



**NATIONAL OPEN UNIVERSITY OF NIGERIA**

**SCHOOL OF SCIENCES**

**COURSE CODE: STT 311**

**COURSE TITLE: PROBABILITY DISTRIBUTION 2**

# **STT 311 PROBABILITY DISTRIBUTION 2**

**Course Developer: MR A. ADELEKE.**

**Osun State College of Technology ESA- OKE**

**(LAGOS CENTRE)**

## **TABLE OF CONTENTS**

### **UNIT 1: PROBABILITY SPACES, MEASURE AND DISTRIBUTION**

- 1.0 Introduction
- 1.1 Objective
- 1.2 Probability space
- 1.3 Sample space and Event
- 1.4 Probability Measure
- 1.5 Theorem on probability space
- 1.6 Probability distribution
- 1.7 Conclusion
- 1.8 Summary
- 1.9 Tutor Marked Assignment
- 1.91 References.

### **UNIT 2 : DISTRIBUTION OF RANDOM VARIABLE SPACES**

- 2.0 Introduction
- 2.1 Objective
- 2.2 Random Variables
- 2.3 Discrete Probability Distributions
- 2.4 Distribution Functions For Random Variables
- 2.5 Distribution Function For Discrete Random Variables.

- 2.6 Continuous Random Variables
- 2.7 Graphical Representation of Random Variables.
- 2.8 Joint Distributions
- 2.9 Independence of Random Variables
- 2.10 Conditional Distributions
- 2.11 Conclusion
- 2.12 Summary
- 2.13 Tutor Marked Assignment (TMA)
- 2.14 Reference / Further Reading / Other References.

### **UNIT 3 : EXPECTATION OF RANDOM VARIABLES.**

- 3.0 Introduction
- 3.1 Objectives
- 3.2 What is Expectation of Random Variables?
- 3.3 Theorems on Expectation
- 3.4 The Variance and Standard Deviation.
- 3.5 Theorems on Variance
- 3.6 Moments
- 3.7 Moments Generating Functions
- 3.8 Theorems on Moment Generating Functions.
- 3.9 Characteristics Function.

3.10 Conclusion

3.11 Summary

3.12 Tutor Marked Assignment (TMA)

3.13 References / Further Reaching / Other Resources.

#### **UNIT 4 : LIMIT THEOREM.**

4.0 Introduction

4.1 Objectives

4.2 Chebyshev's Inequality

4.3 Convergence of Random Variables

4.4 Demovre's Theorem

4.5 Central Limit Theorem

4.6 Klinchine's Theorem

4.7 Conclusion

4.8 Summary

4.9 Tutor Marked Assignment ( TMA)

5.0 References / Further Reading / Other Resources.

## **UNIT 1: PROBABILITY SPACES, MEASURE AND DISTRIBUTION**

### **1.0 INTRODUCTION**

This Unit Focuses on Probability spaces, probability measures and probability distribution for continuous random variables. It gives some basic definition and relevant working examples will be given to make the concept more meaningful for the learners.

### **1.1 OBJECTIVES.**

At the end of this unit, student should be able to:

- Understand the meaning of probability space and its notation.
- Define Sample space and event, and Event Space.
- Discuss Probability Measure and State its Theorems.
- Discuss Probability Distribution for Continuous Random Variables.

### **1.2 PROBABILITY SPACE**

A Probability Space is a triplet Finite measure space  $(\Omega, \mathcal{A}, P[ \cdot ])$  Where  $\Omega$  is a sample space and each  $w \in \Omega$  is called a sample point, and  $P[ \cdot ]$  is a function that has  $\mathcal{A}$  as its domain that is a single term that gives us an expedient way to assume the existence of all three components in its notation.

### 1.3 SAMPLE SPACE AND EVENT

Definition of Sample Space :

The sample space denoted by  $\Omega$ , is the collection or totality of all possible outcomes of a conceptual experiment.

In addition to  $\Omega$ ,  $S$ ,  $Z$ ,  $R$ ,  $E$ ,  $\mu$  and  $A$  are the other symbols to denote sample space.

Event: An event is a subset of the sample space.

Event Space: The class of all events associated with a given experiment is defined to be the event spaces.

### 1.4 PROBABILITY MEASURE

A probability measure is a normed non-negative, countable additive set function defined on the field of all events.

Definition:- A probability measure  $P$  on a  $\sigma$ -field of subset  $A$  of set  $\Omega$  is a real -valued function having domain  $A$  satisfying the following properties

(i).  $P(\Omega) = 1$

(ii).  $P(A) \geq 0$  for all  $A \in \Delta$

(iii) If  $A_n, n = 1, 2, \dots$  are mutually disjoint sets in  $\Lambda$ ,  
then  $P(UA_n) = \sum P(A_n)$ .

A probability space, denoted by  $(\Omega, \mathcal{A}, P)$ .

$P_B(A)$  is a conditional probability measure on  $\Omega$

We say that  $(\Omega, \mathcal{A}, P_B)$  is the probability space obtained by conditioning  $(\Omega, \mathcal{A}, P)$  by the event  $B$ .

If an event  $B$  depends on occurrence of event  $A_1$  or  $A_2$  then

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) \\ &= P(A_1)P(B/A_1) + P(A_2)P(B/A_2) \end{aligned}$$

In general, if an event  $B$  depends on occurrence of events  $A_1, A_2, \dots, A_n$ , then

$$\begin{aligned} P(B) &= \sum_i P(B \cap A_i) \\ &= \sum_i P(A_i)P(B/A_i) \end{aligned}$$

1.5 Theorem: let  $(\Omega, \mathcal{A}, P)$  be a probability space then

$(\Omega, \mathcal{A}, P_B)$  is also a probability space.

Proof:

$$P_B(A) = \frac{P(A \cap B)}{P(B)} \geq 0;$$

$$P_B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = 1$$



Let  $A_1, A_2, \dots$  be disjoint event in  $\Omega$ , then

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) &= \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} \\ &= \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} \end{aligned}$$

Since  $(A_i \cap B), i = 1, 2, \dots$  are disjoint events, we have

$$P_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P_B(A_i) \text{ and so } P_B \text{ is countably additive.}$$

Hence  $(\Omega, \mathcal{F}, P_B)$  is probability space

Lemma

Let  $\{A_n\}$  be a sequence of independent measurable sets

- (i) If  $\sum_{n=1}^{\infty} P(A_n) < \infty$  then  $P(A^c) = 0$
- (ii) If  $\sum_{n=1}^{\infty} P(A_n) = \infty$  then  $P(A^c) = 1$

where  $A^c = \text{Lim sup } A_n$

Proof :

$$(i) \quad A^c = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

Since  $\bigcup_{m=n}^{\infty} A_m$  decrease as  $n$  increases, we have

$$P(A^c) \leq P\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} P(A_m)$$

Since  $\sum P_n < \infty$ , then

$$\lim_{n \rightarrow \infty} \sum_{m=n}^{\infty} P(A_m) = 0$$

$$\begin{aligned} \text{(iii)} \quad P(A^c) &= \left\{ \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c \right\} \leq \sum P \left( \bigcap_{m=n}^{\infty} A_m^c \right) \\ &= \sum_{n=1}^{\infty} \prod_{m=n}^{\infty} (1 - P(A_m)) = 0 \end{aligned}$$

Hence

$$P(A) = 1$$

## 1.6 PROBABILITY DISTRIBUTION

**Definition:-** Let  $x$  be a random variable whose image set (s) is a continuous numbers such as an interval. Then the set

$a \leq x \leq b$  is an event in  $S$ .

The probability  $P(a < x < b)$  defined as

$$P(a < x < b) = \int_a^b f(x) dx$$

Is called continuous random variable, the function  $f$  is called the distribution or continuous probability function or density function of  $X$  and it satisfies the following

- i.  $f(x) > 0$
- ii.  $\int_R f(x) dx = 1$  where  $R = (a, b)$

iii.  $f(x)$  is a non - decreasing function

$$\text{iv. } \int_a^b f(x) dx = f(b) - f(a) = p \quad (a < x < b)$$

for example, if  $x$  is a random variable defined to take any value in interval  $(0,1)$  if a point is taken in this interval say 0.45, the probability that a point picked is

$$\begin{aligned} P(x = 0.45) &= \frac{1}{\text{Uncountable points}} \\ &= \frac{1}{\text{No of points btw the interval}} \\ &= \frac{1}{\infty} = 0 \end{aligned}$$

But it is easier to find the probability of a sub-interval within an interval, it is possible to calculate per  $(x < 0.45)$  etc.

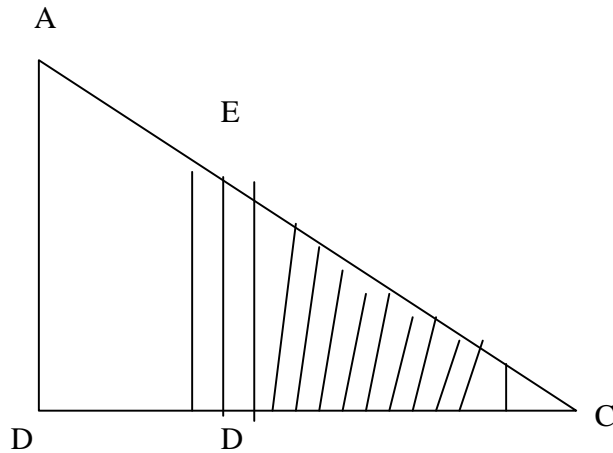
The  $\text{pr}(x < 0.45) = \text{pr}(\text{lenght } 0 \text{ to } 0.45)$

$$= \frac{\text{Ratio of the given Lenght}}{\text{Total lenght}}$$

**Similarly**

$$\text{Pr}(x_1 < x < x_2) = \frac{x_2 - x_1}{\text{Total Lenght}}$$

Example 1.6.1 :- Given the figure below



Find the probability of the shaded portion.

Solution:

$$\Pr(\text{Shaded portion}) = \frac{\text{Area of } \Delta CDE}{\text{Area of } \Delta ABC}$$

Example 1.6.2 :- The length of life measure in hours of a certain rare type of insect is a random, reliable  $x$  with portability density function

$$f(x) = \begin{cases} \frac{3}{4} (2x - x^2) & 0 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

If the amount of food measured in milligrams consumed in a life time by such an insect defined by the function  $g(x) = x^2$ ,

, where  $x$  is the length of life measured in hours, find the expected amount of food that will be consumed by an insect of this type.

