

**PROBABILITY AND RANDOM
PROCESSES
(R20A0408)
LECTURE NOTES**

**B.TECH
(II YEAR – II SEM)
(2022-23)**

Prepared by:

Mr. K.Suresh, Assistant Professor

Mr. K. Rasool Reddy, Assistant Professor

Department of Electronics and Communication Engineering



**MALLA REDDY COLLEGE
OF ENGINEERING & TECHNOLOGY**

(Autonomous Institution – UGC, Govt. of India)

Recognized under 2(f) and 12 (B) of UGC ACT 1956

(Affiliated to JNTUH, Hyderabad, Approved by AICTE - Accredited by NBA & NAAC – 'A' Grade - ISO 9001:2015 Certified)
Maisammaguda, Dhulapally (Post Via. Kompally), Secunderabad – 500100, Telangana State, India

CONTENTS

- ❖ SYLLABUS
- ❖ UNIT-I-PROBABILITY AND RANDOM VARIABLE
- ❖ UNIT-II- DISTRIBUTION AND DENSITY FUNCTIONS AND OPERATIONS ON ONE RANDOM VARIABLE
- ❖ UNIT-III-MULTIPLE RANDOM VARIABLES AND OPERATIONS
- ❖ UNIT-IV-STOCHASTIC PROCESSES-TEMPORAL CHARACTERISTICS
- ❖ UNIT-V- STOCHASTIC PROCESSES-SPECTRAL CHARACTERISTICS
- ❖ UNITWISE IMPORTANT QUESTIONS

PROBABILITY AND RANDOM PROCESSES

Course Objectives:

- To provide mathematical background and sufficient experience so that student can read, write and understand sentences in the language of probability theory.
- To introduce students to the basic methodology of “probabilistic thinking” and apply it to problems.
- To understand basic concepts of Probability theory and Random Variables, how to deal with multiple Random Variables.
- To understand the difference between time averages statistical averages.
- To teach students how to apply sums and integrals to compute probabilities, and expectations.

UNIT I:

Probability and Random Variable

Probability: Set theory, Experiments and Sample Spaces, Discrete and Continuous Sample Spaces, Events, Probability Definitions and Axioms, Joint Probability, Conditional Probability, Total Probability, Bayes’ Theorem, and Independent Events, Bernoulli’s trials.

The Random Variable: Definition of a Random Variable, Conditions for a Function to be a Random Variable, Discrete and Continuous.

UNIT II:

Distribution and density functions and Operations on One Random Variable

Distribution and density functions: Distribution and Density functions, Properties, Binomial, Uniform, Exponential, Gaussian, and Conditional Distribution and Conditional Density function and its properties, problems.

Operation on One Random Variable: Expected value of a random variable, function of a random variable, moments about the origin, central moments, variance and skew, characteristic function, moment generating function.

UNIT III:

Multiple Random Variables and Operations on Multiple Random Variables

Multiple Random Variables: Joint Distribution Function and Properties, Joint density Function and Properties, Marginal Distribution and density Functions, conditional Distribution and density Functions, Statistical Independence, Distribution and density functions of Sum of Two Random Variables.

Operations on Multiple Random Variables: Expected Value of a Function of Random Variables, Joint Moments about the Origin, Joint Central Moments, Joint Characteristic Functions, and Jointly Gaussian Random Variables: Two Random Variables case Properties.

UNIT IV:

Stochastic Processes-Temporal Characteristics: The Stochastic process Concept, Classification of Processes, Deterministic and Nondeterministic Processes, Distribution and Density Functions, Statistical Independence and concept of Stationarity: First-Order Stationary Processes, Second-Order and Wide-Sense Stationarity, Nth-Order and Strict-Sense Stationarity, Time Averages and Ergodicity, Mean-Ergodic Processes, Correlation-Ergodic Processes Autocorrelation Function and Its Properties, Cross-Correlation Function and Its Properties, Covariance Functions and its properties.

Linear system Response: Mean and Mean-squared value, Autocorrelation, Cross-Correlation Functions.

UNIT V:

Stochastic Processes-Spectral Characteristics: The Power Spectrum and its Properties, Relationship between Power Spectrum and Autocorrelation Function, the Cross-Power Density Spectrum and Properties, Relationship between Cross-Power Spectrum and Cross-Correlation Function.

Spectral characteristics of system response: power density spectrum of response, cross power spectral density of input and output of a linear system

TEXT BOOKS:

1. Probability, Random Variables & Random Signal Principles -Peyton Z. Peebles, TMH, 4th Edition, 2001.
2. Probability and Random Processes-Scott Miller, Donald Childers, 2Ed, Elsevier, 2012

REFERENCE BOOKS:

1. Theory of probability and Stochastic Processes-Pradip Kumar Gosh, University Press
2. Probability and Random Processes with Application to Signal Processing - Henry Stark and John W. Woods, Pearson Education, 3rd Edition.
3. Probability Methods of Signal and System Analysis- George R. Cooper, Clive D. MC Gillem, Oxford, 3rd Edition, 1999.
4. Statistical Theory of Communication -S.P. Eugene Xavier, New Age Publications 2003
5. Probability, Random Variables and Stochastic Processes Athanasios Papoulis and S.Unnikrishna Pillai, PHI, 4th Edition, 2002.

UNIT I

Probability and Random Variable

Probability:

- Set theory
- Experiments
- Sample Spaces, Discrete and Continuous Sample Spaces
- Events
- Probability Definitions and Axioms
- Joint Probability
- Conditional Probability
- Total Probability
- Bayes' Theorem
- Independent Events
- Bernoulli's trials

Random Variable:

- Definition of a Random Variable
- Conditions for a Function to be a Random Variable
- Discrete and Continuous Random Variables

UNIT – 1

PROBABILITY AND RANDOM VARIABLE

PROBABILITY

Introduction

It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge. Probability is simply how likely something is to happen. Whenever we're unsure about the outcome of an event, we can talk about the probabilities of certain outcomes —how likely they are. The analysis of events governed by probability is called statistics. Probability theory, a branch of mathematics concerned with the analysis of random phenomena. The outcome of a random event cannot be determined before it occurs, but it may be any one of several possible outcomes. The actual outcome is considered to be determined by chance.

How to Interpret Probability

Mathematically, the probability that an event will occur is expressed as a number between 0 and 1. Notationally, the probability of event A is represented by $P(A)$.

- If $P(A)$ equals zero, event A will almost definitely not occur.
- If $P(A)$ is close to zero, there is only a small chance that event A will occur.
- If $P(A)$ equals 0.5, there is a 50-50 chance that event A will occur.
- If $P(A)$ is close to one, there is a strong chance that event A will occur.
- If $P(A)$ equals one, event A will almost definitely occur.

In a statistical experiment, the sum of probabilities for all possible outcomes is equal to one. This means, for example, that if an experiment can have three possible outcomes (A, B, and C), then $P(A) + P(B) + P(C) = 1$.

Applications

Probability theory is applied in everyday life in risk assessment and in trade on financial markets. Governments apply probabilistic methods in environmental regulation, where it is called pathway analysis

Another significant application of probability theory in everyday life is reliability. Many consumer products, such as automobiles and consumer electronics, use reliability theory in product design to reduce the probability of failure. Failure probability may influence a manufacturer's decisions on a product's warranty.

The range of applications extends beyond games into business decisions, insurance, law, medical tests, and the social sciences. The telephone network, call centers, and airline companies with their randomly fluctuating loads could not have been economically designed without probability theory.

Uses of Probability in real life

- **Sports** – be it basketball or football or cricket a coin is tossed and both teams have 50/50 chances of winning it.
- **Board Games** – The odds of rolling one die and getting an even number there is a 50% chance since three of the six numbers on a die are even.
- **Medical Decisions** – When a patient is advised to undergo surgery, they often want to know the success rate of the operation which is nothing but a probability rate. Based on the same the patient takes a decision whether or not to go ahead with the same.
- **Life Expectancy** – this is based on the number of years the same groups of people have lived in the past.
- **Weather** – when planning an outdoor activity, people generally check the probability of rain. Meteorologists also predict the weather based on the patterns of the previous year, temperatures and natural disasters are also predicted on probability and nothing is ever stated as a surety but a possibility and an approximation.

SET THEORY:

Set: A set is a well defined collection of objects. These objects are called elements or members of the set. Usually uppercase letters are used to denote sets.

The set theory was developed by George Cantor in 1845-1918. Today, it is used in almost every branch of mathematics and serves as a fundamental part of present-day mathematics.

In everyday life, we often talk of the collection of objects such as a bunch of keys, flock of birds, pack of cards, etc. In mathematics, we come across collections like natural numbers, whole numbers, prime and composite numbers.

We assume that,

- the word set is synonymous with the word collection, aggregate, class and comprises of elements.
- Objects, elements and members of a set are synonymous terms.
- Sets are usually denoted by capital letters A, B, C,....., etc.
- Elements of the set are represented by small letters a, b, c, , etc.

If 'a' is an element of set A, then we say that 'a' belongs to A and it is mathematically represented as $a \in A$

If 'b' is an element which does not belong to A, we represent this as $b \notin A$.

Examples of sets:

1. Describe the set of vowels.

If A is the set of vowels, then A could be described as $A = \{a, e, i, o, u\}$.

2. Describe the set of positive integers.

Since it would be impossible to list *all* of the positive integers, we need to use a rule to describe this set. We might say A consists of all integers greater than zero.

3. Set $A = \{1, 2, 3\}$ and Set $B = \{3, 2, 1\}$. Is Set A equal to Set B?

Yes. Two sets are equal if they have the same elements. The order in which the elements are listed does not matter.

4. What is the set of men with four arms?

Since all men have two arms at most, the set of men with four arms contains no elements. It is the null set (or empty set).

5. Set $A = \{1, 2, 3\}$ and Set $B = \{1, 2, 4, 5, 6\}$. Is Set A a subset of Set B ?

Set A would be a subset of Set B if every element from Set A were also in Set B . However, this is not the case. The number 3 is in Set A , but not in Set B . Therefore, Set A is not a subset of Set B .

Some important sets used in mathematics are

N: the set of all natural numbers = $\{1, 2, 3, 4, \dots\}$

Z: the set of all integers = $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$

Q: the set of all rational numbers

R: the set of all real numbers

Z⁺: the set of all positive integers

W: the set of all whole numbers

Types of sets:

1. Empty Set or Null Set:

A set which does not contain any element is called an empty set, or the null set or the void set and it is denoted by $\emptyset$ and is read as phi. In roster form, $\emptyset$ is denoted by $\{\}$. An empty set is a finite set, since the number of elements in an empty set is finite, i.e., 0.

