discrete case} \end{cases}$$

where,  $-\infty < t_1, t_2 < \infty$ .

#### Some Properties

If X, Y are r.v's and  $r_1, r_2$  are integer values, then

I.  $M_{X,Y}(0,0) = 1.$

II.  $M_{X,Y}(t_1, 0) = M_X(t_1).$

III.  $M_{X,Y}(0, t_2) = M_Y(t_2).$

IV.  $\frac{\partial^{r_1}}{\partial t_1^{r_1}} M_{X,Y}(t_1, t_2) \Big|_{t_1=t_2=0} = E(X^{r_1}).$

V.  $\frac{\partial^{r_2}}{\partial t_2^{r_2}} M_{X,Y}(t_1, t_2) \Big|_{t_1=t_2=0} = E(Y^{r_2}).$

VI.  $\frac{\partial^{r_1+r_2}}{\partial t_1^{r_1} \partial t_2^{r_2}} M_{X,Y}(t_1, t_2) \Big|_{t_1=t_2=0} = E(X^{r_1} Y^{r_2}).$  (The  $(r_1 + r_2)$ th joint raw moment)

#### Example 5.3

Consider the joint probability function

		Y		
$f(x,y)$		-2	0	5
X	1	0.15	0.25	0.2
	3	0.2	0.05	0.15

Find  $M(t_1, t_2)$ , then use it to find  $E(X), E(XY)$ .

#### Solution

$$M(t_1, t_2) = \sum_{x=1}^3 \sum_{y=-2}^5 e^{xt_1 + yt_2} f(x, y)$$



$$= 0.15e^{t_1-2t_2} + 0.25e^{t_1} + 0.2e^{t_1+5t_2} + 0.2e^{3t_1-2t_2} + 0.05e^{3t_1} + 0.15e^{3t_1+5t_2}$$

$$E(X) = \frac{\partial}{\partial t_1} M_{X,Y}(t_1, t_2) \Big|_{t_1=t_2=0}$$

$$= 0.15e^{t_1-2t_2} + 0.25e^{t_1} + 0.2e^{t_1+5t_2} + 0.6e^{3t_1-2t_2} + 0.15e^{3t_1} + 0.45e^{3t_1+5t_2} \Big|_{t_1=t_2=0}$$

$$= 1.8.$$

$$E(XY) = \frac{\partial^2}{\partial t_1 \partial t_2} M_{X,Y}(t_1, t_2) \Big|_{t_1=t_2=0} = \frac{\partial}{\partial t_2} [0.15e^{t_1-2t_2} + 0.25e^{t_1} + 0.2e^{t_1+5t_2} +$$

$$0.6e^{3t_1-2t_2} + 0.15e^{3t_1} + 0.45e^{3t_1+5t_2}] \Big|_{t_1=t_2=0}$$

$$= [-0.3e^{t_1-2t_2} + e^{t_1+5t_2} - 1.2e^{3t_1-2t_2} + 2.25e^{3t_1+5t_2}] \Big|_{t_1=t_2=0} = 1.75.$$

## 5.2 Conditional Distributions

### 5.2.1 Conditional Probability function

The probability of the random variable  $X$  under the knowledge provided by the value of  $Y$  is given by

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

Note that  $f_{X|Y=y}(x|y)$  must satisfy

$$\text{I. } f_{X|Y=y}(x|y) \geq 0.$$

$$\text{II. } \begin{cases} \sum_x f_{X|Y=y}(x|y) = 1; & \text{In the discrete case} \\ \int_{-\infty}^{\infty} f_{X|Y=y}(x|y) dx = 1; & \text{In the continuous case} \end{cases}$$

Similarly,

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

### 5.2.2 Conditional Distribution Function

The conditional CDF is as follows

$$F(x|y) = P[X \leq x | Y = y] = \begin{cases} \sum_{s=-\infty}^x f(s|y); & \text{In the discrete case} \\ \int_{-\infty}^x f(u|y) du; & \text{In the continuous case} \end{cases}$$

**Note**

$$\frac{\partial}{\partial x} F(x|y) = f(x|y).$$

**Example 5.4**

If the joint probability function of X and Y is given by  $f(x, y) = 2$ ;  $0 < x < y < 1$ .

Find  $f(x), f(y), f(x|y), f(y|x), F(x|y), F(y|x), f(x|0.5), F(y|0.25)$ .

**Solution**

- $f(x) = \int_x^1 2 dy = 2y|_x^1 = 2 - 2x = 2(1 - x); 0 < x < 1.$
- $f(y) = \int_0^y 2 dx = 2x|_0^y = 2y; 0 < y < 1.$
- $f_{X|Y=y}(x|y) = \frac{f(x,y)}{f(y)} = \frac{2}{2y} = \frac{1}{y}; 0 < x < y < 1.$
- $f_{Y|X=x}(y|x) = \frac{f(x,y)}{f(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x}; 0 < x < y < 1.$
- $F(x|y) = \int_0^x f(x|y) dx = \int_0^x \frac{1}{y} dx = \frac{x}{y}|_0^x = \frac{x}{y}; 0 < x < y < 1.$
- $F(y|x) = \int_x^y f(y|x) dy = \int_x^y \frac{1}{1-x} dy = \frac{y}{1-x}|_x^y = \frac{y-x}{1-x}; 0 < x < y < 1.$
- $f(x|0.5) = \frac{1}{0.5} = 2; 0 < x < \frac{1}{2}.$
- $F(y|0.25) = \frac{y-0.25}{0.75} = \frac{1}{3}(4y - 1); \frac{1}{4} < y < 1.$

**5.2.3 Conditional Expectation**

If X and Y are two r.v's have a joint probability distribution  $f(x, y)$ , then the conditional expectation of any function of X,  $g(x)$  given  $Y = y$  is

$$E[g(X)|Y = y] = \begin{cases} \sum_x g(x)f_{X|Y=y}(x|y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} g(x)f_{X|Y=y}(x|y) dx; & \text{In the continuous case} \end{cases}$$

Similarly for  $E[h(Y)|X = x]$ , where  $h(y)$  is a function of the r.v. Y.

**Special Cases**

I. If  $g(X) = X$ , we get

$$E(X|Y = y) = \begin{cases} \sum_x x f_{X|Y=y}(x|y); & \text{In the discrete case} \\ \int_{-\infty}^{\infty} x f_{X|Y=y}(x|y) dx; & \text{In the continuous case} \end{cases} = \mu_{X|Y},$$

which is the conditional expectation of X given Y (this expectation is considered as a variable of Y).