### Solution

$$\text{Expected amount of food} = E[g(x)]$$

$$= \int_0^2 g(x) f(x) dx$$

$$= \int_0^2 x^2 \left( \frac{3}{4} (2x - x^2) \right) dx$$

$$= 1.2 \text{ mg}$$

Example 1.5.3:- Given a continuous random variable with the probability distribution function

$$f(x) = \int kx^2 \quad 0 \leq x \leq 10$$

0 elsewhere

Find k such that f(x) is then Pdf

### Solution

If F(x) is a Pdf,  $\int_0^1 f(x) dx = 1$

$$\int f(x) dx = \int_0^{10} K x^2 dx$$

$$= \frac{Kx^3}{3} \Big|_0^{10} = k \frac{(10)^3}{3} - k \frac{(0)^3}{3} = 1$$

$$= \frac{k}{3} (10^3 - 0^0) = 1$$

$$\frac{1000K}{3} = 1$$

$$k = \frac{3}{1000}$$

## 1.7 CONCLUSION

In this unit, you have learnt probability space, the notation of its components, the definition of sample space, event and event space.

You also learned probability measure and its main probabilities related theorems. You also learned probability distribution of a continuous random variables with relevant working examples.

## 1.8 SUMMARY

What you have learned in this unit are the following probability distribution concepts.

- i. The meaning of probability space and its notation
- ii. Important definition of sample space, Event and event space .
- iii. Probability measure and its properties
- iv. Part of probability distribution of a continuous random variables

### EXERCISE 1.80.1 (SAE)

The surface area measured in squaremeter of a flat metal disk manufactured by a certain process is a random variable  $x$  with probability density function.

$$f(x) = \int_0^x 6(x - x^2) \quad 0 < x < 1$$

*discuhere*

Find the expected radius measured in meter of a flat metal disk manufactured by this process.

### 1.90. TUTOR MARKED ASSIGNMENT (TMA)

Exercise 1.8.1: The probability function of a random variable  $x$  is given by

$$f(x) = \begin{cases} 2p & x = 1 \\ p & x = 2 \\ 4p & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Where  $P$  is a constant find

(a)  $P(0 \leq x < 3)$ , (b)  $P(x > 1)$ .

## **1.9 REFERENCES / FURTHER READING**

DR. R.A Kasumu (2003) probability theory first edition published by Fatol ventures Lagos.

Alexander M. Mood et al 1974 introduction to the theory of statistics third edition published by McGraw –Hill.



## **UNIT 2: DISTRIBUTION OF RANDOM VARIABLES SPACES**

### **2.0 INTRODUCTION**

This unit concerns with the meaning and classification of random variables into discrete and continuous random variables are high lighted, distributive functions for discrete and continuous random variables and related examples are also given.

The unit further high light Graphical representation, Joint distribution for discrete and continuous random variable , Independence and Conditional probability of random variables and working example on each are given.

### **2.1 OBJECTIVE**

At the end of this unit student should be able to,

- Understand the meaning of random variables.
- Classify random variables into discrete and continuous random variable with example.
- Define and state the properties of distribution function.
- State the distribution function for discrete and continuous random variables and solve example on each.
- Show the graphical representation of random variables.

- State the joint distributions for two random variables which are either both discrete or both continuous.
- State the independent of random variables for independent and dependent events.
- State the conditional probability function for discrete and continuous random variables.
- Solve related problems on the distribution of random variables spaces.

## 2.2 RANDOM VARIABLES

A random variable is a function whose domain of definition is the simple space  $S$  of a random experiment and whose range is a set of real numbers.

### Definition 2

A real valued measurable function

$$X : \Omega \rightarrow R \text{ with respect to } (\Omega, \mathcal{A}, P)$$

Is called a random variables.

Note:

Suppose  $\{X_n\}$  is a sequence of random. Variable

If  $\lim_{n \rightarrow \infty} X_n = X$ , then  $X$  is Random variable

Example : suppose that a coin is tossed twice so that the sample space  $S = (HH, HT, TH, TT)$ . let  $X$  represent the number of heads that can come up. For example  $X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0$

Since the domain of  $X$  is  $S$  and the range consists of real numbers, then  $x$  is a random variable.

A random variable that takes on a finite or countably infinite number of values is called a discrete random variable. While the one which takes a non countable infinite number of values is called a nondiscrete / continuous random variable.

### 2.3 DISCRETE PROBABILITY DISTRIBUTIONS

Let  $X$  be a discrete random variable, and suppose that the possible values that it can assume are given by  $x_1, x_2, x_3, \dots$ , arranged in some order. Suppose also that these values are assumed with probabilities given by

$$p(X = x_k) = f(x_k) \quad K = 1, 2, \dots \dots \dots (1)$$

It is convenient to introduce the probability function, also referred to as probability distribution, given by

$$P(X = x) = f(x) \dots \dots \dots (2)$$

For  $X = x_k$ , this reduces to equation given above while for other values of  $X_1$   $f(x) = 0$ .

In general,  $f(x)$  is a probability function if

1.  $f(x) \geq 0$

2.  $\sum_x f(x) = 1$

Where the sum in equation (2) is taken over all possible values of  $x$

## 2.4 DISTRIBUTION FUNCTION FOR RANDOM VARIABLES

The cumulative distribution function, or briefly distribution function for a random variables  $X$  is defined by

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

Where  $x$  is any real number, that is

$$-\infty < x < \infty$$

The distribution function  $F(x)$  has the following properties:

1.  $f(x)$  is non decreasing [ie,  $f(x) \leq f(y)$  if  $x \leq y$ ]

2.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ;  $\lim_{x \rightarrow \infty} f(x) = 1$

3.  $f(X)$  is continuous from the right  $\left[ \text{ie } \lim_{h \rightarrow 0^+} f(x+h) = f(x) \text{ for all } x \right]$ .

## 2.5 DISTRIBUTION FUNCTION FOR DISCRETE RANDOM VARIABLES

The distribution function for a discrete random variable  $x$  can be obtained from its probability function by noting that, for all

$$p(x) = P(X \leq x) = \sum_{v \leq x} f(v)$$

Where the sum is taken over all values  $\mu$  taken on by  $X$  for which  $\mu \leq x$  if  $X$  takes on only a finite numbers of values  $x_1, x_2, \dots, x_n$ , then the distribution function is given by

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ [f(x_1) + f(x_2)] & x_2 \leq x < x_3 \\ \dots \\ [f(x_1) + \dots + f(x_n)] & x_n \leq x < \infty \end{cases}$$

Example 2.5.1:- find the probability function corresponding to the random variable  $x$  when a coin is tossed twice; assuming that the coin is fair.

**Solution**

$$p(HH) = \frac{1}{4}, \quad P(HT) = \frac{1}{4}, \quad P(TH) = \frac{1}{4}$$

$$\text{and } P(TT) = \frac{1}{4}$$

$$\begin{aligned} \text{Then, } P(X = 0) &= P(TT) = \frac{1}{4} \\ P(X = 1) &= P(HT) + P(TH) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$

$$P(X = 2) = P(HH) = \frac{1}{4}$$

The probability function is given in the table below

x	0	1	2
f(x)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

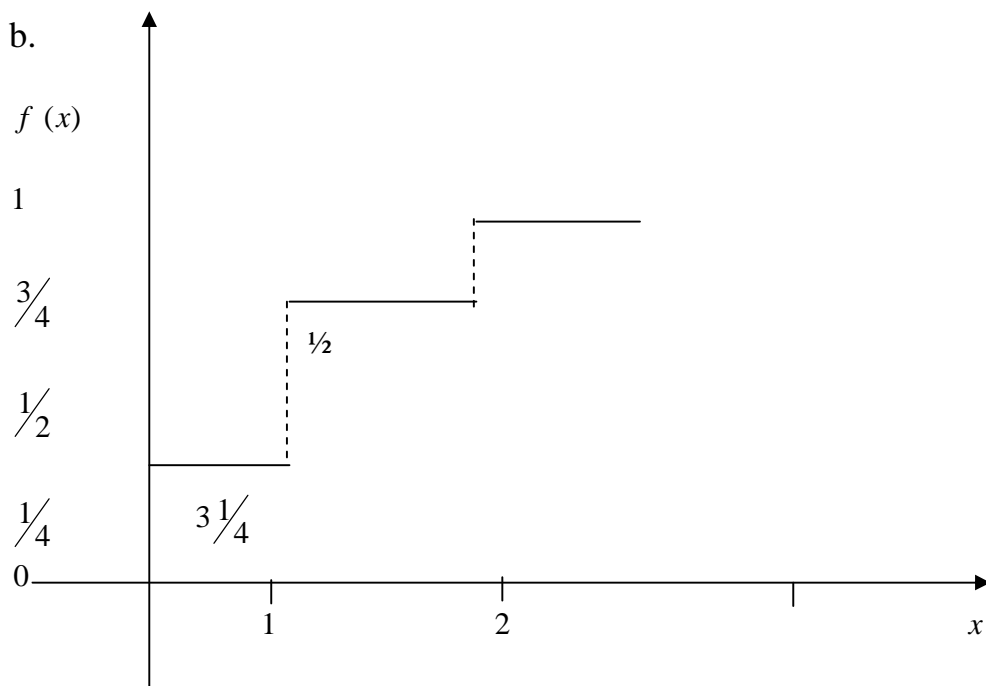
Example 2.5.2 = (a) find the distribution for the random variable X from the working example above (b) obtain its graph

**Solution:**

The distribution function is

$$f(x) = \left\{ \begin{array}{ll} 0 & -\infty < x < 0 \text{ since the table} \\ \frac{1}{4} & 0 \leq x < 1 \text{ since } f(x_1) = \frac{1}{4} \\ \frac{3}{4} & 1 \leq x < 2 \text{ since } f(x_1) + f(x_2) = \frac{3}{4} \\ 1 & 2 \leq x < \infty \text{ since } \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1 \end{array} \right\}$$

$$f(\eta) \left\{ \begin{array}{l} 0 \quad -\infty < x < 0 \quad \text{since } x_1 = 0 \text{ from the above table} \\ \frac{1}{4} \quad 0 \leq x < 1 \quad \text{since } f(x_1) = \frac{1}{4} \\ \frac{3}{4} \quad 1 \leq x < 2 \quad \text{since } f(x_1) + f(x_2) = \frac{3}{4} \\ \quad \quad \quad \text{ie } \frac{1}{4} + \frac{1}{2} = \frac{3}{4} \\ 1 \quad 2 \leq x < \infty \quad \text{since } \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1 \end{array} \right\}$$



The following things about the above distribution function should be noted.

1. The magnitudes of the jumps at 0, 1, 2 are  $\frac{1}{4}$ ,  $\frac{1}{2}$ ,  $\frac{1}{4}$  which are precisely the probabilities ie this fact enables one to obtain the probability function from the distribution function.

2. Because of the appearance of the graph it is often called a staircase function or step function, and the value at the function at an integer is obtained from the higher step; thus the value at 1 is  $\frac{3}{4}$  and not  $\frac{1}{4}$ . This is expressed mathematically by stating that the distribution function is continuous from the right at 0,1,2.
3. As we proceed from left to right in the distribution function is monotonically increasing function.