For example: (a) the set of whole numbers less than 0.

(b) Clearly there is no whole number less than 0.

Therefore, it is an empty set.

(c) $N = \{x : x \in N, 3 < x < 4\}$

• Let $A = \{x : 2 < x < 3, x \text{ is a natural number}\}$

Here A is an empty set because there is no natural number between 2 and 3

2. Singleton Set:

A set which contains only one element is called a singleton set.

For example:

- $A = \{x : x \text{ is neither prime nor composite}\}$

It is a singleton set containing one element, i.e., 1.

- $B = \{x : x \text{ is a whole number, } x < 1\}$

This set contains only one element 0 and is a singleton set.

- Let $A = \{x : x \in \mathbb{N} \text{ and } x^2 = 4\}$

Here A is a singleton set because there is only one element 2 whose square is 4.

- Let $B = \{x : x \text{ is an even prime number}\}$

Here B is a singleton set because there is only one prime number which is even, i.e., 2.

3. Finite Set:

A set which contains a definite number of elements is called a finite set. Empty set is also called a finite set.

For example:

- The set of all colors in the rainbow.
- $N = \{x : x \in \mathbb{N}, x < 7\}$
- $P = \{2, 3, 5, 7, 11, 13, 17, \dots, 97\}$

4. Infinite Set:

The set whose elements cannot be listed, i.e., set containing never-ending elements is called an infinite set.

For example:

- Set of all points in a plane
- $A = \{x : x \in \mathbb{N}, x > 1\}$
- Set of all prime numbers
- $B = \{x : x \in \mathbb{W}, x = 2n\}$

5. Cardinal Number of a Set:

The number of distinct elements in a given set A is called the cardinal number of A. It is denoted by $n(A)$. And read as ‘the number of elements of the set’.

For example:

- $A = \{x : x \in \mathbb{N}, x < 5\}$

$$A = \{1, 2, 3, 4\}$$

$$\text{Therefore, } n(A) = 4$$

- $B = \text{set of letters in the word ALGEBRA}$

$$B = \{A, L, G, E, B, R\}$$

$$\text{Therefore, } n(B) = 6$$

6. Equivalent Sets:

Two sets A and B are said to be equivalent if their cardinal number is same, i.e., $n(A) = n(B)$. The symbol for denoting an equivalent set is ‘ $\leftrightarrow$ ’.

For example:

$$A = \{1, 2, 3\} \text{ Here } n(A) = 3$$

$$B = \{p, q, r\} \text{ Here } n(B) = 3$$

$$\text{Therefore, } A \leftrightarrow B$$

7. Equal sets:

Two sets A and B are said to be equal if they contain the same elements. Every element of A is an element of B and every element of B is an element of A.

For example:

$$A = \{p, q, r, s\}$$

$$B = \{p, s, r, q\}$$

$$\text{Therefore, } A = B$$

8. Disjoint Sets:

Two sets A and B are said to be disjoint, if they do not have any element in common.

For example;

$$A = \{x : x \text{ is a prime number}\}$$

$$B = \{x : x \text{ is a composite number}\}.$$

Clearly, A and B do not have any element in common and are disjoint sets.

9. Overlapping sets:

Two sets A and B are said to be overlapping if they contain at least one element in common.

For example;

$$\bullet A = \{a, b, c, d\}$$

$$B = \{a, e, i, o, u\}$$

$$\bullet X = \{x : x \in \mathbb{N}, x < 4\}$$

$$Y = \{x : x \in \mathbb{I}, -1 < x < 4\}$$

Here, the two sets contain three elements in common, i.e., (1, 2, 3)

10. Definition of Subset:

If A and B are two sets, and every element of set A is also an element of set B, then A is called a subset of B and we write it as $A \subseteq B$ or $B \supseteq A$

The symbol $\subset$ stands for 'is a subset of' or 'is contained in'

- Every set is a subset of itself, i.e., $A \subset A$, $B \subset B$.
- Empty set is a subset of every set.
- Symbol ' $\subseteq$ ' is used to denote 'is a subset of' or 'is contained in'.
- $A \subseteq B$ means A is a subset of B or A is contained in B.
- $B \subseteq A$ means B contains A.

Examples;

1. Let $A = \{2, 4, 6\}$

$B = \{6, 4, 8, 2\}$

Here A is a subset of B

Since, all the elements of set A are contained in set B .

But B is not the subset of A

Since, all the elements of set B are not contained in set A .

2. The set N of natural numbers is a subset of the set Z of integers and we write $N \subset Z$.

3. Let $A = \{2, 4, 6\}$

$B = \{x : x \text{ is an even natural number less than } 8\}$

Here $A \subset B$ and $B \subset A$.

Hence, we can say $A = B$

4. Let $A = \{1, 2, 3, 4\}$

$B = \{4, 5, 6, 7\}$

Here $A \not\subset B$ and also $B \not\subset A$

[$\not\subset$ denotes 'not a subset of']

11. Proper Subset:

If A and B are two sets, then A is called the proper subset of B if $A \subseteq B$ but $A \neq B$. The symbol ' $\subset$ ' is used to denote proper subset. Symbolically, we write $A \subset B$.

For example;

1. $A = \{1, 2, 3, 4\}$

Here $n(A) = 4$

$B = \{1, 2, 3, 4, 5\}$

Here $n(B) = 5$

We observe that, all the elements of A are present in B but the element '5' of B is not present in A.

So, we say that A is a proper subset of B.

Symbolically, we write it as $A \subset B$

Notes:

No set is a proper subset of itself.

Null set or $\emptyset$ is a proper subset of every set.

2. $A = \{p, q, r\}$

$B = \{p, q, r, s, t\}$

Here A is a proper subset of B as all the elements of set A are in set B and also $A \neq B$.

Notes:

No set is a proper subset of itself.

Empty set is a proper subset of every set.

12. Universal Set

A set which contains all the elements of other given sets is called a **universal set**. The symbol for denoting a universal set is **U** or ξ .

For example;

1. If $A = \{1, 2, 3\}$ $B = \{2, 3, 4\}$ $C = \{3, 5, 7\}$

then $U = \{1, 2, 3, 4, 5, 7\}$

[Here $A \subseteq U$, $B \subseteq U$, $C \subseteq U$ and $U \supseteq A$, $U \supseteq B$, $U \supseteq C$]

2. If P is a set of all whole numbers and Q is a set of all negative numbers then the universal set is a set of all integers.

3. If $A = \{a, b, c\}$ $B = \{d, e\}$ $C = \{f, g, h, i\}$

then $U = \{a, b, c, d, e, f, g, h, i\}$ can be taken as universal set.

Operations on sets:

1. Definition of Union of Sets:

Union of two given sets is the set which contains all the elements of both the sets.

To find the union of two given sets A and B is a set which consists of all the elements of A and all the elements of B such that no element is repeated.

The symbol for denoting union of sets is '**U**'.

Some properties of the operation of union:

- | | |
|--|---|
| (i) $A \cup B = B \cup A$ | (Commutative law) |
| (ii) $A \cup (B \cap C) = (A \cup B) \cap C$ | (Associative law) |
| (iii) $A \cup \Phi = A$ | (Law of identity element, is the identity of U) |
| (iv) $A \cup A = A$ | (Idempotent law) |
| (v) $U \cup A = U$ | (Law of U) U is the universal set. |

Note:

$A \cup \Phi = \Phi \cup A = A$ i.e. union of any set with the empty set is always the set itself.

Examples:

1. If $A = \{1, 3, 7, 5\}$ and $B = \{3, 7, 8, 9\}$. Find union of two set A and B.

Solution:

$$A \cup B = \{1, 3, 5, 7, 8, 9\}$$

No element is repeated in the union of two sets. The common elements 3, 7 are taken only once.

2. Let $X = \{a, e, i, o, u\}$ and $Y = \{\phi\}$. Find union of two given sets X and Y.

Solution:

$$X \cup Y = \{a, e, i, o, u\}$$

Therefore, union of any set with an empty set is the set itself.

2. Definition of Intersection of Sets:

Intersection of two given sets is the set which contains all the elements that are common to both the sets.

To find the intersection of two given sets A and B is a set which consists of all the elements which are common to both A and B.

The symbol for denoting intersection of sets is ' $\cap$ '

Some properties of the operation of intersection

$$(i) A \cap B = B \cap A \quad (\text{Commutative law})$$

$$(ii) (A \cap B) \cap C = A \cap (B \cap C) \quad (\text{Associative law})$$

$$(iii) \Phi \cap A = \Phi \quad (\text{Law of } \Phi)$$

$$(iv) U \cap A = A \quad (\text{Law of } U)$$

$$(v) A \cap A = A \quad (\text{Idempotent law})$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (\text{Distributive law})$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad (\text{Distributive law})$$

Note:

$A \cap \Phi = \Phi \cap A = \Phi$ i.e. intersection of any set with the empty set is always the empty set.

Solved examples :

1. If $A = \{2, 4, 6, 8, 10\}$ and $B = \{1, 3, 8, 4, 6\}$. Find intersection of two set A and B.

Solution:

$$A \cap B = \{4, 6, 8\}$$

Therefore, 4, 6 and 8 are the common elements in both the sets.

2. If $X = \{a, b, c\}$ and $Y = \{\phi\}$. Find intersection of two given sets X and Y.

Solution:

$$X \cap Y = \{ \}$$

3. Difference of two sets

If A and B are two sets, then their difference is given by $A - B$ or $B - A$.

- If $A = \{2, 3, 4\}$ and $B = \{4, 5, 6\}$

$A - B$ means elements of A which are not the elements of B.

i.e., in the above example $A - B = \{2, 3\}$

- If A and B are disjoint sets, then $A - B = A$ and $B - A = B$

Solved examples to find the difference of two sets:

1. $A = \{1, 2, 3\}$ and $B = \{4, 5, 6\}$.

Find the difference between the two sets:

(i) A and B

(ii) B and A

Solution:

The two sets are disjoint as they do not have any elements in common.

(i) $A - B = \{1, 2, 3\} = A$

(ii) $B - A = \{4, 5, 6\} = B$

2. Let $A = \{a, b, c, d, e, f\}$ and $B = \{b, d, f, g\}$.

Find the difference between the two sets:

(i) A and B

(ii) B and A

Solution:

(i) $A - B = \{a, c, e\}$

Therefore, the elements a, c, e belong to A but not to B

(ii) $B - A = \{g\}$

Therefore, the element g belongs to B but not A .