II. If  $g(X) = (X - \mu_{X|Y})^2$ , we get

$$E[(X - \mu_{X|Y})^2 | Y = y] = V(X|Y = y) = \sigma_{X|Y}^2,$$

which is the conditional variance of  $X$  given  $Y$

### Corollary

The conditional variance can be expressed as

$$V(X|Y = y) = E(X^2|Y = y) - [E(X|Y = y)]^2 = E(X^2|Y = y) - \mu_{X|Y}^2.$$

### Theorem

Let  $X, Y$  be random variables,  $a, b \in R$ , and  $g : R \rightarrow R$ . Assuming all the following expectations exist, we have

- I.  $E(a|Y) = a$ .
- II.  $E[(aX + b)|Y] = aE(X|Y) + b$ .
- III.  $E[E(X|Y)] = E(X)$ , similarly  $E[E(Y|X)] = E(Y)$ .

### Proof

The first two are not hard to prove, and we leave them to the reader.

Consider (III). We prove the continuous case and leave the discrete case to the reader.

$E(X|Y) = \int_{-\infty}^{\infty} x f_{X|Y=y}(x|y) dx$  (a function of  $Y$ ). Thus,

$$\begin{aligned} E[E(X|Y)] &= \int_{-\infty}^{\infty} E(X|Y) f_Y(y) dy = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} x f_{X|Y=y}(x|y) dx \right] f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot \frac{f(x,y)}{f_Y(y)} \cdot f_Y(y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy \int_{-\infty}^{\infty} x \left[ \int_{-\infty}^{\infty} f(x,y) dy \right] dx \\ &= \int_{-\infty}^{\infty} x f(x) dx = E(X). \end{aligned}$$

### Theorem

If  $X, Y$  have a joint distribution, then the marginal variance of  $X$  can be factored in the form

$$V(X) = E_Y[V(X|Y)] + V_Y[E(X|Y)].$$

Similarly,

$$V(Y) = E_X[V(Y|X)] + V_X[E(Y|X)].$$

### Example 5.5

If  $f(x|y) = \frac{1}{y}; 0 < x < y < 1$ . Find  $E(X|Y), V(X|Y)$ .

### Solution

$$E(X|Y) = \int_0^y x f(x|y) dx = \int_0^y \frac{x}{y} dx = \frac{x^2}{2y} \Big|_0^y = \frac{y}{2}; \quad 0 < y < 1.$$

$$E(X^2|Y) = \int_0^y x^2 f(x|y) dx = \int_0^y \frac{x^2}{y} dx = \frac{x^3}{3y} \Big|_0^y = \frac{y^2}{3}; \quad 0 < y < 1. \text{ Thus,}$$

$$V(X|Y) = E(X^2|Y) - [E(X|Y)]^2 = \frac{y^2}{3} - \frac{y^2}{4} = \frac{y^2}{12}; \quad 0 < y < 1.$$

### 5.2.4 Moment Generating Function for Conditional Distributions

If  $X, Y$  have a joint probability distribution, and  $f(x|y)$  is the conditional probability function of  $X$ , then the moment generating function for the conditional distribution (if it exist) defined as

$$M_{X|Y}(t) = E(e^{tX}|Y = y) = \begin{cases} \sum_x e^{xt} f(x|y) \\ \int_{-\infty}^{\infty} e^{xt} f(x|y) dx \end{cases}$$

#### Properties

I.  $M_{X|Y}(0) = 1.$

II.  $\frac{d^r}{dt^r} M_{X|Y}(t)|_{t=0} = E(X^r|Y = y)$ , which is the  $r^{\text{th}}$  conditional raw moment.

#### Example 5.6

If  $f(x|y) = \frac{1}{y}; 0 < x < y < 1$ . Find  $M_{X|Y}(t)$  then use it to compute  $E(X|Y), V(X|Y)$ .

#### Solution

$$M_{X|Y}(t) = \int_0^y e^{xt} f(x|y) dx = \int_0^y \frac{e^{xt}}{y} dx = \frac{e^{xt}}{yt} \Big|_0^y = \frac{e^{yt}-1}{yt}.$$

Now,

$$\begin{aligned} M_{X|Y}(t) &= \frac{e^{yt}-1}{yt} = \frac{1}{yt} \left( \sum_{r=0}^{\infty} \frac{(yt)^r}{r!} - 1 \right) = \frac{1}{yt} \left( 1 + yt + \frac{y^2 t^2}{2!} + \frac{y^3 t^3}{3!} + \dots + \frac{y^r t^r}{r!} + \dots - 1 \right) \\ &= 1 + \frac{yt}{2!} + \frac{y^2 t^2}{3!} + \dots + \frac{y^r t^r}{(r+1)!} + \dots = 1 + \frac{y}{2} \cdot \frac{t}{1!} + \frac{y^2}{3} \cdot \frac{t^2}{2!} + \dots + \frac{y^r}{(r+1)} \cdot \frac{t^r}{r!} + \dots. \end{aligned}$$

Therefore,

$$\mu'_r = \frac{y^r}{(r+1)}.$$

Hence,

$$E(X|Y) = \mu'_{X|Y} = \frac{y}{2}.$$

$$E(X^2|Y) = \mu'_2 = \frac{y^2}{3}.$$

$$V(X|Y) = \frac{y^2}{12}.$$

## Chapter Six

### Covariance, Correlation, Independence of Variables (Stochastic Independence)

#### 6.1 Covariance of random variables

If random variables  $X$  and  $Y$  are jointly distributed with joint probability function  $f(x, y)$ , then the covariance between  $X$  and  $Y$  is defined as

$$\begin{aligned}\sigma_{X,Y} = \text{Cov}(X, Y) &= E\{[X - E(X)][Y - E(Y)]\} = E[(X - \mu_X)(Y - \mu_Y)] \\ &= E(XY) - E(X)E(Y).\end{aligned}$$

#### Proof

$$\begin{aligned}\sigma_{X,Y} = \text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] = E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y) \\ &= E(XY) - E(X\mu_Y) - E(Y\mu_X) + E(\mu_X\mu_Y) = E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y \\ &= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y = E(XY) - 2\mu_X\mu_Y + \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y).\end{aligned}$$

#### Some Properties for Covariance

If  $X, Y, Z$  are r.v's and  $a, b$  are constants, then

- I.  $\text{Cov}(X, X) = V(X)$ .
- II.  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ .
- III.  $\text{Cov}(X, a) = 0$ .
- IV.  $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$ .
- V.  $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$ , this property can be generalized to
$$\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j).$$
- VI.  $V(X \pm Y) = V(X) + V(Y) \pm 2\text{Cov}(X, Y)$ .