## 2.6 CONTINUOUS RANDOM VARIABLES

A non discrete random variable  $x$  is said to be absolutely continuous, or simply continuous, if its distribution function may be represented as

$$f(x) = P(X \leq x) = \int_{-\infty}^x f(u) du \quad (-\infty < x < \infty)$$

Where the function  $f(x)$  has the following properties

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

It follows from the above that if  $X$  is a continuous random variable, then, the probability that  $X$  takes on only one particular value is zero, whereas the interval probability that  $X$  lies between two different values say  $a$  and  $b$  is given by



$$p(a < x < b) = \int_a^b f(x) dx$$

Example 2.6.1 (a) find the constant  $C$  such that the function

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

is a density function, and

b. Compute  $P(1 < x < 2)$

### Solution

Since  $f(x)$  stratifies property (2) if  $C \geq 0$ , it must satisfy property 2 in order to be a density function

$$\begin{aligned} \text{Now, } \int_{-\infty}^{\infty} f(x) dx &= \int_0^3 cx^2 dx \\ &= \left. \frac{Cx^3}{3} \right|_0^3 \\ &= \frac{C(3)}{3} - C \frac{(0)^3}{3} \\ &= \frac{27C}{3} - \frac{0}{3} = 9C \end{aligned}$$

And since the integral equal to 1, we have  $9C = 1$

$$C = 1/9$$

$$\text{b. } P(1 < x < 2) = \int_1^2 \frac{1}{9} x^2 dx$$

$$\begin{aligned}
&= \frac{x^3}{27} \Big|_1^2 = \frac{(2)^3}{27} - \frac{(1)^3}{27} \\
&= \frac{8}{27} - \frac{1}{27} = \frac{7}{27}
\end{aligned}$$

## 2.7 GRAPHICAL REPRESENTATIONS OF RANDOM VARIABLE

If  $f(x)$  is the density function for random variable  $x$ , then we can represent  $y = f(x)$  graphically by a curve as shown in the figure below.

Since  $f(x) \geq 0$ , the curve cannot fall below the x-axis the entire area bounded by the curve and the x-axis must be 1 because of the second property i.e  $\int_{-\infty}^{\infty} f(x)dx = 1$ .

Geometrically the probability that  $x$  is between  $a$  and  $b$ , i.e  $p(a < x < b)$ , is the represented by the area shown shaded from the first figure below

The distribution function  $F(x) = P(X \leq x)$  is a monotonically increasing function which increases from 0 to a 1 and is represented by a curve as in the second figure.

## 2.8 JOINT DISTRIBUTIONS

Joint distributions can easily be generalized to two more random variables.

We shall consider the typical case of two random variables that are either both discrete or both continuous.

Discrete case: - If  $x$  and  $Y$  are two discrete random variables, we define the joint probability function of  $x$  and  $y$  by

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

Where (1)  $f(x, y) \geq 0$

$$(2) \sum_x \sum_y f(x, y) = 1$$

i.e the sum over all values of  $x$  and  $y$  is 1

Suppose that  $x$  can assume any one of  $m$  values  $x_1, x_2, \dots, x_m$  and  $y$  can assume any none of  $n$  values  $y_1, y_2, \dots, y_n$ .

Then the probability of the event that  $x = x_j$  and  $y = Y_j$  is given by

$$P(x = x_j, Y = y_k) = f(x_j, y_k)$$

A joint probability fraction for  $x$  and  $y$  can be represented by a joint probability table as shown below

$x/Y$	$Y_1$	$Y_2$	-----	$Y_X$	Total
$X_1$	$F(x_1, y_1)$	$F(x_1 y_2)$	-----	$F(x_1, y_n)_{x_1}$	$F_1(x_1, y_n)_{x_1}$
$X_2$	$F(x_2 y_1)$	$F(x_2, y_1)$	-----	$F(x_2, y_n)$	$F_1(x_2)$
$X_m$	$F(x_m y_1)$	$F(x_m y_2)$		$F(x_m y_n)$	$F_1(x_m)$
Totals →	$f_2(y_1)$	$f_2(y_2)$	-----	$f_2(y_n)$	1

The probability that  $X = x_j$  is obtained by adding all entries in the row corresponding to  $x_j$  and is given by

$$P(x = \eta_j) = f_1(x_j) = \sum_{k=1}^2 f(x_j, y_k)$$

For  $J = 1, 2, \dots, m$ , these are indicated by the entry totals in the entrance right hand column or margin from the table above similarly the probability that  $Y = y_k$  is obtained by adding all entries in the column corresponding to  $y_k$  and is given by

$$p(Y = y_k) = f_2(y_k) = \sum_{j=1}^m f(x_j, y_k)$$

For  $k = 1, 2, \dots, n$ , these are indicated by the entry totals in the bottom row or margin of the probability table from the two equations given above  $f_1(x_j)$  and  $f_2(y_k)$  or simply  $f_1(x)$  and  $f_2(y)$

which are obtained from the margin of the table are referred to as the marginal functions of X and y, respectively it should be noted that

$$\sum_{j=1}^m f_1(x_j) = 1 \quad \sum_{k=1}^n f_2(y_k) = 1$$

Which can be written as follow

$$\sum_{j=1}^m \sum_{k=1}^n f(x_j, y_k) = 1$$

This is simply the statement that the total probability of all entries is 1. The grand total of 1 is indicated in the lower right-hand of the probability table.

The joint distribution function of x and y is defined by

$$f(x, y) = P(x \leq x, y \leq y) = \sum_{\mu \leq x} \sum_{\nu \leq y} f(x_\mu, y_\nu)$$

In the probability table  $f(x, y)$  is the sum of all entries for which  $x_j \leq x$  and  $y_k \leq y$ .

### CONTINUOUS CASE:

The case where both variables are continuous is obtained easily by analogy with the discrete case on replacing sums by integrals thus the joint probability function or joint density function for random variables x and y is defined by

$$(1) f(x, y) \geq 0$$

$$(2) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(x, y) dx dy = 1$$

Graphically  $Z = f(x, y)$  represents a surface called the probability surface as indicated in the figure below. The total volume bounded by this surface and the  $xy$  plane is equal to 1 in accordance with property 2 above. The probability that  $x$  lies between  $a$  and  $b$  while  $y$  lies between  $c$  and  $d$  is given graphically by the shaded values of the figure below and mathematically by

$$P(a < x < b, c < y < d) = \int_{x=a}^b \int_{y=c}^d f(x, y) dx dy$$

More generally, if A represents any event, there will be a region  $R_A$  of the  $x, y$  plan that corresponds to it. In such case we can find the probability of A by performing the integration over  $R_A$  i.e

$$P(A) = \int_{R_A} \int f(x, y) dx dy$$

The joint distribution function of x and y in this case is defined by

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{\mu = -\infty}^x \int_{\nu = -\infty}^y f(u, v) du dv$$

It follows in analogy with equation

$$\frac{d F(x)}{d x} = f(x),$$

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y)$$

That is density function is obtained by differentiation the distribution function with respect to  $x$  and  $y$  from the joint distribution equation given above to obtain

$$P(x \leq x) = F_1(x) = \int_{\nu = -\infty}^x \int_{\nu = -\infty}^{\eta} f(u, v) dudv$$

$$P(y \leq y) = f_2(y) = \int_{\mu = -\infty}^{\infty} \int_{\nu = -\infty}^y f(u, v) dudv$$

The two equation above are called the marginal distribution functions or simply the distribution function of x and y respectively.



The derivative of the equations with respect to  $x$  and  $y$  are then called the marginal density functions or simply the density functions, of  $x$  and  $y$  which are given below

$$f_1(x) = \int_{v=-\infty}^{\infty} f(x,v)dv \quad f_2(y) = \int_{u=-\infty}^{\infty} f(u,y) du$$

## 2.9 INDEPENDENCE OF RANDOM VARIABLES

Suppose that  $x$  and  $y$  are discrete random variables. If the events  $X = x$  and  $Y = y$  are independent events for all  $x$  and  $y$ , then we say that  $x$  and  $y$  are independent random variables. In such case,

$$P(X = x, Y = y) = p(X = x) p(Y = y)$$

Or equivalently

$$f(x, y) = f_1(x) f_2(y)$$

Conversely, if for all  $x$  and  $y$  the joint probability function  $f(x, y)$  can be expressed as the product of a function of  $x$  alone and a function of  $y$  alone (which are then the marginal probability function of  $X$  and  $Y$ )  $X$  and  $Y$  are independent. If however,  $f(x, y)$  cannot be so expressed, then  $X$  and  $Y$  are dependent. If  $X$  and  $Y$  are continuous random variables, we say that they are independent random variables if the events  $X \leq x$  and  $Y \leq y$  are independent events for all  $x$  and  $y$ . In such case we can write

$P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$  or equivalently

$$F(x, y) = F_1(x) F_2(y)$$

Where  $f_1(x)$  and  $f_2(y)$  are the (marginal) distribution functions of X and Y, respectively conversely, X and Y are independent random variables if for all  $x$  and  $y$ , their joint distribution function  $F(x, y)$  can be expressed as a product of a function of  $x$  alone and a function of  $y$  alone (which are the marginal distribution of X and Y respectively)

If however,  $f(x, y)$  cannot be expressed, then  $x$  and  $y$  are dependent. For continuous independent random variables, it is also true that the joint density function  $f(x, y)$  is the product of a function of  $x$  alone,  $f_1(x)$  and a function of  $y$  alone,  $f_2(y)$ , and these are the (marginal) density functions of  $x$  and  $y$ , respectively.