4. Complement of a Set

In complement of a set if S be the universal set and A a subset of S then the complement of A is the set of all elements of S which are not the elements of A .

Symbolically, we denote the complement of A with respect to S as A^c or $\bar{A}$ or A'

Some properties of complement sets

(i) $A \cup A' = A' \cup A = U$ (Complement law)

(ii) $(A \cap B)' = \phi$ (Complement law) - The set and its complement are disjoint sets.

(iii) $(A \cup B)' = A' \cap B'$ (De Morgan's law)

(iv) $(A \cap B)' = A' \cup B'$ (De Morgan's law)

(v) $(A')' = A$ (Law of complementation)

(vi) $\Phi' = U$ (Law of empty set - The complement of an empty set is a universal set.

(vii) $U' = \Phi$ and universal set) - The complement of a universal set is an empty set.

For Example; If $S = \{1, 2, 3, 4, 5, 6, 7\}$

$A = \{1, 3, 7\}$ find A' .

Solution:

We observe that 2, 4, 5, 6 are the only elements of S which do not belong to A .

Therefore, $A' = \{2, 4, 5, 6\}$

Algebraic laws on sets:

1. Commutative Laws:

For any two finite sets A and B ;

(i) $A \cup B = B \cup A$

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

2. Associative Laws:

For any three finite sets A , B and C ;

(i) $(A \cup B) \cup C = A \cup (B \cup C)$

(ii) $(A \cap B) \cap C = A \cap (B \cap C)$

Thus, union and intersection are associative.

3. Idempotent Laws:

For any finite set A ;

(i) $A \cup A = A$

(ii) $A \cap A = A$

4. Distributive Laws:

For any three finite sets A , B and C ;

(i) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(ii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Thus, union and intersection are distributive over intersection and union respectively

5. De Morgan's Laws:

For any two finite sets A and B;

$$(i) A - (B \cup C) = (A - B) \cap (A - C)$$

$$(ii) A - (B \cap C) = (A - B) \cup (A - C)$$

De Morgan's Laws can also be written as:

$$(i) (A \cup B)' = A' \cap B'$$

$$(ii) (A \cap B)' = A' \cup B'$$

More laws of algebra of sets:

6. For any two finite sets A and B;

$$(i) A - B = A \cap B'$$

$$(ii) B - A = B \cap A'$$

$$(iii) A - B = A \Leftrightarrow A \cap B = \emptyset$$

$$(iv) (A - B) \cup B = A \cup B$$

$$(v) (A - B) \cap B = \emptyset$$

$$(vi) (A - B) \cup (B - A) = (A \cup B) - (A \cap B)$$

Definition of De Morgan's law:

The complement of the union of two sets is equal to the intersection of their complements and the complement of the intersection of two sets is equal to the union of their complements. These are called **De Morgan's laws**.

For any two finite sets A and B;

$$(i) (A \cup B)' = A' \cap B' \text{ (which is a De Morgan's law of union).}$$

$$(ii) (A \cap B)' = A' \cup B' \text{ (which is a De Morgan's law of intersection).}$$

Venn Diagrams:

Pictorial representations of sets represented by closed figures are called set diagrams or Venn diagrams.

Venn diagrams are used to illustrate various operations like union, intersection and difference.

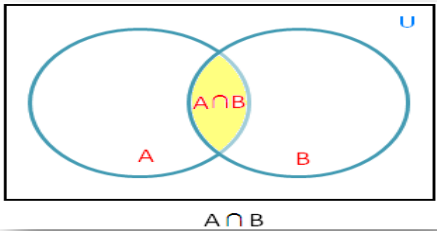
We can express the relationship among sets through this in a more significant way.

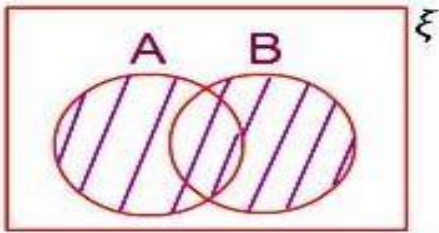
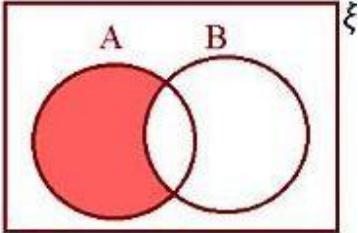
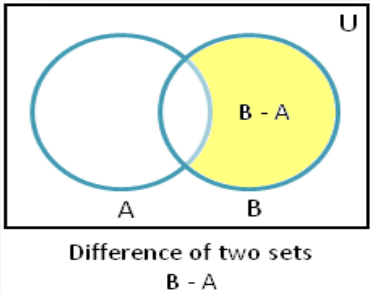
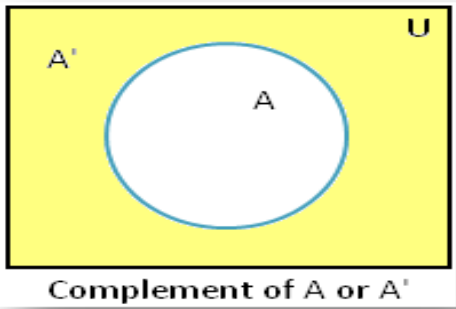
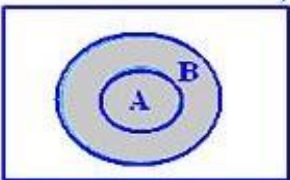
In this,

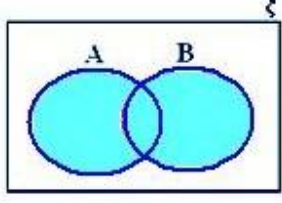
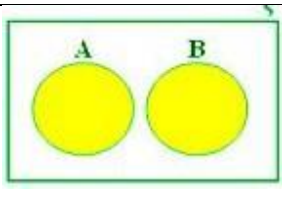
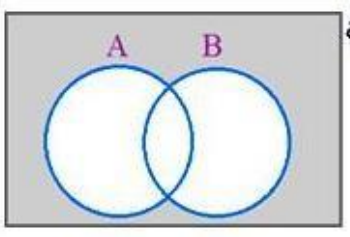
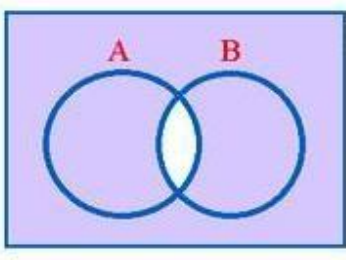
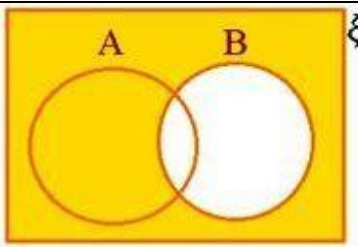
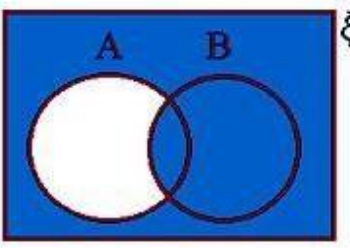
- A rectangle is used to represent a universal set.
- Circles or ovals are used to represent other subsets of the universal set.

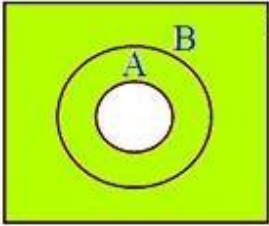
Venn diagrams in different situations

In these diagrams, the universal set is represented by a rectangular region and its subsets by circles inside the rectangle. We represented disjoint set by disjoint circles and intersecting sets by intersecting circles.

S.No	Set &Its relation	Venn Diagram
1	Intersection of A and B	

2	Union of A and B	
3	Difference : A-B	
4	Difference : B-A	
5	Complement of set A	
6	$A \cup B$ when $A \subset B$	

7	$A \cup B$ when neither $A \subset B$ nor $B \subset A$	
8	$A \cup B$ when A and B are disjoint sets	
9	$(A \cup B)'$ (A union B dash)	
10	$(A \cap B)'$ (A intersection B dash)	
11	B' (B dash)	
12	$(A - B)'$ (Dash of sets A minus B)	

13	$(A \subset B)'$ (Dash of A subset B)	
----	---------------------------------------	---

Problems of set theory:

1. Let A and B be two finite sets such that $n(A) = 20$, $n(B) = 28$ and $n(A \cup B) = 36$, find $n(A \cap B)$.

Solution:

Using the formula $n(A \cup B) = n(A) + n(B) - n(A \cap B)$.

then $n(A \cap B) = n(A) + n(B) - n(A \cup B)$

$$= 20 + 28 - 36$$

$$= 48 - 36$$

$$= 12$$

2. If $n(A - B) = 18$, $n(A \cup B) = 70$ and $n(A \cap B) = 25$, then find $n(B)$.

Solution:

Using the formula $n(A \cup B) = n(A - B) + n(A \cap B) + n(B - A)$

$$70 = 18 + 25 + n(B - A)$$

$$70 = 43 + n(B - A)$$

$$n(B - A) = 70 - 43$$

$$n(B - A) = 27$$

Now $n(B) = n(A \cap B) + n(B - A)$

$$= 25 + 27$$

$$= 52$$

3. In a group of 60 people, 27 like cold drinks and 42 like hot drinks and each person likes at least one of the two drinks. How many like both coffee and tea?

Solution:

Let A = Set of people who like cold drinks.

B = Set of people who like hot drinks.

Given

$$(A \cup B) = 60 \quad n(A) = 27 \quad n(B) = 42 \text{ then;}$$

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

$$= 27 + 42 - 60$$

$$= 69 - 60 = 9$$

$$= 9$$

Therefore, 9 people like both tea and coffee.

4. There are 35 students in art class and 57 students in dance class. Find the number of students who are either in art class or in dance class.

- When two classes meet at different hours and 12 students are enrolled in both activities.
- When two classes meet at the same hour.

Solution:

$$n(A) = 35, \quad n(B) = 57, \quad n(A \cap B) = 12$$

(Let A be the set of students in art class.
B be the set of students in dance class.)

$$(i) \text{ When 2 classes meet at different hours } n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

$$= 35 + 57 - 12$$

$$= 92 - 12$$

$$= 80$$

(ii) When two classes meet at the same hour, $A \cap B = \emptyset$ $n(A \cup B) = n(A) + n(B) - n(A \cap B)$

$$= n(A) + n(B)$$

$$= 35 + 57$$

$$= 92$$

5. In a group of 100 persons, 72 people can speak English and 43 can speak French. How many can speak English only? How many can speak French only and how many can speak both English and French?

Solution:

Let A be the set of people who speak English.