#### Proof

II, III, IV are not hard to prove, and we leave them to the reader.

- I.  $\text{Cov}(X, X) = E(X \cdot X) - E(X)E(X) = E(X^2) - [E(X)]^2 = V(X)$ .
- V.  $\text{Cov}(X + Y, Z) = E[(X + Y)Z] - E(X + Y)E(Z)$ 

$$= E(XZ + YZ) - [E(X) + E(Y)]E(Z) = E(XZ) + E(YZ) - E(X)E(Z) - E(Y)E(Z)$$

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

$$\text{VI. } V(X + Y) = \text{Cov}(X + Y, X + Y) \text{ (From I)}$$

$$= \text{Cov}(X, X) + \text{Cov}(X, Y) + \text{Cov}(Y, X) + \text{Cov}(Y, Y) \text{ (From V)}$$

$$= \text{Cov}(X, X) + \text{Cov}(X, Y) + \text{Cov}(X, Y) + \text{Cov}(Y, Y) \text{ (From II)}$$

$$= V(X) + V(Y) + 2\text{Cov}(X, Y). \text{ (From I)}$$

We can prove the case  $V(X - Y)$  by the same way.

## 6.2 Correlation Coefficient

The correlation is a measure of the linear relationship between X and Y. It is obtained by

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

where  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of X and Y respectively.

### Some Properties for Correlation

If X, Y are r.v's and a, b, c, d are constants, then

$$\text{I. } \rho_{X,Y} = \rho_{Y,X}.$$

$$\text{II. } \rho_{X,X} = 1. \text{ (strong positive relationship)}$$

$$\text{III. } \rho_{X,-X} = -1. \text{ (strong negative relationship)}$$

$$\text{IV. } -1 \leq \rho_{X,Y} \leq 1.$$

$$\text{V. } \rho_{(aX \pm b), (cY \pm d)} = \rho_{X,Y}.$$

### Proof

$$\text{VI. } \rho_{(aX \pm b), (cY \pm d)} = \frac{\text{Cov}(aX \pm b, cY \pm d)}{\sqrt{V(aX \pm b)V(cY \pm d)}} = \frac{ac\text{Cov}(X, Y)}{\sqrt{a^2V(X)c^2V(Y)}} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \rho_{X,Y}.$$

### Example 6.1

If  $f(x, y) = \frac{x+y}{8}$ ;  $0 < x, y < 2$ . Find  $\text{Cov}(X, Y)$ ,  $V(X + Y)$ ,  $\rho_{X,Y}$ .

### Solution

From example 5.2 we got

$$E(X) = E(Y) = \frac{7}{6}, \text{ and } V(X) = V(Y) = \frac{11}{36}.$$

$$\blacksquare \quad \text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$\begin{aligned} E(XY) &= \int_0^2 \int_0^2 (xy) \frac{x+y}{8} dx dy = \frac{1}{8} \int_0^2 \int_0^2 (x^2y + xy^2) dx dy \\ &= \frac{1}{8} \int_0^2 \left( \frac{x^3y}{3} + \frac{x^2y^2}{2} \right) \Big|_0^2 dy = \frac{1}{8} \int_0^2 \left( \frac{8y}{3} + 2y^2 \right) dy = \frac{1}{8} \left( \frac{4y^2+2y^3}{3} \right) \Big|_0^2 = \frac{4}{3} = 1.333. \end{aligned}$$

$$\text{Cov}(X, Y) = \frac{4}{3} - \left( \frac{7}{6} \right)^2 = -0.0278.$$

$$\blacksquare \quad \rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} = -\frac{0.0278}{\sqrt{\left(\frac{11}{36}\right)^2}} = -0.0909.$$

$$\blacksquare \quad V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X, Y) = \frac{11}{36} + \frac{11}{36} - 2 \cdot 0.0278 = 0.5556.$$

### 6.3 Independence of random variables

Random variables  $X$  and  $Y$  with cumulative distribution functions  $F(x)$  and  $F(y)$  are said to be independent (or stochastically independent) if and only if the cumulative distribution function of the joint distribution  $F(x, y)$  can be factored in the form

$$F(x, y) = F(x)F(y); \text{ for all } (x, y).$$

Alternatively, stochastic independence can be defined via the probability functions, that,  $X$  and  $Y$  are independent if and only if

$$f(x, y) = f(x)f(y); \text{ for all } (x, y).$$

#### Corollary

If  $X$  and  $Y$  are two independent r.v.'s then

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{f(x)f(y)}{f(y)} = f(x).$$

Similarly,

$$f(y|x) = f(y).$$

#### Note

To prove that any two variables  $X$  and  $Y$  are independent we only need to prove one of the following

$$\text{I. } F(x, y) = F(x)F(y).$$

$$\text{II. } F(x|y) = F(x).$$

$$\text{III. } F(y|x) = F(y).$$

IV.  $f(x, y) = f(x)f(y)$ .

V.  $f(x|y) = f(x)$ .

VI.  $f(y|x) = f(y)$ .

### 6.3.1 Joint Expectation Under Independence Condition

If the two r.v's X and Y are independent, then

I.  $E(XY) = E(X)E(Y)$ ,

II.  $E(X|Y) = E(X)$ ,

III.  $E(Y|X) = E(Y)$ ,

(and vise versa).

#### Example 6.2

If  $f(x, y) = \begin{cases} 1; & 0 < x < 1, 0 < y < 1 \\ 0; & \text{Otherwis} \end{cases}$ , check if X and Y are independent.

#### Solution

First, let find the marginal probability functions

$$f(x) = \int_0^1 1 dy = y|_0^1 = 1, \text{ and}$$

$$f(y) = \int_0^1 1 dx = x|_0^1 = 1.$$

Now, since

$$f(x, y) = 1 = f(x)f(y).$$

Thus, the random variables X and Y are independent.

What is the distribution of X (or Y)?!

#### Example 6.3

Are the random variables X and Y with the following joint probability density table independent?

		Y values			
		0	1	2	3
X values	0	$\frac{1}{8}$	0	0	0
	1	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
	2	0	$\frac{1}{4}$	$\frac{1}{8}$	0
	3	0	$\frac{1}{8}$	0	0



**Solution**

First, let find the marginal probability functions

		Y values				$f_X(x)$
		0	1	2	3	
X values	0	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
	1	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
	2	0	$\frac{1}{4}$	$\frac{1}{8}$	0	$\frac{3}{8}$
	3	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
$f_Y(y)$		$\frac{1}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	1

Now, since

$$f(0,0) = \frac{1}{8} \neq \frac{1}{64} = \frac{1}{8} \cdot \frac{1}{8} = f_X(0)f_Y(0).$$

Thus, the random variables X and Y are not independent (dependent).