## 2.10 CONDITIONAL DISTRIBUTIONS

We already know that if  $P(A) > 0$ ,

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

If  $x$  and  $y$  are discrete random variables and we have the events  $(A: x = x)$

$(B: Y = y)$ , the above equation becomes

$$P(Y = y|X = x) = \frac{f(x, y)}{f_1(x)}$$

Where  $f(x, y) = P(x=X, y=Y)$  is the joint probability fraction and  $f_1(x)$  is the marginal probability function for  $x$ .

$$F(y|x) = \frac{f(x, y)}{f_1(x)}$$

And call it the conditional probability function of  $Y$  given  $X$ . Similarly, the conditional probability function of  $X$  given  $Y$  is

$$F(x|y) = \frac{f(x, y)}{f_2(y)}$$

We can also denote

$f(x_i|y)$  and  $f(y_i|x)$  by  $f_i(x|y)$  and  $f_2(y|x)$  respectively these ideas can easily be extended to the case where  $X, Y$  are continuous random variables. For example the conditional density function of  $Y$  given  $X$  is

$$f(y|x) = \frac{f(x, y)}{f_1(x)}$$

Where  $f(x, y)$  is the joint density function of  $x$  and  $y$ , and  $f_1(x)$  is the marginal density function of  $x$ . Using the equation above we can find the probability of  $y$  being between  $c$  and  $d$  given that  $x < X < x + dx$  is

$$P(c < Y < d | x < X < x + dx) = \int_c^d f(y|x) dy$$

### Example 10.1

2: 101 A random variable  $x$  has the density function

$$f(x) = \frac{C}{(x^2 + 1)}, \text{ where } -\infty < x < \infty.$$

- Find the value of the constant  $C$
- Find the probability that  $X^2$  lies between  $1/3$  and  $1$

### Solution

- We must have  $\int_{-\infty}^{\infty} f(x) dx = 1$  ie

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{c}{x^2 + 1} dx &= C \tan^{-1} x \Big|_{-\infty}^{\infty} \\ &= C \left[ \frac{\pi}{2} - \left( -\frac{\pi}{2} \right) \right] = 1 \\ &= C \pi = 1 \\ C &= \frac{1}{\pi} \end{aligned}$$

- If  $\frac{1}{3} \leq x^2 \leq 1$ , then either  $\frac{\sqrt{3}}{3} \leq x \leq 1$  or

$$-1 \leq x \leq -\frac{\sqrt{3}}{3}$$

Thus the required probability is

$$\begin{aligned}
\frac{1}{\Pi} \int_{-1}^{-\sqrt{3}/3} \frac{dx}{x^2 + 1} + \frac{1}{\Pi} \int_{\sqrt{3}/3}^1 \frac{dx}{x^2 + 1} &= \frac{2}{\Pi} \int_{\sqrt{3}/3}^1 \frac{dx}{x^2 + 1} \\
&= \frac{2}{\Pi} \left[ \tan^{-1}(1) - \tan^{-1}\left(\frac{\sqrt{3}}{3}\right) \right] \\
&= \frac{2}{\Pi} \left( \frac{\Pi}{4} - \frac{\Pi}{6} \right) = \frac{1}{6}
\end{aligned}$$

**Example 2:10:2**

2: 102 find the distribution function corresponding to the density function of the example 2:10:1 given above.

**Solution :**

$$\begin{aligned}
f(x) &= \int_{-\infty}^x f(u) du = \frac{1}{\Pi} \int_{-\infty}^x \frac{du}{u^2 + 1} \\
&= \frac{1}{\Pi} \left[ \tan^{-1} u \int_{-\infty}^x \right]^{-\infty} \\
&= \frac{1}{\Pi} \left[ \tan^{-1} x - \tan^{-1}(-\infty) \right] \\
&= \frac{1}{\Pi} \left[ \tan^{-1} x + \frac{\Pi}{2} \right] = \frac{1}{2} + \frac{1}{\Pi} \tan^{-1} x
\end{aligned}$$

**Example 2:10:3**

2: 10:3 The distribution function for a random variable x is

$$f(x) = \begin{cases} 1 - e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find (a) the density function (b) the probability that  $x > 2$  and (c) the probability that  $-3 < x \leq 4$

**Solution :**

$$f(x) = 1 - e^{-2x} \quad \text{when } x \geq 0$$

a.  $\frac{d}{dx} (f(x)) = 0 - (-2) e^{-2x} = 2e^{-2x}$

$$2 > f(x) = \frac{d}{dx} f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$(b) \quad p(x > 2) = \int_2^{\infty} 2e^{-2x} dx = -e^{-2x} \Big|_2^{\infty}$$

$$= -e^{-2(\infty)}$$

$$= -e^0 + e^{-4} = e^{-4}$$

c.  $p(-3 < x \leq 4) = \int_{-3}^4 f(x) dx = \int_{-3}^0 0 dx + \int_0^4 2e^{-2x} dx$

$$= -e^{-2x} \Big|_0^4 = 1 - e^{-8}$$

OR  $p(-3 < x \leq 4) = P(x \leq 4) - P(x \leq -3)$

$$P(4) - P(-3)$$

$$= (1 - e^{-8}) - (0)$$

$$= 1 - e^{-8}$$

Example 2:10:4 The joint probability function of two discrete random variable X and Y - 1 is given by  $f(x, y) = C(2x + y)$ , where x and y can assume all integers such that

$$0 \leq x \leq 2, \quad 0 \leq y \leq 3, \quad \text{and } f(x, y) = 0 \text{ otherwise}$$

a. find the value of the constant C

b. Find  $p(x = 2, y = 1)$ . (c) find  $p(x \geq 1, y \leq 2)$

**Solution:**

The sample points  $(x, y)$  for which probabilities are different from zero are indicated below

$y/x$	0	1	2	3	Totals
0	0	C	2c	3c	6c
1	2c	3c	4c	5c	14c
2	4c	5c	6c	7c	22c
Totals	6c	9c	12c	15c	42c

The probabilities associated with these points, given by  $C(2x + y)$ , are shown in the table above

Since the grand total must equal to 1 i.e  $42C = 1$

$$C = \frac{1}{42}$$

b.  $P(x = 2, Y = 1) = C(2 \times 2 + 1) + \frac{F(x_1, y)}{\text{grand total}}$

$$\begin{aligned}
 & 5c + \frac{5c}{42c} \\
 & = 5c + \frac{5}{42}
 \end{aligned}$$

$$\begin{aligned}
\text{c. } P(x \geq 1, y \leq 2) &= \sum_{x \geq 1} f \sum_{y \leq 2} f(x, y) \\
&= (2c + 3c + 4c) + (4c + 5c + 6c) \\
&= 24c = \frac{24}{42} = \frac{4}{7}
\end{aligned}$$

Example 2:10:5 Find (a)  $f(y|2)$ ,

(b)  $P(y=1|x=2)$  for the distribution

Find the example 2:10:4 above

### Solution

Using the results from the above example

$$f(y|x) = \frac{f(x, y)}{f_1(x)} = \frac{(2x + y)/42}{f_1(x)}$$

So that with  $x = 2$

$$\begin{aligned}
f(y|2) &= \frac{(4 + y)/42}{\frac{11}{21}} = \frac{4 + y}{11} \cdot \frac{21}{42} \\
&= \frac{4 + y}{22}
\end{aligned}$$

$$\begin{aligned}
\text{(b) } P(y=1|x=2) &= f(1|2) = \frac{4+1}{22} \\
&= \frac{5}{22}
\end{aligned}$$



## **2.11 CONCLUSION**

In this unit, you studied random variables and classification, distributive functions for discrete and continuous random variables. You also learned Graphical representation and joint distribution for discrete and continuous random variables.

Independence and conditional probability of random variables and related working examples are also learned from this its

## **2:12 SUMMARY**

In this unit distribution of random variables spaces that you studied included

- (1) Meaning of random variables and its classification.
- (2) Distribution functions for discrete and continuous random variables.
- (3) Graphical representation of random variables
- (4) Joint distribution for discrete and continuous random variables.
- (5) Independence and conditional probability of random variables
- (6) Worked examples on each concept of random variables.

**Exercise 2.12 .1 (S A E)**

Suppose that a pair of fair dice are to be tossed, and let the random variable  $x$  denote the sum of the points. Obtain the probability distribution for  $x$ .

### 2.13 TUTOR MARKED ASSIGNMENT (TMA)

Exercises 2.131. The joint density function of two continuous random variables  $x$  and  $y$  is

$$f(x, y) = \begin{cases} Cxy & 0 < x < 4, 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of the constant  $c$

- (a) Find  $P(1 < X < 2, 2 < Y < 3)$
- (b) Find  $p(X \geq 3, Y \leq 2)$ .

### 2.14 REFERENCES/ FURTHER READING / OTHER RESOURCES

Marry R Spiegel *etal* (2009) probability and statistics third edition published by mc craw hill

## **UNIT 3 EXPECTATION OF RANDOM VARIABLES.**

### **3.0 INTRODUCTION.**

This is a very important concept in probability and statistics. The unit will focusses on mathematical expectation of random variables. Expected value for discrete and continuous random variables are stated. Variance and Standard Deviation for discrete and continuous random variables are highlighted; also some important theorems on the expectation of random variables are discussed.

Moment and Moment generating functions for random variables are also learned from this unit. Characteristics function of random variables are also learned and relevant working examples on each concept are given to make the unit more meaningful.

### **3.1 OBJECTIVE**

At the end of this unit, student should be able to

1. Define Expectation of random variable
2. Express mathematically the Expected Value of Mean for discrete and continuous random variables.
3. State and prove Theorems on Expectation.

4. State Variance and Standard Deviation for Discrete and Continuous Random Variables.
5. Find the Mathematical Expectation of Moments and Moments Generation Function for Discrete and Continuous random variables.
6. Find the characteristic function of a given random variable.
7. Solve related examples on the mathematical expectation of random variables.

### 3.2 What Is Expectation of Random Variables?

Let  $X$  be a discrete random variable with probability function  $f(x)$ , Then the expected value of  $x$ ,  $E(x)$  is defined to be

$$\begin{aligned}
 E(X) &= \sum_{j=1}^n X_j f(x_j) = \sum X f(x) \\
 &= \sum x f(x) \qquad (1)
 \end{aligned}$$

If  $f(x)$  is a accurate characterization of the population frequency distribution, then

$$E(x) = \mu \text{ ( the population mean)}$$

For a continuous random variable  $x$  having density function  $f(x)$  the expectation of  $x$  is defined as

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{--- (2)}$$

Provided that the integral converges absolutely. Where  $f(x)$  is the value of its probability density at  $x$  e.g If  $x$  is the member of point roll with a balance die.  $F(x) = \frac{1}{6}$  for  $x = 1, 2, 3, 4, 5, 6$  and its mathematical

$$\begin{aligned} \text{expectation is } E(x) &= 1x\frac{1}{6} + 2x\frac{1}{6} + 3x\frac{1}{6} + 4x\frac{1}{6} + 5x\frac{1}{6} + 6x\frac{1}{6} \\ &= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} \\ &= \frac{21}{6} = 3\frac{1}{2} \quad (3.5) \end{aligned}$$

Also if  $X$  has the uniform density function

$f(x) = \frac{1}{2}$  for  $2 < x < 4$  and  $f(x) = 0$  elsewhere. Then

$$E(x) = \int_2^4 x \cdot \frac{1}{2} dx = \frac{1}{4} x^2 \Big|_2^4 = 3$$

In many scientific problem, we are interested not only the expected value of a random variable  $X$  but also in the expected value of random variable related to  $x$ . Thus, we might be interested in random variable  $y$  whose values are related to those of  $X$  by  $y = g(x)$

$$\text{The } E[g(x)] = \sum g(x) f(x) \quad \text{--- (3)}$$

Where  $x$  is discrete and  $E[(g(x))] = \int_{-\infty}^{\infty} g(x) f(x) dx$

For continuous case using the above example find the expectation of

$g(x) = x^2$  for the number of points rolls with a balance die

### Solution

$$E[g(x)] = \sum x^2 f(x)$$

$$\begin{cases} f(x) = 1/6, & x = 1, 2, 3, 4, 5, 6 \\ f(x) = 0, & \text{elsewhere} \end{cases}$$

$$\begin{aligned} \therefore E[g(x)] &= 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} \\ &= 15 \frac{1}{6} \end{aligned}$$