B be the set of people who speak French.

A - B be the set of people who speak English and not French.

B - A be the set of people who speak French and not English.

$A \cap B$ be the set of people who speak both French and English.

Given,

$$n(A) = 72 \quad n(B) = 43 \quad n(A \cup B) = 100$$

$$\text{Now, } n(A \cap B) = n(A) + n(B) - n(A \cup B)$$

$$= 72 + 43 - 100$$

$$= 115 - 100$$

$$= 15$$

Therefore, Number of persons who speak both French and English = 15

$$n(A) = n(A - B) + n(A \cap B)$$

$$\Rightarrow n(A - B) = n(A) - n(A \cap B)$$

$$= 72 - 15$$

$$= 57$$

$$\text{and } n(B - A) = n(B) - n(A \cap B)$$

$$= 43 - 15$$

$$= 28$$

Therefore, Number of people speaking English only = 57

Number of people speaking French only = 28

Probability Concepts

Before we give a definition of probability, let us examine the following concepts:

1. Experiment:

In probability theory, an **experiment** or **trial** (see below) is any procedure that can be infinitely repeated and has a well-defined set of possible outcomes, known as the sample space. An experiment is said to be *random* if it has more than one possible outcome, and *deterministic* if it has only one. A random experiment that has exactly two (mutually exclusive) possible outcomes is known as a Bernoulli trial.

Random Experiment:

An experiment is a random experiment if its outcome cannot be predicted precisely. One out of a number of outcomes is possible in a random experiment. A single performance of the random experiment is called a **trial**.

Random experiments are often conducted repeatedly, so that the collective results may be subjected to statistical analysis. A fixed number of repetitions of the same experiment can be thought of as a **composed experiment**, in which case the individual repetitions are called **trials**. For example, if one were to toss the same coin one hundred times and record each result, each toss would be considered a trial within the experiment composed of all hundred tosses.

Mathematical description of an experiment:

A random experiment is described or modeled by a mathematical construct known as a probability space. A probability space is constructed and defined with a specific kind of experiment or trial in mind.

A mathematical description of an experiment consists of three parts:

1. A sample space, Ω (or S), which is the set of all possible outcomes.
2. A set of events, where each event is a set containing zero or more outcomes.
3. The assignment of probabilities to the events—that is, a function P mapping from events to probabilities.

An *outcome* is the result of a single execution of the model. Since individual outcomes might be of little practical use, more complicated *events* are used to characterize groups of outcomes. The collection of all such events is a *sigma-algebra*. Finally, there is a need to specify each event's likelihood of happening; this is done using the *probability measure* function, P .

2. **Sample Space:** The sample space S is the collection of all possible outcomes of a random experiment. The elements of S are called **sample points**.

- A sample space may be *finite*, *countably infinite* or *uncountable*.
- A finite or countably infinite sample space is called a *discrete sample space*.
- An uncountable sample space is called a *continuous sample space*.

Ex:1. For the coin-toss experiment would be the results –Head and –Tail, which we may represent by $S = \{H, T\}$.

Ex. 2. If we toss a die, one sample space or the set of all possible outcomes is

$$S = \{1, 2, 3, 4, 5, 6\}$$

The other sample space can be

$$S = \{\text{odd}, \text{even}\}$$

Types of Sample Space:

1. Finite/Discrete Sample Space:

Consider the experiment of tossing a coin twice.

The sample space can be

$S = \{HH, HT, TH, TT\}$ the above sample space has a finite number of sample points. It is called a finite sample space

2. Countably Infinite Sample Space:

Consider that a light bulb is manufactured. It is then tested for its life length by inserting it into a socket and the time elapsed (in hours) until it burns out is recorded. Let the measuring instrument is capable of recording time to two decimal places, for example 8.32 hours.

Now, the sample space becomes countably infinite i.e.

$$S = \{0.0, 0.01, 0.02\}$$

The above sample space is called a countable infinite sample space.

3. Un Countable/ Infinite Sample Space:

If the sample space consists of unaccountably infinite number of elements then it is called

Un Countable/ Infinite Sample Space.

3. **Event:** An event is simply a set of possible outcomes. To be more specific, an event is a subset A of the sample space S .

- $A \subseteq S$
- For a discrete sample space, all subsets are events.

Ex: For instance, in the coin-toss experiment the events $A=\{\text{Heads}\}$ and $B=\{\text{Tails}\}$ would be mutually exclusive.

An event consisting of a single point of the sample space ' S ' is called a simple event or elementary event.

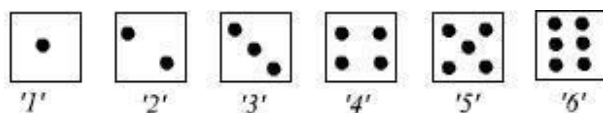
Some examples of event sets:

Example 1: tossing a fair coin

The possible outcomes are **H (head)** and **T (tail)**. The associated sample space is $S = \{H, T\}$. It is a finite sample space. The events associated with the sample space S are: $S, \{H\}, \{T\}$ and $\emptyset$.

Example 2: Throwing a fair die:

The possible 6 outcomes are:



The associated finite sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Some events are

A = The event of getting an odd face = { '1', '3', '5' }.

B = The event of getting a six = { '6' }

And so on.

Example 3: Tossing a fair coin until a head is obtained

We may have to toss the coin any number of times before a head is obtained. Thus the possible outcomes are:

H, TH, TTH, TTTH,

How many outcomes are there? The outcomes are countable but infinite in number. The countably infinite sample space is $S = \{H, TH, TTH, \dots\}$.

Example 4 : Picking a real number at random between -1 and +1

The associated Sample space is

$$S = \{s \mid s \in \mathbb{R}, -1 \leq s \leq 1\} = [-1, 1]$$

Clearly S is a continuous sample space.

Example 5: Drawing cards

Drawing 4 cards from a deck: Events include all spades, sum of the 4 cards is (assuming face cards have a value of zero), a sequence of integers, a hand with a 2, 3, 4 and 5. There are many more events.

Types of Events:

1. Exhaustive Events:

A set of events is said to be exhaustive, if it includes all the possible events.

Ex. In tossing a coin, the outcome can be either Head or Tail and there is no other possible outcome. So, the set of events { H , T } is exhaustive.

2. Mutually Exclusive Events:

Two events, A and B are said to be mutually exclusive if they cannot occur together.

i.e. if the occurrence of one of the events precludes the occurrence of all others, then such a set of events is said to be mutually exclusive.

If two events are mutually exclusive then the probability of either occurring is

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B).$$

Ex. In tossing a die, both head and tail cannot happen at the same time.

3. Equally Likely Events:

If one of the events cannot be expected to happen in preference to another, then such events are said to be Equally Likely Events. (Or) Each outcome of the random experiment has an equal chance of occurring.

Ex. In tossing a coin, the coming of the head or the tail is equally likely.

4. Independent Events:

Two events are said to be independent, if happening or failure of one does not affect the happening or failure of the other. Otherwise, the events are said to be dependent.

If two events, A and B are independent then the joint probability is

$$P(A \text{ and } B) = P(A \cap B) = P(A)P(B),$$

5. Non-. Mutually Exclusive Events:

If the events are not mutually exclusive then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Probability Definitions and Axioms:

1. Relative frequency Definition:

Consider that an experiment E is repeated n times, and let A and B be two events associated with E . Let n_A and n_B be the number of times that the event A and the event B occurred among the n repetitions respectively.

The relative frequency of the event A in the ' n ' repetitions of E is defined as

$$f(A) = n_A / n$$

$f(A) = n_A / n$

The Relative frequency has the following properties:

$$0 \leq f(A) \leq 1$$

2. $f(A) = 1$ if and only if A occurs every time among the n repetitions.

If an experiment is repeated n times under similar conditions and the event A occurs in n_A times, then the probability of the event A is defined as

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

Limitation:

Since we can never repeat an experiment or process indefinitely, we can never know the probability of any event from the relative frequency definition. In many cases we can't even obtain a long series of repetitions due to time, cost, or other limitations. For example, the probability of rain today can't really be obtained by the relative frequency definition since today can't be repeated again.

2. The classical definition:

Let the sample space (denoted by S) be the set of all possible distinct outcomes to an experiment. The probability of some event is

$$\frac{\text{number of ways the event can occur}}{\text{number of outcomes in } S},$$

provided all points in S are equally likely. For example, when a die is rolled the probability of getting a 2 is $\frac{1}{6}$ because one of the six faces is a 2.

Limitation:

What does "equally likely" mean? This appears to use the concept of probability while trying to define it! We could remove the phrase "provided all outcomes are equally likely", but then the definition would clearly be unusable in many settings where the outcomes in S did not tend to occur equally often.

Example1: A fair die is rolled once. What is the probability of getting a '6' ?

Here $S = \{ '1', '2', '3', '4', '5', '6' \}$ and $A = \{ '6' \}$

$$\therefore N = 6 \text{ and } N_A = 1$$

$$\therefore P(A) = \frac{1}{6}$$

Example2: A fair coin is tossed twice. What is the probability of getting two 'heads'?

Here $S = \{ HH, TH, HT, TT \}$ and $A = \{ HH \}$.

Total number of outcomes is 4 and all four outcomes are equally likely.