(We can use any pair other than (0,0) to reject that X and Y are independent).

**6.3.2 Covariance Under Independence Condition**

If X and Y are two independent r.v.'s then

- I.  $Cov(X, Y) = 0$ . (But the converse is not true in general)
- II.  $V(X \pm Y) = V(X) + V(Y)$ .

**Proof**

- I.  $Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$ .
- II.  $V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X, Y) = V(X) + V(Y) \pm 0$  (From I)  
 $= V(X) + V(Y)$ .

**6.3.3 Correlation Under Independence Condition**

If X and Y are two independent r.v.'s then

$$\rho_{X,Y} = Corr(X, Y) = 0;$$

but the converse is not true in general.

**Example 6.4**

Let the joint probability density function of X and Y is

		Y values			$f(x)$
		-1	0	1	
X values	-1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{5}{16}$
	0	$\frac{3}{16}$	0	$\frac{3}{16}$	$\frac{6}{16}$
	1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{5}{16}$
$f(y)$		$\frac{5}{16}$	$\frac{6}{16}$	$\frac{5}{16}$	1

- I. Are X and Y independent.
- II. Find  $Cov(X, Y)$  and  $\rho_{X,Y}$ .

**Solution**

- I. Since,

$$f(-1, -1) = \frac{1}{16} = 0.0625 \neq 0.0977 = \frac{25}{256} = \frac{5}{16} \cdot \frac{5}{16} = f_X(-1)f_Y(-1).$$

Therefore, X and Y are dependent.

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

$$E(X) = E(Y) = -1 \cdot \frac{5}{16} + 0 \cdot \frac{6}{16} + 1 \cdot \frac{5}{16} = 0.$$

$$E(XY) = -1 \cdot -1 \cdot \frac{1}{16} - 1 \cdot 0 \cdot \frac{3}{16} - 1 \cdot 1 \cdot \frac{1}{16} + \dots + 1 \cdot 1 \cdot \frac{1}{16} = 0. \text{ Thus,}$$

$$Cov(X, Y) = 0 - 0 = 0. \text{ Therefore,}$$

$$\rho_{X,Y} = 0.$$

**Note**  $Cov(X, Y) = \rho_{X,Y} = 0$  even that X and Y are dependent.

## Chapter Seven

### Distributions of Functions of Random Variables (Transformations)

In this chapter, we will study how to find the distribution of a function of a random variable with known distribution, which is called transformations of variables.

#### 7.1 Discrete Case

##### 7.1.1 The Case of One Variable

Suppose that  $X$  is a discrete random variable with probability function  $f(x)$ . If  $g(x)$  is a function of  $x$ , and  $Y$  is a random variable defined by the equation  $Y = g(X)$ , then  $Y$  is a discrete random variable with probability function  $f(y) = \sum_{y=g(x)} f(x)$  - given a value of  $y$ , find all values of  $x$  for which  $y = g(x)$ , (say,  $g(x_1) = g(x_2) = \dots = g(x_t) = y$ ), and then  $g(y)$  is the sum of those  $f(x_i)$  probabilities.

##### Corollary

If  $X$  and  $Y$  are independent random variables, and  $g$  and  $h$  are functions, then the random variables  $g(x)$  and  $h(y)$  are independent.

##### There are two cases

- I. One-to-one correspondence.
- II. Not one-to-one correspondence.

However, we will focus on the first case.

If  $g$  is a one-to-one function, then the inverse image of a single value is itself a single value. For instance,  $g(x) = x^3$ , this inverse function is the cube root, while  $g(x) = x^2$ , this inverse function is the square root which may results in two values.

##### Steps to Obtain $f_Y(y)$ for One-To-One Functions

- I. Compute  $Y$  values that corresponding to  $X$  values,  $y = g^{-1}(x_1), g^{-1}(x_2), \dots$
- II. Find the inverse  $x = g^{-1}(y)$
- III.  $f_Y(y) = P(Y = y) = P(g(X) = y) = P(X = g^{-1}(y)) = f_X(g^{-1}(y))$ .

Hence, the pmf of Y is

$$f_Y(y) = f_X(g^{-1}(y)); \quad y = g(x_1), g(x_2), \dots$$

### Example 7.1

If the r.v. X has pmf  $f_X(x) = \frac{x}{15}; x = 1, 2, 3, 4, 5$ . Find the pmf of the r.v. Y whrer  $Y = X - 3$ .

#### Solution

Note that  $Y = g(X) = X - 3$  is a one-to-one function. Thus,

- $x = 1, 2, 3, 4, 5 \Rightarrow y = (1 - 3), (2 - 3), (3 - 3), (4 - 3), (5 - 3) = -2, -1, 0, 1, 2$ .
- $Y = X - 3 \Rightarrow g^{-1}(Y) = X = Y + 3 \Rightarrow g^{-1}(y) = x = y + 3$ .
- $f_Y(y) = f_X(g^{-1}(y)) = f_X(y + 3) = \frac{y+3}{15}$ .

Thus, ;  $f_Y(y) = \frac{y+3}{15}; y = -2, -1, 0, 1, 2$ .

### Example 7.2

If the r.v. X has pmf  $f_X(x) = \frac{1}{3}; x = 0, 1, 2$ . Find the pmf of the r.v. Y whrer  $Y = X^3$ .

#### Solution

Note that  $Y = g(X) = X^3$  is a one-to-one function. Thus,

- $x = 0, 1, 2 \Rightarrow y = 0^3, 1^3, 2^3 = 0, 1, 8$ .
- $Y = X^3 \Rightarrow g^{-1}(Y) = X = \sqrt[3]{Y} = Y^{\frac{1}{3}} \Rightarrow g^{-1}(y) = x = y^{\frac{1}{3}}$ .
- $f_Y(y) = f_X(g^{-1}(y)) = f_X\left(y^{\frac{1}{3}}\right) = \frac{1}{3}$ .

Thus, ;  $f_Y(y) = \frac{1}{3}; y = 0, 1, 8$ .