Similarly for the random variable with the uniform density function

$f(x) = \frac{1}{2}$  for  $2 < x < 4$  and  $f(x) = 0$  elsewhere, we get

$$\begin{aligned} E[g(x)] &= E(x^2) = \int_2^4 x^2 f(x) dx \\ &= \int_2^4 x^2 \left(\frac{1}{2}\right) dx \\ &= \int_2^4 \frac{1}{2} x^2 dx = \frac{1}{6} x^3 \Big|_2^4 \\ &= 9 \frac{1}{3} \end{aligned}$$

### 3.3 THEOREMS ON EXPECTATION

Theorem 3.31: If  $c$  is any constant, then  $E(cX) = cE(X)$

Also,  $E[cg(x)] = cE[g(x)]$

Theorem 3.3.2 : If  $x$  and  $y$  are any random variables, then

$$E(X + Y) = E(x) + E(Y)$$

**Theorem 3.3.3:** If  $x$  and  $y$  are independent random variables, then

$$E(XY) = E(X)E(Y)$$

**THEOREM 3.3.4**

$$E[(ax + b)^n] = \sum_{t=0}^n \binom{n}{t} a^{n-t} b^t E(x^{n-t})$$

For instance, if  $n = 1$ ,

$$E[(ax + b)^1] = \sum_{t=0}^1 \binom{1}{t} a^{1-t} b^t E(x^{1-t})$$

$$= \binom{1}{0} a E(x) + \binom{1}{1} b E(1)$$

$$= a E(x) + b$$

If  $n = 2$

$$E[(ax + b)^2] = \sum_{t=0}^2 \binom{2}{t} a^{2-t} b^t E(x^{2-t})$$

$$= \binom{2}{0} a^2 E(x^2) + \binom{2}{1} ab E(x) + \binom{2}{2} b^2 E(1)$$

$$= a^2 E(x^2) + 2ab E(x) + b^2$$

**Theorem 3.3.4** can easily be proved by mathematical induction

If  $z$  is a random variable whose values are related to those of  $x$  and  $y$  random variable  $x$  and  $y$  by means of equation  $z = g(x, y)$

The mathematical Expectation is written as

$$E[g(x, y)] = \sum_x \sum_y g(x, y) f(x, y) \quad \text{---(4)}$$

$$\text{Or } E[g(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad \text{---(5)}$$

In equation (4)  $f(x, y)$  is the value of joint probability function of  $x$  and  $y$  at  $x, y$  while in equation (5)  $f(x, y)$  corresponds to the value of the joint probability density

### 3.4 THE VARIANCE AND STANDARD DEVIATION

We have already noted that the expectation of a random variable  $X$  is often called the means and is denoted by  $\mu$  another quality of great importance in probability and statistics is called variance and is defined by  $\text{Var}(x) = E(x - \mu)^2$  .....(6)

The variance is a non-negative number. The positive squaeroot of the variance is called the Standard Deviatiation and is given by

$$\sigma_x = \sqrt{\text{Var}(x)} = \sqrt{E((X - x)^2)} \dots\dots\dots(7)$$

The standard derivation is often denoted by  $\sigma$  instead of  $\sigma_x$ , and the variance in such case is  $\sigma^2$

If  $x$  is a discrete random variable taking the values  $X_1, X_2, \dots, X_n$  and having probability function  $f(x)$ , then the variance is given by

$$\alpha_x^2 = E(X - \mu)^2 = \sum_{j=1}^n (x_j - \mu)^2 f(x_j) = \sum (x - \mu)^2 f(x) - \dots\dots\dots(8)$$

In the special case of (8) where the probabilities are all equal, we have  $\sigma^2 = [(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2] / n - \dots\dots\dots(9)$



Which is the variance for a set of n numbers  $x_1, \dots, x_n$

If X takes on an infinite number of values

$$X_1, x_2, \dots, \text{then } \sigma^2_x = E(X - \mu)^2 = \int_{-\infty}^{\infty} (X - \mu)^2 f(x) dx \dots \dots \dots (10)$$

Provided that the integral converges.

### 3.5. THEOREMS ON VARIANCE

#### Theorems 3.5.1

$$\begin{aligned} \sigma^2 &= E (x - \mu)^2 = E (x^2) - \mu^2 = \\ &= E (x^2) - (E(x))^2 \\ \text{where } \mu &= E (x). \end{aligned}$$

Theorem 3.5.2. If C is any constant,

$$\text{Var}(c x) = C^2 \text{Var} (x) - \dots \dots \dots (1)$$

Theorem 3.5.3 The quantity  $E ((x - a)^2)$  is a minimum when  $a = \mu = E (x)$

Theorem 3.5.4 If X and Y are independent random variables

$\text{Var} (X + Y) = \text{Var} (X) + \text{Var} (Y)$  or

$$\sigma^2_{x+y} = \sigma_x^2 + \sigma_y^2 \dots \dots \dots (12)$$

$\text{Var}(X - Y) = \text{Var} (X) + \text{Var} (Y)$  or

$$\sigma^2_{x-y} = \sigma_x^2 + \sigma_y^2 \dots \dots \dots (13)$$

### 3.6 MOMENTS

The rth moment of a random variable X about the mean  $\mu$ , also called the rth central moment, is defined as

$$\mu_r = E (x - \mu)^r \dots\dots\dots(14)$$

Where r = 0, 1, 2, ----- it follows that  $\mu_0 = 1, \mu_1 = 0$  and  $\mu_2 = \sigma^2$  i.e second moment about the mean is the variance.

We have, assuming absolute convergence.

$$\mu_r = \sum (x - \mu)^r f(x) \text{ (discrete) } \dots\dots\dots(15)$$

var iabl;e

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx \text{ (continuous) } \dots\dots\dots(16)$$

var iable

The rth moment of x about the origin, also called the rth raw moment, is defined as

$$\mu^1_r = E (x^r) \dots\dots\dots(17)$$

The zero moment and the first moment about the mean are respectively 1

and 0 Since  $\mu_0 = E [(x - \mu)^0] = E (1) = 1$

And  $\mu^1 = E [(x - \mu)^1] = E(x) = E(\mu) = \mu - \mu = 0$

The second movement called the variance and is dented by  $\sigma^2$ .

$$\mu_2 = E [(x - \mu)^2] = \sigma^2$$

This indicate the strength on dispersion of the distribution generated.

Generally moment about the mean describe the shape of the distribution of a random variable.

### 3.7 MOMENT GENERATING FUNCTIONS

Although the moment of some distribution can be determined directly by evaluating the necessary integral or sum. There exist on alternative technique which often provide considerable signification. This technique is based on the moment generating function which is given by

$$M_x(t) = E(e^{tx}) = \sum_{(discrete\ variable)} e^{tx} f(x) \dots\dots\dots(18)$$

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \dots\dots\dots(19)$$

$$but\ e^{tx} = 1 + tx + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + \dots\dots\dots + \frac{t^r x^r}{r!} \dots\dots$$

$$This\ for\ the\ discrete\ case\ M_x(t) = \sum \left[ 1 + tx + \frac{t^2 x^2}{2!} + \dots\dots\dots + \frac{t^r x^r}{r!} \dots\dots \right] f(x)$$

$$\Rightarrow M_x(t) = \sum f(x) + t \sum x f(x) + \frac{t^2}{2!} \sum x^2 f(x) + \dots\dots\dots - \frac{t^r}{r!} \sum x^r f(x) + \dots\dots$$

$$M_x(t) = 1 + \mu^1 t + \mu^2 \frac{t^2}{2!} + \dots\dots\dots - \mu^r \frac{t^r}{r!} + \dots\dots\dots(20)$$

Thus if we expand  $M_x(t)$  as a power series in  $t$ , the coefficient of  $t^r$  is  $\mu^r$ , which is the  $r$ th moment about the origin of distribution  $X$ .

You observed that the maclaurins series of a function  $M_x(t)$  with coefficient

$\frac{t^r}{r!}$  is the rth derivation of the function with respect to  $t$  at  $t = 0$ .

Hence another way of determine the movement of a distribution is given by variation

$$\mu_r^1 = \left[ \frac{d^r M_x(t)}{dt^r} \right]_{t=0}$$

### 3.8 THEOREMS ON MOMENT GENERATING FUNCTION

Theorems 3.8.1 If  $M_x(t)$  is the moment generating function of the random variable  $x$  and  $a$  and  $b(\neq 0)$  are constants, then the moment generating

function of  $\left(\frac{x+a}{b}\right)$  is  $M_{\left(\frac{x+a}{b}\right)}(t) = e^{at/b} M_x\left(\frac{t}{b}\right)$

Theorem 3.8.2 :- If  $X$  and  $Y$  are independent random variables having moment generating fractions  $M_x(t)$  and  $M_y(t)$ , respectively, then

$$M_{x+y}(t) = M_x(t)M_y(t).....(21)$$

i.e The moment generating function of a sum of independent random variables is equal to the product of their moment generating functions.

### 3.9 CHARACTERISTIC FUNCTIONS

Characteristic function of a random variable  $X(\omega)$  defined on  $(\Omega, \mathcal{A}, p)$  provides a powerful and application tool in the theory of probability.

Characteristic function has one important advantages over moment generating function because it can be need to prove both the weak law of large numbers and the Central limit theorem which will be treated in the next unit.

Definition: Let  $X$  be a random variable with probability distribution function. The characteristic function of  $x$  is defined for real  $t$  by

$$\begin{aligned}\psi(t) &= \int_{\mathcal{R}} e^{itx} dF(x) = \int e^{itx} p(dx) \\ &= E(e^{itx}) = E(e^{itx})\end{aligned}$$

Where  $e^{itx} = \cos t x + i \sin t x$

$$E(e^{itx}) = E(\cos t x) + i E(\sin t x)$$

Properties of characteristic function

(a) (i)  $\psi(t)$  is informally continuous on the real line.

(ii)  $\psi(0) = 1$

(iii)  $|\phi(t)| \leq 1$  for all  $t$ .