Only outcome favourable to A is $\{ HH \}$

$$\therefore P(A) = \frac{1}{4}$$

Probability axioms:

Given an event E in a sample space S which is either finite with N elements or countably infinite with $N = \infty$ elements, then we can write

$$S \equiv \left(\bigcup_{i=1}^N E_i \right),$$

and a quantity $P(E_i)$, called the probability of event E_i , is defined such that

Axiom1: The probability of any event A is positive or zero. Namely $P(A) \geq 0$. The probability measures, in a certain way, the difficulty of event A happening: the smaller the probability, the more difficult it is to happen. i.e

$$. 0 \leq P(E_i) \leq 1$$

Axiom2: The probability of the sure event is 1. Namely $P(\Omega)=1$. And so, the probability is always greater than 0 and smaller than 1: probability zero means that there is no possibility for it to happen (it is an impossible event), and probability 1 means that it will always happen (it is a sure event).i.e

$$P(S) = 1.$$

Axiom3: The probability of the union of any set of two by two incompatible events is the sum of the probabilities of the events. That is, if we have, for example, events A, B, C , and these are two by two incompatible, then $P(A \cup B \cup C) = P(A) + P(B) + P(C)$. i.e Additivity:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2), \text{ where } E_1 \text{ and } E_2 \text{ are mutually exclusive.}$$

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i) \text{ for } n = 1, 2, \dots, N \text{ where } E_1, E_2, \dots \text{ are mutually exclusive (i.e., } E_1 \cap E_2 = \emptyset).$$

Main properties of probability: If A is any event of sample space S then

1. $P(A) + P(\bar{A}) = 1$
2. Since $A \cup \bar{A} = S$, $P(A \cup \bar{A}) = 1$
3. The probability of the impossible event is 0, i.e $P(\emptyset) = 0$
4. If $A \subset B$, then $P(A) \leq P(B)$.
5. If A and B are two incompatible events, and therefore, $P(A - B) = P(A) - P(A \cap B)$. and $P(B - A) = P(B) - P(A \cap B)$.
6. Addition Law of probability:
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

PERMUTATIONS and COMBINATIONS:

S.No.	PERMUTATIONS	COMBINATIONS:
1	Arrangement of things in a specified order is called permutation. Here all things are taken at a time	In permutations, the order of arrangement of objects is important. But, in combinations, order is not important, but only selection of objects.
2	Arrangement of 'r' things taken at a time from 'n' things ,where $r < n$ in a specified order in called r-permutation.	
3	Consider the letters a,b and c . Considering all the three letters at a time, the possible permutations are ABC ,a c b , b c a , b a c , c b a and c a b .	
4	The number of permutations taking r things at a time from 'n' available things is denoted as $p(n, r)$ or $n P_r$	The number of combinations taking r things at a time from 'n' available things is denoted as $C(n, r)$ or $n C_r$
5	$n P_r = \frac{n!}{n-r!}$	$n C_r = \frac{n!}{r!(n-r)!}$

Joint Probability:

Joint probability is the likelihood of more than one event occurring at the same time.

If a sample space consists of two events A and B which are not mutually exclusive, and then the probability of these events occurring jointly or simultaneously is called the Joint Probability. In other words the joint probability of events A and B is equal to the relative frequency of the joint occurrence.

For 2 events A and B it is denoted by $P(A \cap B)$ and joint probability in terms of conditional probability is

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

If A and B are independent events then

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

If A and B are mutually exclusive events then

$$P(A \cap B) = 0$$

From addition law joint probability of two events can be represented as

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

Conditional probability

The answer is the *conditional probability of B given A* denoted by $P(B/A)$. We shall develop the concept of the conditional probability and explain under what condition this conditional probability is same as $P(B)$.

Notation $P(B/A)$ = Conditional probability of B given A

Let us consider the case of *equiprobable* events discussed earlier. Let N_{AB} sample points be favourable for the joint event $A \cap B$

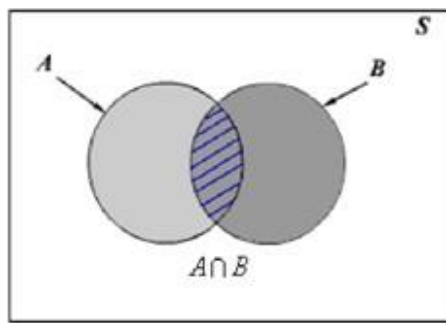


Figure 1

$$\begin{aligned}
 P(B/A) &= \frac{\text{Number of outcomes favourable to A and B}}{\text{Number of outcomes in A}} \\
 &= \frac{n(AB)}{n(A)} = \frac{\frac{n(AB)}{n}}{\frac{n(A)}{n}} = \frac{P(A \cap B)}{P(A)}
 \end{aligned}$$

This concept suggests us to define conditional probability. The probability of an event B under the condition that another event A has occurred is called the *conditional probability of B given A* and defined by

$$P(B/A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) \neq 0$$

We can similarly define the *conditional probability of A given B*, denoted by $P(A/B)$. From the definition of conditional probability, we have the joint probability $P(A \cap B)$ of two events A and B as follows

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

Example 1 Consider the example tossing the fair die. Suppose

A = event of getting an even number = $\{2, 4, 6\}$

B = event of getting a number less than 4 = $\{1, 2, 3\}$

$\therefore A \cap B = \{2\}$

$$\therefore P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/6}{3/6} = \frac{1}{3}$$

Example 2 A family has two children. It is known that at least one of the children is a girl. What is the probability that both the children are girls?

A = event of at least one girl

B = event of two girl

$S = \{gg, gb, bg, bb\}$, $A = \{gg, gb, bg\}$ and $B = \{gg\}$

$A \cap B = \{gg\}$

$$\therefore P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Properties of Conditional probability:

1. If $B \subseteq A$, then $P(B/A) = 1$ and $P(A/B) \geq P(A)$

We have, $A \cap B = B$

$$\therefore P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

and

$$\begin{aligned} P(A/B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(A)P(B/A)}{P(B)} \\ &= \frac{P(A)}{P(B)} \\ &\geq P(A) \end{aligned}$$

2. Since $P(A \cap B) \geq 0, P(A) > 0$

$$\therefore P(B/A) = \frac{P(A \cap B)}{P(A)} \geq 0$$

3. We have, $\therefore P(S/A) = \frac{P(S \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$

4. Chain Rule of Probability/Multiplication theorem:

$$P(A_1 \cap A_2 \dots A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \dots P(A_n/A_1 \cap A_2 \dots \cap A_{n-1})$$

We have,

$$\begin{aligned} (A \cap B \cap C) &= (A \cap B) \cap C \\ P(A \cap B \cap C) &= P(A \cap B)P(C/A \cap B) \\ &= P(A)P(B/A)P(C/A \cap B) \end{aligned}$$

$$\therefore P(A \cap B \cap C) = P(A)P(B/A)P(C/A \cap B)$$

We can generalize the above to get the *chain rule of probability for n events as*

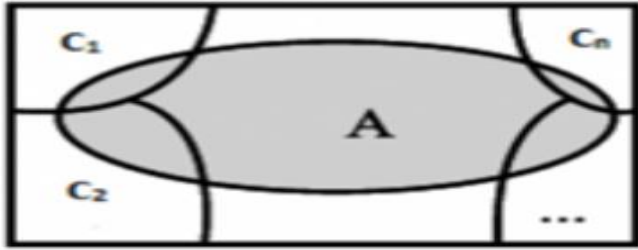
$$P(A_1 \cap A_2 \dots A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \dots P(A_n/A_1 \cap A_2 \dots \cap A_{n-1})$$

Total Probability theorem:

Statement:

Let events $C_1, C_2 \dots C_n$ form partitions of the sample space S , where all the events have non- zero probability of occurrence. For any event, A associated with S , according to the total probability theorem, $P(A)$

$$P(A) = \sum_{k=0}^n P(C_k)P(A|C_k)$$



Proof: From the figure 2, $\{C_1, C_2, \dots, C_n\}$ is the partitions of the sample space S such that, $C_i \cap C_k = \emptyset$, where $i \neq k$ and $i, k = 1, 2, \dots, n$ also all the events $C_1, C_2 \dots C_n$ have non zero probability. Sample space S can be given as,

$$S = C_1 \cup C_2 \cup \dots \cup C_n$$

For any event A , $A = A \cap S$

$$\begin{aligned} &= A \cap (C_1 \cup C_2 \cup \dots \cup C_n) \\ &= (A \cap C_1) \cup (A \cap C_2) \cup \dots \cup (A \cap C_n) \dots \dots (1) \end{aligned}$$

Here, C_i and C_k are disjoint for $i \neq k$. since they are mutually independent events which imply that $A \cap C_i$ and $A \cap C_k$ are also disjoint for all $i \neq k$. Thus,

$$\begin{aligned} P(A) &= P[(A \cap C_1) \cup (A \cap C_2) \cup \dots \cup (A \cap C_n)] \\ &= P(A \cap C_1) + P(A \cap C_2) + \dots + P(A \cap C_n) \dots \dots (2) \end{aligned}$$

We know that,

$$P(A \cap C_i) = P(C_i) P(A|C_i) \text{ (By multiplication rule of probability) } \dots (3)$$

Using (2) and (3), (1) can be rewritten as,

$$P(A) = P(C_1)P(A|C_1) + P(C_2)P(A|C_2) + P(C_3)P(A|C_3) + \dots + P(C_n)P(A|C_n)$$

Hence, the theorem can be stated in form of equation as,

$$P(A) = \sum_{k=0}^n P(C_k)P(A|C_k)$$

Example:

A person has undertaken a mining job. The probabilities of completion of job on time with and without rain are 0.42 and 0.90 respectively. If the probability that it will rain is 0.45, then determine the probability that the mining job will be completed on time.

Solution:

Let A be the event that the mining job will be completed on time and B be the event that it rains. We have,

$$P(B) = 0.45,$$

$$P(\text{no rain}) = P(B') = 1 - P(B) = 1 - 0.45 = 0.55$$

From given data

$$P(A|B) = 0.42$$

$$P(A|B') = 0.90$$

Since, events B and B' form partitions of the sample space S, by total probability theorem, we have

$$P(A) = P(B) P(A|B) + P(B') P(A|B')$$

$$= 0.45 \times 0.42 + 0.55 \times 0.9$$

$$= 0.189 + 0.495 = 0.684$$

So, the probability that the job will be completed on time is 0.684.

Bayes' Theorem:

Statement:

Let $\{E_1, E_2, \dots, E_n\}$ be a set of events associated with a sample space S , where all the events $E_1, E_2, \dots, E_n$ have nonzero probability of occurrence and they form a partition of S . Let A be any event associated with S , then according to Bayes theorem,

$$P(E_i | A) = \frac{P(E_i)P(A|E_i)}{\sum_{k=1}^n P(E_k)P(A|E_k)}$$

Proof: According to conditional probability formula,

$$P(E_i | A) = P(E_i \cap A) / P(A) \dots\dots\dots(1)$$

Using multiplication rule of probability,

$$P(E_i \cap A) = P(E_i)P(A|E_i) \dots\dots\dots(2)$$

Using total probability theorem

$$P(A) = \sum_{k=1}^n P(E_k)P(A|E_k) \dots\dots\dots(3)$$

Substitute 2 and 3 in 1 then will get

$$P(E_i | A) = \frac{P(E_i)P(A|E_i)}{\sum_{k=1}^n P(E_k)P(A|E_k)}$$

Example1:

In a binary communication system a zero and a one is transmitted with probability 0.6 and 0.4 respectively. Due to error in the communication system a zero becomes a one with a probability 0.1 and a one becomes a zero with a probability 0.08. Determine the probability (i) of receiving a one and (ii) that a one was transmitted when the received message is one

Solution:

Let S is the sample space corresponding to binary communication. Suppose T_0 be event of

Transmitting 0 and T_1 be the event of transmitting 1 and R_0 and R_1 be corresponding events of receiving 0 and 1 respectively.