## 7.1.2 The Case of Two Variables

Suppose the two discrete r.v.'s  $(X_1, X_2)$  has joint probability function  $f_{X_1 X_2}(x_1, x_2)$  and joint sample space  $\Omega_{X_1 X_2}$ . Let  $(Y_1, Y_2)$  be some function of  $(X_1, X_2)$  defined by  $Y_1 = g_1(X_1, X_2)$  and  $Y_2 = g_2(X_1, X_2)$  with the single-valued inverse given by  $X_1 = g_1^{-1}(Y_1, Y_2)$  and  $X_2 = g_2^{-1}(Y_1, Y_2)$ . Let  $\Omega_{Y_1 Y_2}$  be the sample space of  $Y_1, Y_2$ . Then, the joint probability function of  $(Y_1, Y_2)$  is given by

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)).$$

**Example 7.3**

Let the two r.v.'s  $X_1, X_2$  have a joint probability function as follow

		$x_2$			
$f(x_1, x_2)$		0	1	2	3
$x_1$	0	0.06	0.07	0.11	0.07
	1	0.08	0.09	0.12	0.09
	2	0.06	0.08	0.10	0.07

Find the pmf of the r.v.  $Y$  where  $Y = X_1 + X_2$ .

**Solution**

- $x_1 = 0, 1, 2$  &  $x_2 = 0, 1, 2, 3 \Rightarrow y = 0, 1, 2, 3, 4, 5$ .
- We will compute the values of  $f_Y(y)$  by equivalency as follow
- $f_Y(y) = P(Y = y)$ , Thus,
 
$$f_Y(0) = P(X_1 = 0, X_2 = 0) = 0.06,$$

$$f_Y(1) = P(X_1 = 0, X_2 = 1) + P(X_1 = 1, X_2 = 0) = 0.07 + 0.08 = 0.15,$$

$$f_Y(2) = P(X_1 = 0, X_2 = 2) + P(X_1 = 1, X_2 = 1) + P(X_1 = 2, X_2 = 0)$$

$$= 0.11 + 0.09 + 0.06 = 0.26,$$

$$f_Y(3) = P(X_1 = 0, X_2 = 3) + P(X_1 = 1, X_2 = 2) + P(X_1 = 2, X_2 = 1)$$

$$= 0.07 + 0.12 + 0.08 = 0.27,$$

$$f_Y(4) = P(X_1 = 1, X_2 = 3) + P(X_1 = 2, X_2 = 2) = 0.09 + 0.1 = 0.19,$$

$$f_Y(5) = P(X_1 = 2, X_2 = 3) = 0.07.$$

Therefore,

$y$	0	1	2	3	4	5
$f_Y(y)$	0.06	0.15	0.26	0.27	0.19	0.07

**7.2 Continuous Case**

There are three techniques to compute the distribution of function of random variable:

- Method of distribution function. ( $F(x)$ )
- Method of change-of-variable. (One-to-One transformation)
- Method of moment-generating function. ( $M_X(t)$ )

### 7.2.1 Distribution Function Method (CDF)

Let  $X_1, \dots, X_n \sim f(x_1, \dots, x_n)$  and  $Y = g(X_1, \dots, X_n)$ . Then we follow the following steps to obtain  $f_Y(y)$  by using the CDF technique

- I. Find  $F_X(x)$ . (If it not given)
- II. Find y rang in terms of x.
- III. Compute  $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$  over the region where  $Y \leq y$ .
- IV. Compute  $f_Y(y) = \frac{dF_Y(y)}{dy}$  (by derivation the CDF).

#### Example 7.4

Let the probability density function of a random variable X is

$$f_X(x) = \begin{cases} 2x & ; 0 < x < 1 \\ 0 & ; \text{other wise} \end{cases}$$

Use the CDF method to find the probability density function of the random variable  $Y = 8X^3$ .

#### Solution

- $F_X(x) = \int_0^x 2x \, dx = x^2 \Big|_0^x = x^2$
- The rang of y:  $0 < x < 1 \Rightarrow 0 < x^3 < 1 \Rightarrow 0 < 8x^3 < 8 \Rightarrow 0 < y < 8$ .
- $F_Y(y) = P(Y \leq y) = P(8X^3 \leq y) = P\left(X^3 \leq \frac{y}{8}\right) = P\left(X \leq \frac{1}{2}y^{\frac{1}{3}}\right) = F_X\left(\frac{1}{2}y^{\frac{1}{3}}\right)$   
 $= \left(\frac{1}{2}y^{\frac{1}{3}}\right)^2 = \frac{1}{4}y^{\frac{2}{3}}; \quad 0 < y < 8.$
- $f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{2}{12}y^{-\frac{1}{3}} = \frac{y^{-\frac{1}{3}}}{6}; \quad 0 < y < 8.$

#### Example 7.5

Let  $X \sim \text{Exp}(\theta)$  i.e.  $f_X(x) = \theta e^{-\theta x}; x \geq 0$ . Use the CDF method to find the distribution of the random variable  $Y = e^X$ .

#### Solution

- Since  $X \sim \text{Exp}(\theta)$  then  $F_X(x) = 1 - e^{-\theta x}$ .
- The rang of y:  $0 \leq x \leq \infty \Rightarrow e^0 \leq e^x \leq e^\infty \Rightarrow 1 \leq y \leq \infty$ .
- $F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(\ln(e^X) \leq \ln(y)) = P(X \leq \ln(y)) = F_X(\ln(y)) = 1 - e^{-\theta \ln(y)} = 1 - e^{\ln(y^{-\theta})} = 1 - y^{-\theta}; \quad 1 \leq y \leq \infty.$

$$f_y(y) = \frac{dF_y(y)}{dy} = \theta y^{-(\theta+1)}; 1 \leq y \leq \infty.$$

### Example 7.6

Let  $X \sim N(\mu, \sigma)$ . Use the CDF method to find the distribution of the random variable

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

### Solution

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

**Note** At this example we notice that it is difficult to compute  $F_X(x)$ , therefore, we will use the differentiation that  $\frac{d}{dy} F_Y(y) = f_Y(y)$ .

- The rang of z:  $-\infty < x < \infty \Rightarrow \frac{-\infty - \mu}{\sigma} < \frac{x - \mu}{\sigma} < \frac{\infty - \mu}{\sigma} \Rightarrow -\infty < z < \infty$ .
- $F_Z(z) = P(Z \leq z) = P\left(\frac{X - \mu}{\sigma} \leq z\right) = P(X \leq \sigma z + \mu) = F_X(\sigma z + \mu)$

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \frac{dF_X(\sigma z + \mu)}{dz} = \frac{dF_X(\sigma z + \mu)}{dx} \cdot \frac{dx}{dz}$$

$$= f_X(\sigma z + \mu) \cdot \sigma, \text{ (by using the chain rule) thus,}$$

$$f_Z(z) = \sigma \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\sigma z + \mu - \mu)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\sigma z)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}; -\infty \leq Z \leq \infty;$$

i.e.  $Z \sim N(0,1)$ .