Since  $|e^{itx}| = 1$

Proof:

$$\begin{aligned}
 \text{(i) } |\phi(t+h) - \phi(t)| &= \left| \int_{-\infty}^{\infty} \{e^{i(t+h)x} - e^{itx}\} p(dx) \right| \\
 &\leq \int_{-\infty}^{\infty} |e^{i(t+h)x} - e^{itx}| p(dx) \\
 &= \int_{-\infty}^{\infty} |e^{itx}| |e^{ihx} - 1| p(dx) \leq \int_{-\infty}^{\infty} |e^{ihx} - 1| p(dx)
 \end{aligned}$$

By the dominated convergence theorem, we have

$$\lim_{h \rightarrow 0} \int_{-\infty}^{\infty} |e^{itx}| |e^{ihx} - 1| p(dx) = \int_{-\infty}^{\infty} \lim_{h \rightarrow 0} |e^{ihx} - 1| p(dx) = 0$$

**Note:**

$$\text{(i) } |e^{ihx} - 1| \leq 2$$

(ii) The limit tends to zero independently of  $t$ .

Thus,  $|\phi(t+h) - \phi(t)| \rightarrow 0$  independent of  $t$ .

Hence  $\phi(t)$  is uniformly continuous on the real line

$$\text{(iii) } |e^{itx}| = |\cos tx + i \sin tx| = \sqrt{\cos^2 tx + \sin^2 tx} = 1$$

$$|\phi(t)| = \left| \int e^{itx} p(dx) \right| \leq \int |e^{itx}| p(dx) = \int p(dx) = 1$$

(b) The characteristic function of the sum of independent random variables is the product of their characteristic functions

Proof :

Let  $S_n = X_1 + X_2 + \dots + X_n$  where  $X_1, X_2, \dots, X_n$  are independent random variables.

Then

$$\begin{aligned} \psi_{sn}(t) &= E[e^{its_n}] = E[e^{it(x_1 + \dots + x_n)}] \\ &= E(e^{itx_1}) E(e^{itx_2}) \dots E(e^{itx_n}) \\ &= \psi_{x_1}(t) \psi_{x_2}(t) \dots \psi_{x_n}(t) \text{ for all real } t. \end{aligned}$$

If  $(X_i)$  are independent and identically distributed then

$$\psi_{sn}(t) = |\psi_x(t)|^n$$

(c) Unlike moment generating functions,  $\psi_x(t)$

is finite for all variable  $x$  and all real number  $t$ . The reason being that  $e^{it}$  is bounded while it is unbounded for  $-\infty < |x| < \infty$

d. The distribution function of  $X$  and hence the Pdf, if it exists can be obtained from the characteristic function using an "Inversion formula:",

If  $X$  is integer valued random variable then  $f_x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-in} \psi_x(t) dt$

If  $x$  is a continuous random variable, then  $f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it} \psi_x(t) dt$

assuming  $\int_{-\infty}^{\infty} |\psi_x(t)| dt < \infty$

e. Properties of characteristic function enable us to prove both the weak law of large numbers and the central limit theorem.

Properties (c) (d) and (e) are important advantages of functions over moment generating function.

(f) If two random variables have the same characteristic function they have the same distribution function.

(g) If  $x$  has finite  $n^{\text{th}}$  moment, then  $\mu_x^{(n)}(t)$  exists and is continuous in  $t$ .

$$\mu_x^{(n)}(t) = d^n \frac{E(e^{itx})}{dt^n} = E\{(ix)^n e^{itx}\}$$

$$\text{Thus } E(x^n) = \frac{\mu_x^{(n)}(0)}{i^n}$$

**Example 3.9.1:** let  $x$  have an exponential distribution with parameter  $\beta$  find the characteristic function of  $x$

**Solution**

$$\begin{aligned} \psi_x(t) &= E(e^{itx}) = \int_0^{\infty} e^{itx} \beta e^{-\beta x} dx \\ &= \beta \int_0^{\infty} e^{-(\beta-it)x} dx = \frac{\beta}{\beta-it} \end{aligned}$$

**Example 3.9.2** Let  $x$  be uniformly distributed on  $(-1,1)$ . Find the characteristic function of  $X$ .



**Solution**

$$\begin{aligned} \psi_x(t) = E &= \int_{-1}^1 e^{itx} \frac{1}{2} dx \\ &= \frac{1}{2} \left. \frac{e^{itx}}{it} \right|_{-1}^1 \quad t \neq 0 \\ &= \frac{1}{2} \left[ \frac{e^{it} - e^{-it}}{it} \right] = \frac{\sin t}{t} \end{aligned}$$

$$e^{it} = \cos t + i \sin t,$$

**Note:**  $e^{-it} = \cos t - i \sin t,$

$$e^{it} - e^{-it} = 2i \sin t.$$

**Example 3.9.3 :** Find the characteristic function of the random variable X having density function give by

$$f(x) = \begin{cases} 1/2a & |x| \leq a \\ 0 & \text{otherwise} \end{cases}$$

**Solution**

$$\begin{aligned} E(e^{itx}) &= \int_{-\infty}^{\infty} e^{itx} f(x) dx = \frac{1}{2a} \int_{-a}^a e^{itx} dx \\ &= \frac{1}{2a} \left. \frac{e^{itx}}{it} \right|_{-a}^a = \frac{e^{ita} - e^{-ita}}{2iat} \\ &= \frac{\sin at}{at} = \frac{\sin \vartheta}{\vartheta} \end{aligned}$$

Using Euler's formula with  $\vartheta = at$

**Example 3.9.4 :** Find the expectation of a discrete random variable x whose probability function is given by

$$f(x) = \left(\frac{1}{2}\right)^x \quad [x = 1, 2, 3, \dots -]$$

**Solution**

We have  $E(x) = \sum_{x=1}^{\infty} x \left(\frac{1}{2}\right)^x = \frac{1}{2} + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{8}\right) + \dots$

To find the sum,

Let  $S = \frac{1}{2} + 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{8}\right) + 4\left(\frac{1}{16}\right) + \dots$

Then  $\frac{1}{2} S = \frac{1}{4} + 2\left(\frac{1}{8}\right) + 3\left(\frac{1}{16}\right) + \dots$

By subtracting  $\frac{1}{2} S = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots = 1$

Therefore,  $S = 2$

**Example 3.9.5: A Continuous Random Variable X Has Probability**

**Density given by**  $f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \leq 0 \end{cases}$

Find (a)  $E(x)$  (b)  $E(x^2)$

(a)  $E(x) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x (2e^{-2x}) dx$

$= 2 \int_0^{\infty} x e^{-2x} dx$

$= 2 \int_0^{\infty} (x) \left(\frac{e^{-2x}}{-2}\right) - (1) \left(\frac{e^{-3x}}{4}\right) \Big|_0^{\infty} = \frac{1}{2}$

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

(b)  $= 2 \left[ (x^2) \left(\frac{e^{-2x}}{-2}\right) - (2x) \left(\frac{e^{-2x}}{4}\right) + \left(4 \left(\frac{e^{-2x}}{8}\right)\right) \Big|_0^{\infty} = \frac{1}{2} \right]$

**Example 3.9.6** Find (a) the variance, (b) The standard deviation of the sum obtained in tossing a pair of fair dice.

**Solution**

$$E(x) = E(y) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \dots + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

$$\begin{aligned} \text{Hence } E(x^2) = E(y^2) &= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + \dots + 6^2\left(\frac{1}{6}\right) \\ &= 1\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + \dots + 36\left(\frac{1}{6}\right) \\ &= \frac{91}{6} \end{aligned}$$

$$\begin{aligned} \text{Then } \text{Var}(X) = \text{Var}(Y) &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} \\ &= \frac{35}{12}. \end{aligned}$$

And since X and Y are independent

$$\begin{aligned} \text{var}(x + y) &= \text{var}(x) + \text{var}(y) \\ &= \frac{35}{12} + \frac{35}{12} = \frac{70}{12} \\ &= \frac{35}{6} \end{aligned}$$

(c) Standard Deviation =  $\sqrt{\text{variance}}$

$$\text{ie } \sigma_{x+y} = \sqrt{\text{var}(x+y)} = \sqrt{\frac{35}{6}}$$

**Example 3.9.7:** The random variable x can assume the values 1 and -1 with probability  $\frac{1}{2}$  each. Find (a) the moment generating function (b) the first four moment about the origin.

**Solution:**

$$(a) \quad E\left(e^{itx}\right) = e^{t(1)} \left(\frac{1}{2}\right) + e^{t(-1)} \left(\frac{1}{2}\right) \\ = \frac{1}{2} \left(e^t + e^{-t}\right) \dots\dots\dots(1)$$

$$(b) \quad \text{we have } e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \dots\dots\dots(2)$$

$$\text{then from (1)} \quad \frac{1}{2} \left(e^t + e^{-t}\right) = 1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \dots\dots\dots$$

$$\text{But from (2)} \quad M_x(t) = 1 + \mu^1_1 t + \mu^1_2 \frac{t^2}{2!} + \mu^1_3 \frac{t^3}{3!} + \mu^1_4 \frac{t^4}{4!} + \dots\dots\dots$$

Then comparing (1) and (2) we have

$$\mu = 0, \quad \mu^1_2 = 1, \quad \mu^1_3 = 0, \quad \mu^1_4 = 1$$

The odd moment are all zero, and the even moments are all one

**Example 3.9.8:** A random variable  $x$  has density function given by

$$f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find (a) the moment generating function,

(b) the first four moments about the origin

**Solution:**

$$(a) \quad M_x(t) = E\left(e^{tx}\right) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ = \int_{\infty}^{\infty} e^{tx} (2e^{-2x}) dx \\ = 2 \int_{\infty}^{\infty} e^{(t-2)x} dx \\ = 2e \left. \frac{(t-2)^x}{t-2} \right|_0^{\infty}$$

$$= \frac{2}{2-t} \text{ assuming } t < 2$$

(b) If  $|t| < 2$  we have

$$\frac{2}{2-t} = \frac{1}{1-t/2} = 1 + \frac{t}{2} + \frac{t^2}{4} + \frac{t^3}{8} + \frac{t^4}{16}$$

$$\text{But } M_x(t) = 1 + \mu_1 + \mu_2^1 \frac{t}{2!} + \mu_3^1 \frac{t^3}{3!} + \mu_4^1 \frac{t^4}{4!} + \dots$$

Therefore, on comparing terms  $\mu = 1/2$

$$\mu_2^1 = \frac{1}{2}, \mu_3^1 = \frac{3}{4}, \mu_4^1 = \frac{3}{2}$$

### 3.10 CONCLUSION

In this unit you have learnt mathematical expectation of random variables for discrete and continuous random variables. You also learned Variance and Standard Deviation for discrete and continuous random variables.

Some important theorems on Expectation, Variance and Standard Deviation are stated.

Moment, Moment Generating functions and Characteristic function are fully treated and related working example on each concept are easily shown to make the learning of the unit more meaningful.