Given $P(T_0) = 0.6$, $P(T_1) = 0.4$, $P(R_1 / T_0) = 0.1$ and $P(R_0 / T_1) = 0.08$.

$$\begin{aligned} \text{(i) } P(R_1) &= \text{Probability of receiving 'one'} \\ &= P(T_1)P(R_1 / T_1) + P(T_0)P(R_1 / T_0) \\ &= 0.4 \times 0.92 + 0.6 \times 0.1 \\ &= 0.448 \end{aligned}$$

(ii) Using the Baye's rule

$$\begin{aligned} P(T_1 / R_1) &= \frac{P(T_1)P(R_1 / T_1)}{P(R_1)} \\ &= \frac{P(T_1)P(R_1 / T_1)}{P(T_1)P(R_1 / T_1) + P(T_0)P(R_1 / T_0)} \\ &= \frac{0.4 \times 0.92}{0.4 \times 0.92 + 0.6 \times 0.1} \\ &= 0.8214 \end{aligned}$$

Example 2: In an electronics laboratory, there are identically looking capacitors of three makes A_1 , A_2 and A_3 in the ratio 2:3:4. It is known that 1% of A_1 , 1.5% of A_2 and 2% of A_3 are defective. What percentages of capacitors in the laboratory are defective? If a capacitor picked at defective is found to be defective, what is the probability it is of make A_3 ?

Let D be the event that the item is defective. Here we have to find $P(D)$ and $P(A_3 / D)$.

$$\begin{aligned} P(A_1) &= \frac{2}{9}, P(A_2) = \frac{1}{3} \text{ and } P(A_3) = \frac{4}{9} \\ P(D / A_1) &= 0.01, P(D / A_2) = 0.015 \text{ and } P(D / A_3) = 0.02 \end{aligned}$$

$$\begin{aligned} \therefore P(D) &= P(A_1)P(D / A_1) + P(A_2)P(D / A_2) + P(A_3)P(D / A_3) \\ &= \frac{2}{9} \times 0.01 + \frac{1}{3} \times 0.015 + \frac{4}{9} \times 0.02 \\ &= 0.0167 \end{aligned}$$

and

$$\begin{aligned} P(A_3 / D) &= \frac{P(A_3)P(D / A_3)}{P(D)} \\ &= \frac{\frac{4}{9} \times 0.02}{0.0167} \\ &= 0.533 \end{aligned}$$

Independent events

Two events are called *independent* if the probability of occurrence of one event does not affect the probability of occurrence of the other. Thus the events A and B are independent if

$$P(B|A) = P(B) \text{ and } P(A|B) = P(A)$$

where $P(A)$ and $P(B)$ are assumed to be non-zero.

Equivalently if A and B are independent, we have

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

or

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

Joint probability is the product of individual probabilities.

Two events A and B are called statistically *dependent* if they are not independent. Similarly, we can define the independence of n events. The events $A_1, A_2, \dots, A_n$ are called independent if and only if

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$$

$$P(A_i \cap A_j \cap A_k \cap \dots A_n) = P(A_i)P(A_j)P(A_k) \dots P(A_n)$$

Example: Consider the example of tossing a fair coin twice. The resulting sample space is given by $S = \{HH, HT, TH, TT\}$ and all the outcomes are equiprobable.

Let $A = \{TH, TT\}$ be the event of getting _tail' in the first toss and $B = \{TH, HH\}$ be the event of getting _head' in the second toss. Then

$$P(A) = \frac{1}{2} \text{ and } P(B) = \frac{1}{2}$$

Again, $(A \cap B) = \{TH\}$ so that

$$P(A \cap B) = \frac{1}{4} = P(A)P(B)$$

Hence the events A and B are independent.

Problems:

Example 1. A dice of six faces is tailored so that the probability of getting every face is proportional to the number depicted on it.

a) What is the probability of extracting a 6?

In this case, we say that the probability of each face turning up is not the same, therefore we cannot simply apply the rule of Laplace. If we follow the statement, it says that the probability of each face turning up is proportional to the number of the face itself, and this means that, if we say that the probability of face 1 being turned up is k which we do not know, then:

$$P(\{1\})=k, P(\{2\})=2k, P(\{3\})=3k, P(\{4\})=4k,$$

$$P(\{5\})=5k, P(\{6\})=6k.$$

Now, since $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ form an events complete system, necessarily

$$P(\{1\})+P(\{2\})+P(\{3\})+P(\{4\})+P(\{5\})+P(\{6\})=1$$

Therefore

$$k+2k+3k+4k+5k+6k=1$$

which is an equation that we can already solve:

$$21k=1$$

thus

$$k=1/21$$

And so, the probability of extracting 6 is $P(\{6\})=6k=6 \cdot (1/21)=6/21$.

b) What is the probability of extracting an odd number?

The cases favourable to event A = "to extract an odd number" are: $\{1\}, \{3\}, \{5\}$. Therefore, since they are incompatible events,

$$P(A)=P(\{1\})+P(\{3\})+P(\{5\})=k+3k+5k=9k=9 \cdot (1/21)=9/21$$

Example2: Roll a red die and a green die. Find the probability the total is 5.

Solution: Let (x,y) represent getting x on the red die and y on the green die.

$$S = \{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & \cdots & (1,6) \\ (2,1) & (2,2) & (2,3) & \cdots & (2,6) \\ (3,1) & (3,2) & (3,3) & \cdots & (3,6) \\ & \text{---} & \text{---} & \text{---} & \\ (6,1) & (6,2) & (6,3) & \cdots & (6,6) \end{array} \}$$

Then, with these as simple events, the sample space is

The sample points giving a total of 5 are (1,4) (2,3) (3,2), and (4,1).

$$\text{Therefore } P(\text{total is 5}) = \frac{4}{36}$$

Example3: Suppose the 2 dice were now identical red dice. Find the probability the total is 5.

Solution : Since we can no longer distinguish between (x,y) and (y,x) , the only distinguishable points in S are

$$S = \{ \begin{array}{cccc} (1,1) & (1,2) & (1,3) & \cdots & (1,6) \\ & (2,2) & (2,3) & \cdots & (2,6) \\ & & (3,3) & \cdots & (3,6) \\ & & & \cdots & \cdots \\ & & & & (6,6) \end{array} \} :$$

Using this sample space, we get a total of 5 from points (1,4) and (2,3) only. If we assign equal probability $\frac{1}{21}$ to each point (simple event) then we get $P(\text{total is 5}) = \frac{2}{21}$.

Example4: Draw 1 card from a standard well-shuffled deck (13 cards of each of 4 suits - spades, hearts, diamonds, and clubs). Find the probability the card is a club.

Solution 1: Let $S = \{ \text{spade, heart, diamond, club} \}$. (The points of S are generally listed between brackets { }.) Then S has 4 points, with 1 of them being "club", so $P(\text{club}) = \frac{1}{4}$.

Solution 2: Let $S = \{ \text{each of the 52 cards} \}$. Then 13 of the 52 cards are clubs, so

$$P(\text{club}) = \frac{13}{52} = \frac{1}{4}.$$

Example 5: Suppose we draw a card from a deck of playing cards. What is the probability that we draw a spade?

Solution: The sample space of this experiment consists of 52 cards, and the probability of each sample point is $1/52$. Since there are 13 spades in the deck, the probability of drawing a spade is

$$P(\text{Spade}) = (13)(1/52) = 1/4$$

Example 6: Suppose a coin is flipped 3 times. What is the probability of getting two tails and one head?

Solution: For this experiment, the sample space consists of 8 sample points.

$$S = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

Each sample point is equally likely to occur, so the probability of getting any particular sample point is $1/8$. The event "getting two tails and one head" consists of the following subset of the sample space.

$$A = \{TTH, THT, HTT\}$$

The probability of Event A is the sum of the probabilities of the sample points in A. Therefore,

$$P(A) = 1/8 + 1/8 + 1/8 = 3/8$$

Example7: An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn *without replacement* from the urn. What is the probability that both of the marbles are black?

Solution: Let A = the event that the first marble is black; and let B = the event that the second marble is black. We know the following:

- In the beginning, there are 10 marbles in the urn, 4 of which are black. Therefore, $P(A) = 4/10$.
- After the first selection, there are 9 marbles in the urn, 3 of which are black. Therefore, $P(B|A) = 3/9$.

Therefore, based on the rule of multiplication:

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

$$P(A \cap B) = (4/10) * (3/9) = 12/90 = 2/15$$

RANDOM VARIABLE

INTRODUCTION

In many situations, we are interested in numbers associated with the outcomes of a random experiment. In application of probabilities, we are often concerned with numerical values which are random in nature. For example, we may consider the number of customers arriving at a service station at a particular interval of time or the transmission time of a message in a communication system. These random quantities may be considered as real-valued function on the sample space. Such a real-valued function is called real random variable and plays an important role in describing random data. We shall introduce the concept of random variables in the following sections.

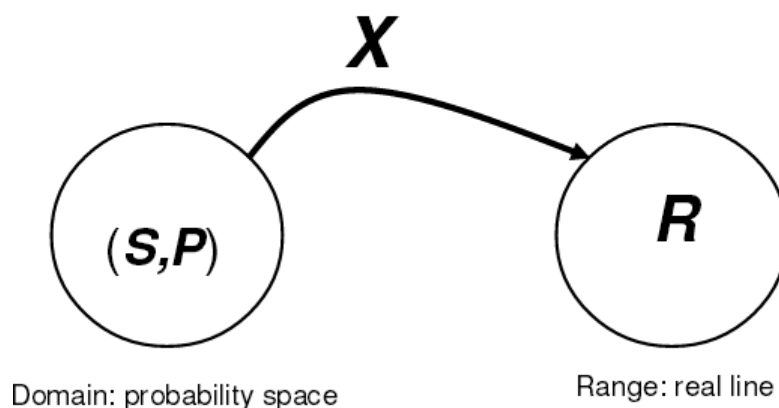
Random Variable Definition

A random variable is a function that maps outcomes of a random experiment to real numbers. (or)

A random variable associates the points in the sample space with real numbers

A (real-valued) random variable, often denoted by X (or some other capital letter), is a function mapping a probability space $(S; P)$ into the real line R . This is shown in Figure 1. Associated with each point s in the domain S the function X assigns one and only one value $X(s)$ in the range R . (The set of possible values of $X(s)$ is usually a proper subset of the real line; i.e., not all real numbers need occur. If S is a finite set with m elements, then $X(s)$ can assume at most an m different value as s varies in S .)