## 7.2.2 Change-of-Variable Method

### 7.2.2.1 One Variable

#### Definition

Let  $X$  be a continuous random variable with probability density function  $f(x)$  defined over the rang  $c_1 < x < c_2$ , and, let  $Y = g(X)$  be an invertible function of  $X$  with inverse function  $X = g^{-1}(Y)$ . Then, using the change-of-variable technique, the probability density function of  $Y$  is

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{dg^{-1}(y)}{dy} \right|$$

defined over the rang  $g^{-1}(c_1) < y < g^{-1}(c_2)$ .

### Example 7.7

Use the change-of-variable method to find the distribution of the random variable  $Y$  in Example 7.4.

**Solution**

- $g^{-1}(y) = \frac{y^{\frac{1}{3}}}{2}$ .
- The rang of y:  $0 < x < 1 \Rightarrow 0 < x^3 < 1 \Rightarrow 0 < 8x^3 < 8 \Rightarrow 0 < y < 8$ .
- $f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{dg^{-1}(y)}{dy} \right| = f_X\left(\frac{y^{\frac{1}{3}}}{2}\right) \cdot \left| \frac{1}{6}y^{-\frac{2}{3}} \right| = 2\left(\frac{y^{\frac{1}{3}}}{2}\right) \cdot \frac{1}{6}y^{-\frac{2}{3}}$   
 $= \frac{1}{6}y^{-\frac{1}{3}}; \quad 0 < y < 8.$

**Example 7.8**

If  $X \sim \text{Uniform}(2,5)$ . Use the change-of-variable method to find the distribution of the random variable  $Y = \frac{X}{1+X}$ .

**Solution**

- $f_X(x) = \frac{1}{3}$ .
- $Y = \frac{X}{1+X} \Rightarrow Y + YX = X \Rightarrow Y = X - YX \Rightarrow Y = X(1 - Y) \Rightarrow X = \frac{Y}{1-Y}$ . Hence,  
 $g^{-1}(y) = \frac{Y}{1-Y}$ .
- The rang of y:  $2 < x < 5 \Rightarrow 3 < 1 + x < 6 \Rightarrow \frac{2}{3} < \frac{x}{1+x} < \frac{5}{6} \Rightarrow \frac{2}{3} < y < \frac{5}{6}$ .
- $f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{dg^{-1}(y)}{dy} \right| = f_X\left(\frac{y}{1-y}\right) \cdot \left| \frac{1}{(1-y)^2} \right| = \frac{1}{3}(1-y)^{-2}$   
 $= \frac{(1-y)^{-2}}{3}; \quad \frac{2}{3} < y < \frac{5}{6}.$

**Example 7.9**

Let  $X \sim \text{Exp}(\lambda)$  i.e.  $f_X(x) = \lambda e^{-\lambda x}; x \geq 0$ . Use the change-of-variable method to find the distribution of the random variable  $Y = X^{\frac{1}{\beta}}$ .

**Solution**

- $Y = X^{\frac{1}{\beta}} \Rightarrow X = Y^{\beta}$ . Hence,  $g^{-1}(y) = Y^{\beta}$ .
- The rang of y:  $0 \leq x \leq \infty \Rightarrow 0^{\frac{1}{\beta}} \leq x^{\frac{1}{\beta}} \leq \infty^{\frac{1}{\beta}} \Rightarrow 0 \leq y \leq \infty$ .
- $f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{dg^{-1}(y)}{dy} \right| = f_X(y^{\beta}) \cdot |\beta y^{\beta-1}| = \lambda \beta y^{\beta-1} e^{-\lambda y^{\beta}}; \quad y > 0, \beta > 0;$   
i.e.  $Y \sim \text{Weibull Distribution}$ .



### 7.2.2.2 Two Variables

#### Definition

Suppose the two continuous r.v.'s  $(X_1, X_2)$  has joint probability function  $f_{X_1 X_2}(x_1, x_2)$  and joint sample space  $\Omega_{X_1 X_2}$ . Let  $(Y_1, Y_2)$  be some function of  $(X_1, X_2)$  defined by  $Y_1 = g_1(X_1, X_2)$  and  $Y_2 = g_2(X_1, X_2)$  with the single-valued inverse given by  $X_1 = g_1^{-1}(Y_1, Y_2)$  and  $X_2 = g_2^{-1}(Y_1, Y_2)$ . Let  $\Omega_{Y_1 Y_2}$  be the sample space of  $Y_1, Y_2$ . Then, we usually find  $\Omega_{Y_1 Y_2}$  by considering the image of  $\Omega_{X_1 X_2}$  under the transformation  $(Y_1, Y_2)$ .

The joint pdf  $Y_1$  and  $Y_2$  is

$$f_{Y_1 Y_2}(y_1, y_2) = |J| f_{X_1 X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)),$$

where  $|J|$  refers to the absolute value of the Jacobian " $J$ " which is given by

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_2} \\ \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_2} \end{vmatrix}$$

#### Example 7.10

Let  $X_1$  and  $X_2$  are two independent random variables having exponential distributions with parameters  $\lambda_1$  and  $\lambda_2$  respectively. Find the distribution of  $Y_1 = X_1 + X_2$  and  $Y_2 = \frac{X_1}{X_1 + X_2}$ .

#### Solution

- $f_{X_1}(x_1) = \lambda_1 e^{-\lambda_1 x_1}$  ;  $x_1 \geq 0$  and  $f_{X_2}(x_2) = \lambda_2 e^{-\lambda_2 x_2}$  ;  $x_2 \geq 0$ .

Since  $X_1$  and  $X_2$  are independent, hence

$$f(x_1, x_2) = \lambda_1 e^{-\lambda_1 x_1} \cdot \lambda_2 e^{-\lambda_2 x_2} = \lambda_1 \lambda_2 e^{-(\lambda_1 x_1 + \lambda_2 x_2)}$$

- $Y_1 = X_1 + X_2$  &  $Y_2 = \frac{X_1}{X_1 + X_2} \Rightarrow X_1 = Y_1 Y_2 \Rightarrow X_2 = Y_1 - Y_1 Y_2 = Y_1(1 - Y_2)$ .

Hence,  $g_1^{-1}(y_1, y_2) = y_1 y_2$  &  $g_2^{-1}(y_1, y_2) = y_1(1 - y_2)$ .