### 3.11 SUMMARY

In this unit expectation of random variables that you have studied included the following:

- Meaning of Expectation for Discrete and Continuous random variables
- Mathematical Expectation for discrete and continuous random variables.
- Expected Value for Variance and Standard Deviation .
- Theorems on the Expectation of Random Variables
- Moment, Moment Generating Function for Discrete and Continuous Random Variables
- Characteristic Functions of Random Variables
- Working examples on the Mathematical Expectation of Random Variables

#### Exercise: 3.11.1 (ASE)

The density function of a random variables X is given by

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

find (a)  $E(x)$  (b)  $E(x^2)$

### 3.1.2 TUTOR MARKED ASSIGNMENT (TMA)

Exercises 3.12.1: if  $X$  is random variable of Exercise 3.11.1 above find

$$E(3x^2 - 2x)$$

Exercises 3.12.2 A random variable  $X$  has  $E(x) = 2$ ,  $E(x^2) = 8$  find

(a)  $\text{Var}(x)$  (b)  $\sigma_x$

### 3.13 Reference / Further Reading / other resources

Murray R. Spiegel et al (2009) Probability and statistics.

Third edition published by McGraw Hill Dr R.A. Kasumu (2003)

probability theory first edition published by FATOL VENTURES

LAGOS, DR S.A Okunuga (1998) Probability Distribution 2 lecturer

Materials.

## **LIMIT THEOREM**

### **4.0 INTRODUCTION**

The purpose of this unit is to acquaint the students with the limit theorems on Chebyshev's inequality, convergence, weak laws of large numbers, strong law of large numbers. Some of the theorems are proved and related working examples are shown

### **4.1 OBJECTIVES**

At the end of this unit student should be able to

- State and prove Chebyshev's inequality
- Define Convergence of random variables
- State and prove some theorems on convergence in measure
- State and prove weak law of large numbers
- State the strong law large numbers

### **4.2 CHEBYSHEV'S INEQUALITY**

This is an important theorem in probability and statistics that reveals a general property of discrete or continuous random variable having finite mean and variance known under the name of Chebyshev's inequality. Theorem 4.21 suppose that  $X$  is a random variable (discrete or continuous)



Having mean  $\mu$  and variance  $\sigma^2$ , which are finite. Then if epsilon ( $\epsilon$ ) is a positive number,

$$P(|x - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

or with  $\epsilon = k\sigma$

$$P(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

**Proof :**

We shall proof for continuous random variables. A proof for discrete variables is similar it integrals are replaced by sums.

If  $f(x)$  is the density function of X, then

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Since the integrand is nonnegative, the value of the integral can only decrease when the range of integration is diminished.

Therefore,

$$\begin{aligned} \sigma^2 &\geq \int_{|x - \mu| \geq \epsilon} (x - \mu)^2 f(x) dx \geq \int_{|x - \mu| \geq \epsilon} \epsilon^2 f(x) dx \\ &= \epsilon^2 \int_{|x - \mu| \geq \epsilon} f(x) dx \end{aligned}$$

But the last integral is equal to  $P(|x - \mu| \geq \epsilon)$

Hence,

$$P(|x - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

### 4.3 CONVERGENCE OF RANDOM VARIABLE

**Definition 1** A sequence of random variables  $X_n$  is said to converge in distribution or in law to  $X$  (we write  $X_n \xrightarrow{L} X$ ) if the corresponding sequence of distribution functions  $F_n, F_n \rightarrow F$  as  $n \rightarrow \infty$ .

In this case  $F_n$  is a distribution function of  $F$ .

**Example:** Consider the random variable  $X_n$  which is a Binomial

*Bi (n, p), then*  
 random variable *Bi (1, p), Bi (2, p), Bi (3, p) ..... → X*  
*Bi (n, p) tends to normal*

Furthermore if the corresponding  $F$  distribution gives:

$$F_n = \Pr(X \leq x) = \sum \binom{n}{x} p^x q^{n-x} \rightarrow F$$

If  $X_n \xrightarrow{L} X$  ie  $F_n(X) \rightarrow F$  distribution function of  $X$  Then  
 (1)  $C_n(t) \rightarrow C(t)$

Where  $C_n(t)$  stands for corresponding characteristic function of random variable  $X_n$

(2) for any bounded continuous function

$$\int g d F_n(x) \rightarrow \int g d F$$

**Definition 2:** The sequence  $\{X_n\}$  of random variables is said to converge to a random variable  $X$  if

$P \{ |X_n - X| > \varepsilon \} \rightarrow 0$  as  $n \rightarrow \infty$  for an  $\varepsilon > 0$ . we indicate this by  $X_n \xrightarrow{P} X$

**Definition 3:** The sequence  $\{X_n\}$  of random variables is said to converge in mean square to a random variable  $X$  if

$\varepsilon \{ |X_n - X|^2 \} \rightarrow 0$  as  $n \rightarrow \infty$ . we indicate this by  $X_n \xrightarrow{m.s} X$

*Definition 4:* The sequence  $\{X_n\}$  of random variable is said to converge with probability one (or almost surely, a.s.) to a constant  $C$  if  $P \left\{ \lim_{n \rightarrow \infty} X_n = c \right\} = 1$

We indicate this by  $X_n \rightarrow c$  or  $X_n \xrightarrow{X_n \text{ w.p.}} C$

or equivalently,  $\lim P \left\{ \sup_{n \geq N} |X_n - C| > \varepsilon \right\} = 0$ , for every  $\varepsilon > 0$ .

**Note:**

1. In general  $f_n \rightarrow f$  does not imply that  $\{X_n\}$  converges to a random variable for example, Suppose  $X$  is  $N(0,1)$  and for all  $n$ , let  $X_n = -X$

Then  $X_n$  is  $N(0,1)$  That is,  $F_n(x) = F(x)$

$F_n(x) = (f(x))$  for all  $n$ . But

$$P \{ |X_n - x| \geq \varepsilon \} = P \{ |-2x| \geq \varepsilon \} = P \left\{ |x| \geq \frac{\varepsilon}{2} \right\} \rightarrow 0 \text{ as } n \rightarrow \infty$$

**Hence**  $X_n \xrightarrow{P} X$

(ii) If  $X_n \xrightarrow{P} X$  then  $F_n \rightarrow F$ . letting  $\varepsilon \rightarrow 0$ ,

We see that  $F_n(x) \rightarrow F(x)$

#### 4.4 DEMOVRE'S THEOREM

Let  $X_n$  be  $Bi(n, p)$  then,

$$Y_n = \frac{X_n - np}{\sqrt{npq}} \rightarrow N(0,1) \text{ as } n \rightarrow \infty$$

*Proof*  $\Pr(X_n = r) = \binom{n}{r} P^r q^{n-r}$

$$C_n(t) \Big|_{X_n} = E(e^{itx})$$

$$= \sum_{x=0}^{\infty} e^{itx} \Pr(X_n = x)$$

$$= \sum_{\mu=0}^{\infty} e^{itx} \binom{n}{r} P^x q^{n-x}$$

$$= \sum_{\mu=0}^{\infty} \binom{n}{r} (e^{itx} P)^x q^{n-x}$$

$$= (Pe^{it} + q)^n$$

Therefore characteristic function of  $Y_n = E(e^{iY_n})$

$$= E \left[ e^{\frac{itX_n}{\sqrt{npq}} - \frac{itnp}{\sqrt{npq}}} \right]$$

$$= e^{-it \sqrt{\frac{np}{q}}} E \left( \frac{e^{itx}}{\sqrt{npq}} \right)$$

$$\text{i.e } C_{Y_n}(t) = e^{-it \sqrt{\frac{np}{q}}} \left( p e^{\frac{it}{mpq} + q} \right)^n$$

$$\rightarrow \log C y_n(t) = -it \sqrt{\frac{np}{q}} + n \log \left( Pe^{\frac{it}{\sqrt{npq}}} + 1 - p \right)$$

$$= -it \sqrt{\frac{np}{q}} + n \log \left[ 1 + p \left( e^{\frac{it}{\sqrt{npq}}} - 1 \right) \right]$$

But  $e^z = 1 + z + \frac{z^2}{2!} + \dots$

And  $\log(1+z) = z - \frac{z^2}{2!} + \frac{z^3}{3!} - \dots$

$$\therefore \log \left( y_n(t) = it \sqrt{\frac{np}{q}} \right) + n \log \left\{ 1 + p \left( \frac{it}{\sqrt{npq}} + \frac{(it)^2}{2npq} + \frac{(it)^3}{\sqrt{npq}3!} + \dots \right) \right\} \Rightarrow \text{order}$$

$$= -it \sqrt{\frac{np}{q}} + n \log \left\{ 1 + p \left( \frac{it}{\sqrt{npq}} - \frac{t^2}{2npq} + 0 \left( \frac{1}{n^3} \right) \dots - \frac{p^2}{2} \left[ \frac{-t^2}{npq} + 0 \left( \frac{1}{n^2} \right) \right] \right) \right\}$$

$$= -it \sqrt{\frac{np}{q}} + it \sqrt{\frac{np}{q}} - \frac{nt^2p}{2npq} + \frac{np^2t^2}{2npq} + O \left( \frac{1}{n^{1/2}} \right)$$

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

$$C y(t) \rightarrow e^{-\frac{t^2}{2}} \text{ as } n \rightarrow \infty$$

but  $e^{-\frac{t^2}{2}}$  is characteristic function of a standard normal distribution  $N(0,1)$

$$\therefore -Y_n \xrightarrow{L} N(0,1)$$

### Theorem

If  $X_n$  is binomial  $Bi(n, p)$  as  $n \rightarrow \infty, p \rightarrow 0$ , such that

$$np = \lambda, X_n \xrightarrow{L} \text{Poisson}(\lambda)$$

$$\text{Recall: } C_{X_n}(t) = (pe^{it} + q)^n$$

$$\left( \frac{\lambda}{n} e^{it} + \left(1 - \frac{\lambda}{n}\right) \right)^n = \left( 1 + \frac{\lambda}{n} (e^{it} - 1) \right)^n$$

$$\lim \left( 1 + \frac{z}{n} \right)^n = e^z$$

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow \lambda}} C_{X_n}(t) = \lim_{n \rightarrow \infty} \left[ 1 + \frac{\lambda}{n} (e^{it} - 1) \right]^n \\ = e^{\lambda e^{it} - \lambda}$$

Which is the characteristic function of poisson ( $\lambda$ ), as  $n \rightarrow \infty, p \rightarrow 0, np = \lambda$ .