A random variable: a function



Example1

A fair coin is tossed 6 times. The number of heads that come up is an example of a random variable.

HHTTHT – 3, THHTTT -- 2.

These random variables can only take values between 0 and 6.

The Set of possible values of random variables is known as its

Range. **Example2**

A box of 6 eggs is rejected once it contains one or more broken eggs. If we examine 10 boxes of eggs and define the random variables X_1 , X_2 as

1 X_1 - the number of broken eggs in the 10

boxes 2 X_2 - the number of boxes rejected

Then the range of X_1 is $\{0, 1, 2, 3, 4, \dots, 60\}$ and X_2 is $\{0, 1, 2, \dots, 10\}$

Figure 2: A (real-valued) function of a random variable is itself a random variable, i.e., a function mapping a probability space into the real line.

Example 3 Consider the example of tossing a fair coin twice. The sample space is $S = \{HH, HT, TH, TT\}$ and all four outcomes are equally likely. Then we can define a random variable X as follows

Sample Point	Value of the random Variable
HH	0
HT	1
TH	2
TT	3

Here $R_X = \{0, 1, 2, 3\}$.

Example 4 Consider the sample space associated with the single toss of a fair die. The sample space is given by $S = \{1, 2, 3, 4, 5, 6\}$.

If we define the random variable X that associates a real number equal to the number on the face of the die, then $R_X = \{1, 2, 3, 4, 5, 6\}$.

Types of random variables:

There are two types of random variables, *discrete* and *continuous*.

1. Discrete random variable:

A **discrete random variable** is one which may take on only a countable number of distinct values such as 0, 1, 2, 3, 4, Discrete random variables are usually (but not necessarily) counts. If a random variable can take only a finite number of distinct values, then it must be discrete
(Or)

A random variable X is called a **discrete random variable** if $F_X(x)$ is piece-wise constant. Thus $F_X(x)$ is flat except at the points of jump discontinuity. If the sample space S is discrete the random variable X defined on it is always discrete.

• A discrete random variable has a finite number of possible values or an infinite sequence of countable real numbers.

– X: number of hits when trying 20 free throws.

– X: number of customers who arrive at the bank from 8:30 – 9:30 AM Mon--Fri.

– E.g. Binomial, Poisson...

2. Continuous random variable:

A **continuous random variable** is one which takes an infinite number of possible values. Continuous random variables are usually measurements. E

A continuous random variable takes all values in an interval of real numbers.

(or)

X is called a **continuous random variable** if $F_X(x)$ is an absolutely continuous function of x . Thus $F_X(x)$ is continuous everywhere on $\mathbb{R}$ and $F_X'(x)$ exists everywhere except at finite or countably infinite points

3. Mixed random variable:

X is called a **mixed random variable** if $F_X(x)$ has jump discontinuity at countable number of points and increases continuously at least in one interval of X . For a such type RV X .

Conditions for a Function to be a Random Variable:

1. Random variable must not be a multi valued function. i.e two or more values of X cannot assign for Single outcome.
2. $P(X=\infty)=P(X=-\infty)=0$
3. The probability of the event $(X \leq x)$ must be equal to the sum of probabilities of all the events Corresponding to $(X \leq x)$.

Bernoulli's trials:

Bernoulli Experiment with n Trials Here are the rules for a Bernoulli experiment.

1. The experiment is repeated a fixed number of times (n times).
2. Each trial has only two possible outcomes, "success" and "failure". The possible outcomes are exactly the same for each trial.
3. The probability of success remains the same for each trial. We use p for the probability of success (on each trial) and $q = 1 - p$ for the probability of failure.
4. The trials are independent (the outcome of previous trials has no influence on the outcome of the next trial).
5. We are interested in the random variable X where X = the number of successes. Note the possible values of X are 0, 1, 2, 3, . . . , n.

An experiment in which a single action, such as flipping a coin, is repeated identically over and over.

The possible results of the action are classified as "success" or "failure". The binomial probability formula is used to find probabilities for Bernoulli trials.

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k q^{n-k}$$

n = number of trials

k = number of successes

n - k = number of failures

p = probability of success in one trial

q = 1 - p = probability of failure in one trial

Problem 1:

If the probability of a bulb being defective is 0.8, then what is the probability of the bulb not being defective.

Solution:

Probability of bulb being defective, $p = 0.8$

Probability of bulb not being defective, $q = 1 - p = 1 - 0.8 = 0.2$

Problem 2:

10 coins are tossed simultaneously where the probability of getting head for each coin is 0.6. Find the probability of getting 4 heads.

Solution:

Probability of getting head, $p = 0.6$

Probability of getting head, $q = 1 - p = 1 - 0.6 = 0.4$

Probability of getting 4 heads out of 10,

$$P(X = 4) = C_4^{10} (0.6)^4 (0.4)^6 = 0.111476736$$

UNIT II

Distribution and density functions and Operations on One Random Variable

Distribution and density functions:

- Distribution function and its Properties
- Density function and its Properties
- Important Types of Distribution and density functions
 - Binomial
 - Uniform
 - Exponential
 - Gaussian
- Conditional Distribution
- Conditional Density function and its properties
- Problems

Operation on One Random Variable:

- Expected value of a random variable, function of a random variable
- Moments about the origin
- Central moments - variance and skew
- Characteristic function
- Moment generating function.

Probability Distribution

The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values. It is also sometimes called the probability function or the probability mass function.

More formally, the probability distribution of a discrete random variable X is a function which gives the probability $p(x_i)$ that the random variable equals x_i , for each value x_i :

$$p(x_i) = P(X=x_i)$$

It satisfies the following conditions:

- a. $0 \leq p(x_i) \leq 1$
- b. $\sum p(x_i) = 1$

Cumulative Distribution Function

All random variables (discrete and continuous) have a cumulative distribution function. 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:

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

for

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

For a discrete random variable, the cumulative distribution function is found by summing up the probabilities as in the example below.

For a continuous random variable, the cumulative distribution function is the integral of its probability density function.

Example

Discrete case : Suppose a random variable X has the following probability distribution $p(x_i)$:

x_i	0	1	2	3	4	5
$p(x_i)$	1/32	5/32	10/32	10/32	5/32	1/32

This is actually a binomial distribution: $Bi(5, 0.5)$ or $B(5, 0.5)$. The cumulative distribution function $F(x)$ is then:

x_i	0	1	2	3	4	5
$F(x_i)$	1/32	6/32	16/32	26/32	31/32	32/32

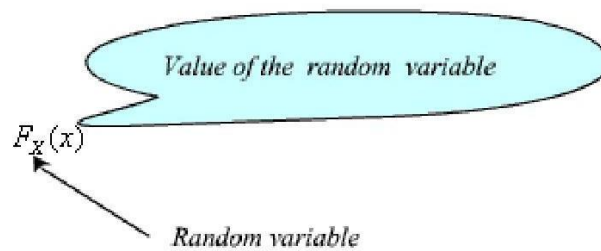
$F(x)$ does not change at intermediate values. For example:

$$F(1.3) = F(1) = 6/32 \text{ and } F(2.86) = F(2) = 16/32$$

Probability Distribution Function

The probability $P(\{X \leq x\}) = P(\{s \mid X(s) \leq x, s \in S\})$ is called the *probability distribution function* (also called the *cumulative distribution function*, abbreviated as *CDF*) of X and denoted by $F_X(x)$. Thus

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



Properties of the Distribution Function

- $0 \leq F_X(x) \leq 1$

- $F_X(x)$ is a non-decreasing function of X . Thus, if $x_1 < x_2$, then $F_X(x_1) < F_X(x_2)$

$$x_1 < x_2$$

$$\Rightarrow \{X(s) \leq x_1\} \subseteq \{X(s) \leq x_2\}$$

$$\Rightarrow P\{X(s) \leq x_1\} \leq P\{X(s) \leq x_2\}$$

$$\therefore F_X(x_1) < F_X(x_2)$$

□ $F_X(x)$ is right continuous.

$$F_X(x^+) = \lim_{\substack{h \rightarrow 0 \\ h > 0}} F_X(x+h) = F_X(x)$$

$$\begin{aligned} \text{Because, } \lim_{\substack{h \rightarrow 0 \\ h > 0}} F_X(x+h) &= \lim_{\substack{h \rightarrow 0 \\ h > 0}} P\{X(s) \leq x+h\} \\ &= P\{X(s) \leq x\} \\ &= F_X(x) \end{aligned}$$

- $F_X(-\infty) = 0$

Because, $F_X(-\infty) = P(\mathcal{S} | X(\mathcal{S}) \leq -\infty) = P(\emptyset) = 0$.

4. $F_X(\infty) = 1$
5. $P(\{x_1 < X \leq x_2\}) = F_X(x_2) - F_X(x_1) | \leq \infty = P(\mathcal{S}) = 1$

We have,

$$\begin{aligned}
 F_X(x^-) &= \lim_{\substack{h \rightarrow 0 \\ h > 0}} F_X(x-h) \\
 &= \lim_{h \rightarrow 0} P\{X(\mathcal{S}) \leq x-h\} \\
 6. \quad P(\{X > x\}) &= P(\{x < X < \infty\}) = 1 - F_X(x) \\
 &= P\{X(\mathcal{S}) \leq x\} - P(X(\mathcal{S}) = x) \\
 &= F_X(x) - P(X = x)
 \end{aligned}$$

Example: Consider the random variable X in the above example. We have

Value of the random Variable $X = x$	$P(\{X = x\})$
0	1/4
1	1/4
2	1/4
3	1/4

For $x < 0$,

$$F_X(x) = P(\{X \leq x\}) = 0$$

For $0 \leq x < 1$,

$$F_X(x) = P(\{X \leq x\}) = P(\{X = 0\}) = \frac{1}{4}$$

For $1 \leq x < 2$,

$$\begin{aligned}
 F_X(x) &= P(\{X \leq x\}) \\
 &= P(\{X = 0\} \cup \{X = 1\}) \\
 &= P(\{X = 0\}) + P(\{X = 1\}) \\
 &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}
 \end{aligned}$$

For $2 \leq x < 3$,

$$\begin{aligned}
 F_X(x) &= P(\{X \leq x\}) \\
 &= P(\{X = 0\} \cup \{X = 1\} \cup \{X = 2\}) \\
 &= P(\{X = 0\}) + P(\{X = 1\}) + P(\{X = 2\}) \\
 &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}
 \end{aligned}$$

For $x \geq 3$,

$$\begin{aligned}
 F_X(x) &= P(\{X \leq x\}) \\
 &= P(S) \\
 &= 1
 \end{aligned}$$

For $2 \leq x < 3$,

$$\begin{aligned} F_X(x) &= P(\{X \leq x\}) \\ &= P(\{X = 0\} \cup \{X = 1\} \cup \{X = 2\}) \\ &= P(\{X = 0\}) + P(\{X = 1\}) + P(\{X = 2\}) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \end{aligned}$$

Thus we have seen that given $F_X(x)$, $-\infty < x < \infty$, we can determine the probability of any event involving values of the random variable X . Thus $F_X(x) \forall x \in \mathbb{X}$ is a complete description of the random variable X .