- The rang of  $y_1$  &  $y_2$ :  $0 \leq \frac{x_1}{x_1 + x_2} \leq 1 \Rightarrow 0 \leq y_2 \leq 1$ ;

$$0 \leq y_1 y_2 \leq \infty \Rightarrow 0 \leq y_1 \leq \infty.$$

- $J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1 y_2 - y_1(1 - y_2) = -y_1$ .

- $f_{Y_1 Y_2}(y_1, y_2) = |J| f_{X_1 X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) = |-y_1| f(y_1 y_2, y_1(1 - y_2))$

$$\begin{aligned}
&= \lambda_1 \lambda_2 y_1 e^{-(\lambda_1 y_1 y_2 + \lambda_2 y_1 (1-y_2))} \\
&= \lambda_1 \lambda_2 y_1 e^{-[(\lambda_1 - \lambda_2) y_1 y_2 + \lambda_2 y_1]} ; 0 \leq y_1, 0 \leq y_2 \leq 1.
\end{aligned}$$

**Example 7.11**

Let  $X_1$  and  $X_2$  are two independent random variables having Gamma distributions with parameters  $\alpha = 2$  and  $\beta = 1$ . Find the distribution of  $X_1 + X_2$ .

**Solution**

- $f_{X_1}(x_1) = x_1 e^{-x_1} ; x_1 \geq 0$  and  $f_{X_2}(x_2) = x_2 e^{-x_2} ; x_2 \geq 0$ .

Since  $X_1$  and  $X_2$  are independent, hence

$$f(x_1, x_2) = x_1 e^{-x_1} \cdot x_2 e^{-x_2} = x_1 x_2 e^{-(x_1 + x_2)}.$$

- Let  $Y_1 = X_1$  &  $Y_2 = X_1 + X_2 \Rightarrow X_1 = Y_1$ .

Also,  $X_2 = Y_2 - Y_1$ .

Hence,  $g_1^{-1}(y_1, y_2) = y_1$  &  $g_2^{-1}(y_1, y_2) = y_2 - y_1$ .

- The rang of  $y_1$  &  $y_2$ :  $0 \leq x_1 = y_1 \leq \infty$  &  $0 \leq x_2 = y_2 - y_1 \leq \infty \Rightarrow 0 \leq y_1 \leq y_2$ .

- $J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1 - 0 = 1.$

- $f_{Y_1 Y_2}(y_1, y_2) = |J| f(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) = |1| f(y_1, y_2 - y_1)$   
 $= y_1 (y_2 - y_1) e^{-(y_1 + y_2 - y_1)} = y_1 (y_2 - y_1) e^{-y_2} ; y_2 \geq y_1 \geq 0.$

- $f_{Y_2}(y_2) = \int_0^{y_2} y_1 (y_2 - y_1) e^{-y_2} dy_1 = e^{-y_2} \int_0^{y_2} (y_1 y_2 - y_1^2) dy_1$   
 $= e^{-y_2} \left( \frac{y_1^2 y_2}{2} - \frac{y_1^3}{3} \right) \Big|_0^{y_2} = e^{-y_2} \left[ \left( \frac{y_2^3}{2} - \frac{y_2^3}{3} \right) - 0 \right] = \frac{y_2^3}{6} e^{-y_2} ; y_2 \geq 0,$

i.e.  $Y_2 = X_1 + X_2 \sim \text{Gamma}(4, 1)$ .

**7.2.3 Moment-Generating Function Method**

Let  $X$  &  $Y$  are two random variables where  $M_X(t)$ ,  $M_Y(t)$  exist and equal, then, depending on the uniqueness of the moment generating function of a random variable  $X$  and  $Y$  have the same distribution.

**Properties (From chapter two)**

- If  $Y = aX + b$ , then  $M_Y(t) = e^{bt}M_X(at)$ .
- If  $X$  and  $Y$  are two independent r.v's, then  $M_{X+Y}(t) = M_X(t)M_Y(t)$ .

**Example 7.12 (Sum of Independent Gammas).**

Let  $X_i \sim \text{Gamma}(\alpha_i, \beta)$   $i = 1, \dots, n$ . Independent random variables. Use the moment generating function to find the distribution of  $Y = \sum_{i=1}^n X_i$ .

**Solution**

- $M_{X_i}(t) = \left(\frac{\beta}{\beta-t}\right)^{\alpha_i}, i = 1, \dots, n$ .
- $M_Y(t) = M_{\sum_{i=1}^n X_i}(t) = E(e^{t(X_1+\dots+X_n)}) = E(e^{tX_1} \dots e^{tX_n}) = E(e^{tX_1}) \dots E(e^{tX_n})$   

$$= M_{X_1}(t) \dots M_{X_n}(t) = \left(\frac{\beta}{\beta-t}\right)^{\alpha_1} \dots \left(\frac{\beta}{\beta-t}\right)^{\alpha_n} = \left(\frac{\beta}{\beta-t}\right)^{\sum_{i=1}^n \alpha_i}$$

Thus,  $Y = \sum_{i=1}^n X_i \sim \text{Gamma}(\sum_{i=1}^n \alpha_i, \beta)$ .

**Example 7.13 (Linear Function of Independent Normal r.v's).**

Let  $X_i \sim N(\mu_i, \sigma_i^2); i = 1, \dots, n$ . Independent random variables. Use the moment generating function to find the distribution of  $Y = \sum_{i=1}^n a_i X_i$ .

**Solution**

- $M_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}, i = 1, \dots, n$ .
- $M_Y(t) = M_{\sum_{i=1}^n a_i X_i}(t) = E(e^{t(a_1 X_1 + \dots + a_n X_n)}) = E(e^{ta_1 X_1} \dots e^{ta_n X_n})$   

$$= E(e^{ta_1 X_1}) \dots E(e^{ta_n X_n}) = M_{X_1}(a_1 t) \dots M_{X_n}(a_n t)$$
  

$$= e^{\mu_1 a_1 t + \frac{\sigma_1^2 a_1^2 t^2}{2}} \dots e^{\mu_n a_n t + \frac{\sigma_n^2 a_n^2 t^2}{2}} = e^{t \sum_{i=1}^n \mu_i a_i + \frac{t^2}{2} \sum_{i=1}^n \sigma_i^2 a_i^2}.$$

Thus,  $Y = \sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n \mu_i a_i, \sqrt{\sum_{i=1}^n \sigma_i^2 a_i^2}\right)$ .