#### 4.5 CENTRAL LIMIT THEOREM

Let  $X_1, X_2, \dots, X_n$  be independently and identically distributed random variables each with mean  $\mu$  and variance  $\sigma^2$ .

$$\text{Let } \bar{Y}_n = \left( \frac{X_1 + X_2 + \dots + X_n}{\sqrt{n} \sigma} - \frac{\mu}{\sigma} \right)$$

$$= \frac{\bar{x} - \mu \sqrt{n}}{\sigma}$$

Then  $\bar{Y}_n \xrightarrow{L} N(0,1)$

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$n\bar{X} = X_1 + X_2 + \dots + X_n$$

$$\bar{X} = \frac{n\bar{X} - \mu}{\sqrt{n}\sigma}$$

Proof:

Let  $C(t)$  be the characteristics function of  $X_1$  so that  $C(t) = E(e^{itx})$

Since the random variable are identical

$$\Rightarrow E(e^{itx_1}) = E(e^{itx_2}) = \dots = E(e^{itx_n}) = C(t)$$

$$C_{\bar{Y}_n}(t) = E(e^{it\bar{y}_n})$$

$$= E\left[e^{it\left(\frac{X_1 + X_2 + \dots + X_n - \mu}{\sqrt{n}\sigma}\right)}\right]$$

$$= E\left[\frac{e^{itx_1}}{\sqrt{n}\sigma} + \frac{itx_2}{\sqrt{n}\sigma} + \dots + \frac{itx_n}{\sqrt{n}\sigma} - \frac{it\mu}{\sqrt{n}\sigma}\right]$$

$$= e^{-\frac{it\mu}{\sqrt{n}\sigma}} E\left[e^{\frac{itx_1}{\sqrt{n}\sigma}}\right] E\left[e^{\frac{itx_2}{\sqrt{n}\sigma}}\right] \dots E\left[e^{\frac{itx_n}{\sqrt{n}\sigma}}\right]$$

$$C_{\bar{Y}_n}(t) = e^{-\frac{it\mu}{\sqrt{n}\sigma}} \left[C\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n$$

But  $C(t) = 1 + \alpha_1 it + \alpha_2 \frac{(it)^2}{2!} + \dots$

Where  $d_1 = \mu^1 = \mu$

$$\sigma_2 = \mu_2 = \mu^2 + \sigma^2$$

$$\therefore C_{\bar{Y}_n}(t) = e^{-\frac{it\mu}{\sqrt{n}\sigma}} \left[1 + i \frac{t\mu}{\sqrt{n}\sigma} + \left(\frac{it}{\sqrt{n}\sigma}\right)^2 \frac{\mu^2}{\sqrt{2!}} + 0(t^3)\right]$$

Taking log to get  $e$

$$\log_e C_{\bar{Y}_n}(t) = -\frac{it\mu}{\sqrt{n}\sigma} + n \log \left( 1 + \frac{it\mu}{\sqrt{n}\sigma} - \frac{t^2\mu^2}{2\sigma n} + o(t^3) \right)^n$$

But  $\log(1+X) = X - \frac{X^2}{2} + \frac{X^3}{3} - \dots$

$$\therefore \log C_{\bar{Y}_n}(t) = -it \sqrt{\frac{n}{\sigma}} \mu + n \int \frac{it\mu}{\sqrt{n}\sigma} - \left( \frac{t^2\mu^2}{2\sqrt{n}\sigma} + \frac{(it\mu)^3}{2\sigma^2 n} + \dots \right)$$

$$= -\frac{t^2}{2\sigma^2} (\mu^2_2 - \mu^2) \text{ as } n \rightarrow \infty$$

$$= -\underline{t^2}$$

$$\therefore C_{\bar{Y}_n}(t) \rightarrow e^{-\frac{t^2}{2}} \text{ (remove the log)}$$

Hence  $\bar{Y}_n \xrightarrow{L} N(0,1)$

Definition 1: A sequence of random variable  $X_n$  is said to “converge in probability”(weakly) to the constant  $C$  if the limit

$$P(|X_n - C| > k) = 0 \text{ where } C > 0 \text{ This is written as}$$

$$X_n \xrightarrow{P} C$$

Definition 2: A sequence of random variable  $X_n$  is said to converge in probability to  $X$  if  $X_n \rightarrow X \xrightarrow{P} 0$



#### 4.6 KHINCHINE'S THEOREM

(Weak law of large Numbers) let  $X_1, X_2, \dots, X_n$  be a sequence of independently and identically distributed random variable each with mean  $\mu$

$$\text{Let } \bar{X}_n = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

$$\text{Then } \bar{X}_n \xrightarrow{P} \mu$$

Proof:

$$C(t) = E(e^{itx_1}) \\ = 1 + it\mu + o(t^2)$$

$$\text{Let } C_{\bar{X}}(t) = E(e^{it\bar{x}_n}) \\ = E(e^{it(X_1 + X_2 + \dots + X_n)}) \\ = E(e^{itx_1}) E(e^{itx_2}) \dots E(e^{itx_n})$$

$$= \left[ C\left(\frac{t}{n}\right) \right]^n = E(e^{itx_1}) \dots E(e^{itx_n}) \\ = \left[ 1 + \frac{it\mu}{n} + o\left(\frac{t^2}{n^2}\right) \right]^n$$

$$\Rightarrow C_{\bar{X}_n}(t) \rightarrow e^{it\mu} \text{ as } n \rightarrow \infty$$

$$\text{hence, } \bar{X}_n \xrightarrow{P} \mu$$

$$\text{meaning } \lim_{n \rightarrow \infty} P[|x - \mu| > k] = 0$$

$e^{it\mu}$  is the characteristics function of a random variables taking the value of  $\mu$  with probability 1

**THEOREM 4**

The strong law of large Numbers let  $\{X_i\}$  be a sequence of independent random variables such that

$$E(X_n) = 0, \text{Var}(X_n) = \sigma^2 < \infty \text{ and } \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < \infty,$$

Then the sequence  $\frac{1}{n} \sum_{i=1}^n X_i$  converge to 0 almost surely i.e)

Proof:

Let  $Y_n = \frac{X_n}{n} \quad n = 1, 2, \dots$

$$E(Y_n) = 0 \text{ and } \sum \text{Var}(Y_n) = \sum \frac{\sigma^2}{n^2} < \infty,$$

$$\sum_{n=1}^{\infty} Y_n = \sum_{n=1}^{\infty} \frac{X_n}{n}$$

Converge almost surely and hence

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum X_i = 0 \text{ a.e}$$

**Corollary 1**

If  $\{X_i\}$  is a sequence of independent identically distributed random variables such that  $E(X_i) = 0$

and variance such that  $\sigma^2 < \infty$ , then

$$\frac{S_n}{n} = \frac{\sum_{i=1}^n X_i}{n} \rightarrow 0 \text{ almost surely}$$

Corollary  $\Pi$

Let everything be as from the above theorem except that

$$E(X_n) = \mu \text{ and } \text{Var}(X_n) \text{ for all } n.$$

Then  $\frac{S_n}{n} \rightarrow \mu$  almost surely

**Example 4.61**

$$\text{Suppose } X_n = \begin{cases} 0 & \text{with Prob } \frac{1}{n} \\ 1 & \text{with prob } 1 - \frac{1}{n} \end{cases}$$

$X = 1$  with probability 1.

**Solution.**

Consider  $|X_n - X|$ . The only possible values of  $|x_n - x|$  are 0 and 1. So

$$|X_n - X| = \begin{cases} 0, & \text{with prob } \frac{1}{n} \\ 1 & \text{with prob } \frac{1}{n} \end{cases}$$

$$P \{ |X_n - X| \leq X \} = \begin{cases} 0 & X < 0 \\ 1 - \frac{1}{n} & 0 \leq X < 1 \\ 1 & X \geq 1 \end{cases}$$

$$\lim P \left\{ \left| \frac{X_n - X}{n} \right| \leq \frac{X}{n} \right\} = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

$$X_n \xrightarrow{P} X$$

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

$F_n(x) \rightarrow F(x)$  for all  $x$

**Note:**  $X_n \xrightarrow{d} X_n \rightarrow X$

### Example 4.6.2

Let  $x_1, x_2, \dots, X_n$  be a sequence of independent identically distributed poisson random variable with parameter  $\lambda$ . Then

$$E \left\{ \left| \bar{X}_n - \lambda \right|^2 \right\} = \text{Var}(\bar{X}_n) = \frac{\lambda}{n} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Therefore,  $\bar{X}_n \xrightarrow{\text{Meansquare}} \lambda$  or  $\bar{X}_n \xrightarrow{\text{quadratic mean}} \lambda$

### Example 4.

Let  $X_1, X_2, \dots, X_n$  be a sequence of i i d

Poison random variables. Than by chebychev's inequality

$$\begin{aligned} P \left\{ \left| \bar{X}_n - \lambda \right| > \ell \right\} &\geq \frac{\text{Var}(\bar{X}_n)}{\ell^2} \\ &= \frac{\lambda}{n\ell^2} \rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned}$$

Therefore,

$$\bar{X}_n \xrightarrow{P} \lambda$$

### EXAMPLE 4.6.3

Convergence in mean square implies convergence in probability.

By chebychev's inequality

$$P \left\{ |X_n - X| > \ell \right\} \leq E \left| \frac{(|X_n - X|^2)}{\ell^2} \right| \rightarrow 0 \text{ as } n \rightarrow \infty$$

Thus  $X_n \xrightarrow{P} X$

In general, convergence in rth mean implies convergence in probability.

## 4.7 CONCLUSION

In this unit you have learned chebyshev's inequality, the proving and the application of the theorem.

You have also learned convergence of random variables with different definitions, Demovre's theorem, and Central limit theorem by using characteristics function for finding convergence.

Moreover, the weak and strong law of large numbers are discussed the and related working examples on the theorems are treated.

## 4.80 SUMMARY

In this unit the following concepts have been learned.

2. Chebyshev's Inequality
3. Convergence of Random Variables
4. Demovre's Theorem
5. Central limit Theorem
6. Weak law of Large Numbers
7. Strong law of Large Numbers
8. Relevant examples on the theorems are treated.

### EXERCISE: 4.80.1 (S A E)

A random variable  $X$  has mean 3 and variance 2. Use Chebyshev's Inequality to obtain an upper bound for

$$(a) P(|x - 3| \geq 2), \quad (b) P(|x - 3| \geq | \cdot |).$$

### EXERCISE 4.80.2

Show that the (weak) law of large numbers can be stated as.

$$\lim P\left(\left|\frac{S_n}{n} - \mu\right| < \ell\right) = 1$$

#### 4.90 TUTOR MARKED ASSIGNMENT (TMA)

- (a) Show that the sequence  $X_n$  of random variable  $\theta$  is said to converge
- (i) In mean square to Random Variable  $X$ .
  - (ii) With probability one almost surely to a constant  $c$ .
- (b) State and Prove Central limit Theorem.

5.  $\phi$  References / Further Reading Dr RA kasumu (2003) Probability Theory first edition published by Fatol Ventures Lagos Dr S.A Okunuga (1998) Probability Distribution 2 lecture Materials Murray R. Spiegel *et al* Schaum's Outhies (2009) Probability and Statistics Published by Mc Graw Hill.