**Example 7.14**

Use the moment generating function to find the distribution of  $Z^2$  where  $Z \sim N(0,1)$ .

**Solution**

- $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}; -\infty < z < \infty$ .
- $M_{Z^2}(t) = E(e^{tZ^2}) = \int_{-\infty}^{\infty} e^{tZ^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{1-2t}{2}\right)z^2} dz$

$$= \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-\left(\frac{1-2t}{2}\right)z^2} dz \quad (\text{Symmetric about } 0)$$

Let  $u = z^2 \Rightarrow z = u^{1/2}$ ,  $dz = \frac{1}{2}u^{-\frac{1}{2}} du$  and  $0 \leq u$ . Hence

$$\begin{aligned} M_{Z^2}(t) &= \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-\left(\frac{1-2t}{2}\right)u} \cdot \frac{1}{2}u^{-\frac{1}{2}} du = \frac{1}{\sqrt{2\pi}} \int_0^\infty u^{-\frac{1}{2}} e^{-\left(\frac{1-2t}{2}\right)u} du \\ &= \frac{\Gamma\left(\frac{1}{2}\right)\left(\frac{2}{1-2t}\right)^{\frac{1}{2}}}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\Gamma\left(\frac{1}{2}\right)\left(\frac{2}{1-2t}\right)^{\frac{1}{2}}} u^{-\frac{1}{2}} e^{-\left(\frac{1-2t}{2}\right)u} du = \frac{\Gamma\left(\frac{1}{2}\right)\left(\frac{2}{1-2t}\right)^{\frac{1}{2}}}{\sqrt{2\pi}} \cdot (1) = \frac{\sqrt{\pi}\left(\frac{2}{1-2t}\right)^{\frac{1}{2}}}{\sqrt{2\pi}} \\ &= (1-2t)^{-\frac{1}{2}}. \end{aligned}$$

Thus,  $Z^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$ .

### Example 7.15 (Sum of two exponential r.v's).

Let  $X_1, X_2$  are two independent random variables have the same exponential distribution with parameter  $\theta$ .  $f_{X_i}(x_i) = \theta e^{-\theta x_i}$ ,  $x_i \geq 0$ . Use the moment generating function to find the distribution of  $X_1 + X_2$ .

#### Solution

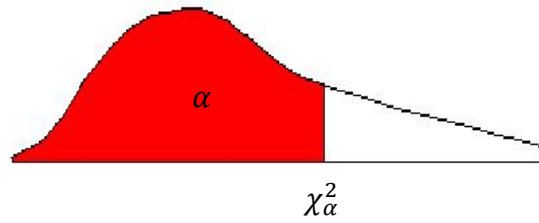
- $M_{X_i}(t) = \frac{\theta}{\theta-t}$ .
- $M_U(t) = M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t) = \frac{\theta}{\theta-t} \cdot \frac{\theta}{\theta-t} = \left(\frac{\theta}{\theta-t}\right)^2$ .

Thus,  $X_1 + X_2 \sim \text{Gamma}(2, \theta)$ .

**Table 4.1:** The Incomplete Gamma Function:  $F(x; \alpha) = \int_0^x \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy$

$x$	$\alpha$									
	1	2	3	4	5	6	7	8	9	10
1	0.6320	0.2640	0.0800	0.0190	0.0040	0.0010	0.0000	0.0000	0.0000	0.0000
2	0.8650	0.5940	0.3230	0.1430	0.0530	0.0170	0.0050	0.0010	0.0000	0.0000
3	0.9500	0.8010	0.5770	0.3530	0.1850	0.0840	0.0340	0.0120	0.0040	0.0010
4	0.9820	0.9080	0.7620	0.5670	0.3710	0.2150	0.1110	0.0510	0.0210	0.0080
5	0.9930	0.9600	0.8750	0.7350	0.5600	0.3840	0.2380	0.1330	0.0680	0.0320
6	0.9980	0.9830	0.9380	0.8490	0.7150	0.5540	0.3940	0.2560	0.1530	0.0840
7	0.9990	0.9930	0.9700	0.9180	0.8270	0.6990	0.5500	0.4010	0.2710	0.1700
8	1.0000	0.9970	0.9860	0.9580	0.9000	0.8090	0.6870	0.5470	0.4070	0.2830
9		0.9990	0.9940	0.9790	0.9450	0.8840	0.7930	0.6760	0.5440	0.4130
10		1.0000	0.9970	0.9900	0.9710	0.9330	0.8700	0.7800	0.6670	0.5420
11			0.9990	0.9950	0.9850	0.9620	0.9210	0.8570	0.7680	0.6590
12			1.0000	0.9980	0.9920	0.9800	0.9540	0.9110	0.8450	0.7580
13				0.9990	0.9960	0.9890	0.9740	0.9460	0.9000	0.8340
14				1.0000	0.9980	0.9940	0.9860	0.9680	0.9380	0.8910
15					0.9990	0.9970	0.9920	0.9820	0.9630	0.9300

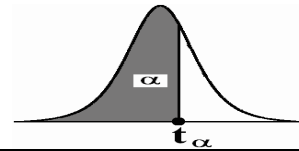
## Appendix A.

**Table 4.2:** Lower Critical values for Chi-square distribution

$df \backslash \alpha$	.005	.01	.025	.05	.10	.90	.95	.975	.99	.995
1	.00004	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.6	5.23	6.26	7.26	8.55	22.31	25	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

**Table 4.3:** Area to the Left of the Z score for Standard Normal Distribution.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

**Table 4.4:** Lower Critical Values for t Distribution.

$v = df$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
35	1.3062	1.6896	2.0301	2.4377	2.7238
40	1.3030	1.6840	2.0210	2.4230	2.7040
45	1.3006	1.6794	2.0141	2.4121	2.6896
50	1.2987	1.6759	2.0086	2.4033	2.6778
60	1.2958	1.6706	2.0003	2.3901	2.6603
70	1.2938	1.6669	1.9944	2.3808	2.6479
80	1.2922	1.6641	1.9901	2.3739	2.6387
90	1.2910	1.6620	1.9867	2.3685	2.6316
100	1.2901	1.6602	1.9840	2.3642	2.6259
120	1.2886	1.6577	1.9799	2.3578	2.6174
140	1.2876	1.6558	1.9771	2.3533	2.6114
160	1.2869	1.6544	1.9749	2.3499	2.6069
180	1.2863	1.6534	1.9732	2.3472	2.6034
200	1.2858	1.6525	1.9719	2.3451	2.6006
$\infty$	1.282	1.645	1.960	2.326	2.576