Example 5 Consider the random variable X defined by

$$\begin{aligned} F_X(x) &= 0, & x < -2 \\ &= \frac{1}{8}x + \frac{1}{4}, & -2 \leq x < 0 \\ &= 1, & x \geq 0 \end{aligned}$$

Find a) $P(X = 0)$.

b) $P\{X \leq 0\}$.

c) $P\{X > 2\}$.

d) $P\{-1 < X \leq 1\}$.

Solutio

$$\begin{aligned} \text{a) } P(X = 0) &= F_X(0^+) - F_X(0^-) \\ &= 1 - \frac{1}{4} = \frac{3}{4} \end{aligned}$$

$$\begin{aligned} \text{b) } P\{X \leq 0\} &= F_X(0) \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{c) } P\{X > 2\} &= 1 - F_X(2) \\ &= 1 - 1 = 0 \end{aligned}$$

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.

More formally, the probability density function, $f(x)$, of a continuous random variable X is the derivative of the cumulative distribution function $F(x)$:

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

Since $F(x) = P(X \leq x)$ it follows that:

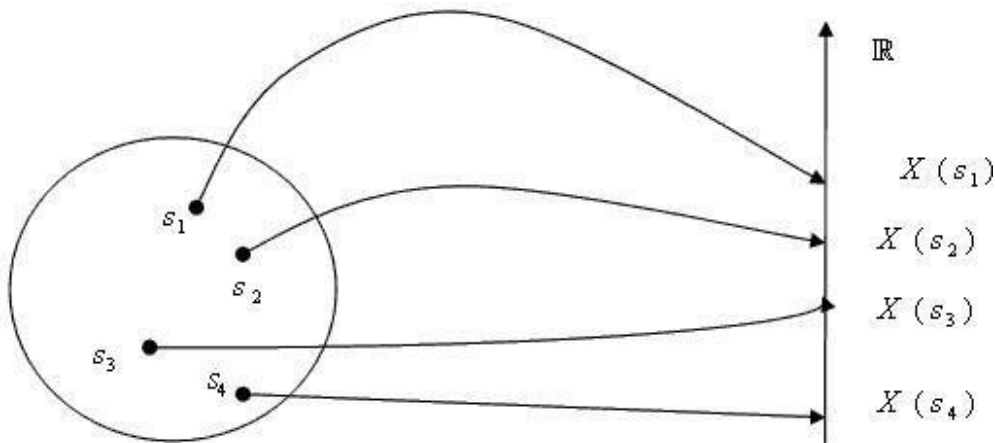
$$\int f(x) dx = F(b) - F(a) = P(a < X < b)$$

If $f(x)$ is a probability density function then it must obey two conditions:

- that the total probability for all possible values of the continuous random variable X is 1:

$$\int f(x) dx = 1$$

- that the probability density function can never be negative: $f(x) > 0$ for all x .

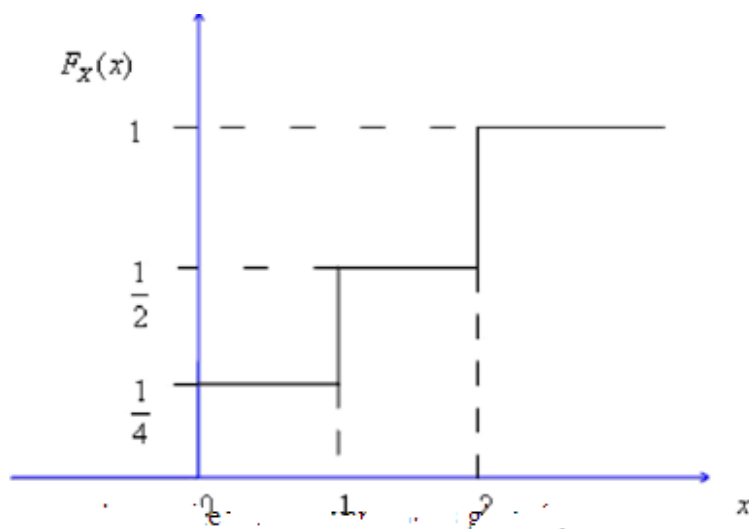


Example 1

Consider the random variable X with the distribution function

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4} & 0 \leq x < 1 \\ \frac{1}{2} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

The plot of the $F_X(x)$ is shown in Figure 7 on next page.



The probability mass function of the random variable is given by

Value of the random variable $X = x$	$p_X(x)$
0	$\frac{1}{4}$
1	$\frac{1}{4}$
2	$\frac{1}{2}$

Properties of the Probability Density Function

1. $f_X(x) \geq 0$.----- This follows from the fact that $F_X(x)$ is a non-decreasing function

2.
$$F_X(x) = \int_{-\infty}^x f_X(u) du$$

3.
$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$

4.
$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f_X(x) dx$$

Other Distribution and density functions of Random variable:

1. Binomial random variable

Suppose X is a discrete random variable taking values from the set $\{0, 1, \dots, n\}$. X is called a binomial random variable with parameters n and $0 \leq p \leq 1$ if

$$p_X(k) = {}^nC_k p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$$

where

$${}^nC_k = \frac{n!}{k!(n-k)!}$$

The trials must meet the following requirements:

- the total number of trials is fixed in advance;
- there are just two outcomes of each trial; success and failure;
- the outcomes of all the trials are statistically independent;
- all the trials have the same probability of success.

As we have seen, the probability of k successes in n independent repetitions of the Bernoulli trial is given by the binomial law. If X is a discrete random variable representing the number of successes in this case, then X is a binomial random variable. For example, the number of heads in n independent tossing of a fair coin is a binomial random variable.

- The notation $X \sim B(n, p)$ is used to represent a binomial RV with the parameters n and

$$\sum_{k=0}^n p_X(k) = \sum_{k=0}^n {}^nC_k p^k (1-p)^{n-k} = [p + (1-p)]^n = 1.$$

- The sum of n independent identically distributed Bernoulli random variables is a binomial random variable.
- The binomial distribution is useful when there are two types of objects - good, bad; correct, erroneous; healthy, diseased etc

Example1: In a binary communication system, the probability of bit error is 0.01. If a block of 8 bits are transmitted, find the probability that

- Exactly 2 bit errors will occur
- At least 2 bit errors will occur
- More than 2 bit errors will occur
- All the bits will be erroneous

Suppose X is the random variable representing the number of bit errors in a block of 8 bits. Then $X \sim B(8, 0.01)$.

Therefore,

- Probability that exactly 2 bit errors will occur

$$\begin{aligned} &= p_X(2) \\ &= {}^8C_2 \times 0.01^2 \times 0.99^6 \\ &= 0.0026 \end{aligned}$$

- Probability that at least 2 bit errors will occur

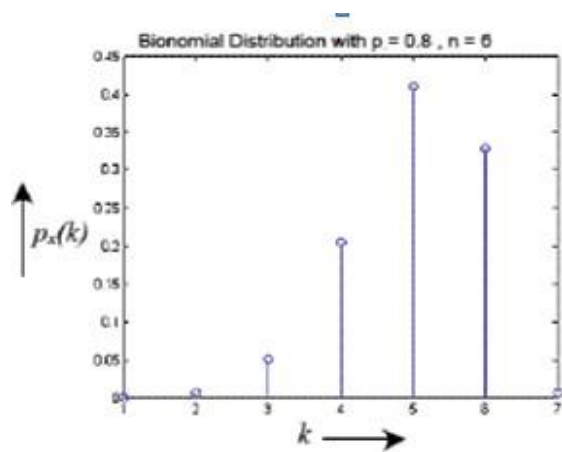
$$\begin{aligned} &= p_X(0) + p_X(1) + p_X(2) \\ &= 0.99^8 + {}^8C_1 \times 0.01^1 \times 0.99^7 + {}^8C_2 \times 0.01^2 \times 0.99^6 \\ &= 0.9999 \end{aligned}$$

- Probability that more than 2 bit errors will occur

$$\begin{aligned} &= 1 - \sum_{k=0}^2 p_X(k) \\ &= 1 - 0.9999 \\ &= 0.0001 \end{aligned}$$

- Probability that all 8 bits will be erroneous

$$\begin{aligned} &= p_X(8) \\ &= 0.01^8 = 10^{-16} \end{aligned}$$



The probability mass function for a binomial random variable with $n = 6$ and $p = 0.8$

Mean and Variance of the Binomial Random Variable

We have

$$\begin{aligned} EX &= \sum_{k=0}^n k p_X(k) \\ &= \sum_{k=0}^n k {}^n C_k p^k (1-p)^{n-k} \\ &= 0 \times q^n + \sum_{k=1}^n k {}^n C_k p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \frac{n-1!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-1-(k-1)} \\ &= np \sum_{k_1=0}^{n-1} \frac{n-1!}{k_1!(n-1-k_1)!} p^{k_1} (1-p)^{n-1-k_1} \quad (\text{Substituting } k_1 = k-1) \\ &= np(p+1-p)^{n-1} \\ &= np \end{aligned}$$

Similarly

$$\begin{aligned}
 EX^2 &= \sum_{k=0}^n k^2 p_X(k) \\
 &= \sum_{k=0}^n k^2 {}^nC_k p^k (1-p)^{n-k} \\
 &= 0^2 \times q^n + \sum_{k=1}^n k^2 {}^nC_k p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
 &= np \sum_{k=1}^n (k-1+1) \frac{n-1!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-1-(k-1)} \\
 &= np \sum_{k=1}^n (k-1) \frac{n-1!}{(k-1)!(n-1-k+1)!} p^{k-1} (1-p)^{n-1-(k-1)} + np \sum_{k=1}^n \frac{n-1!}{(k-1)!(n-1-k+1)!} p^{k-1} (1-p)^{n-1-(k-1)} \\
 &= np \times (n-1)p + np \\
 &= n(n-1)p^2 + np
 \end{aligned}$$

Where

$$\sum_{k=1}^n (k-1) \frac{(n-1)!}{(k-1)!(n-1-k+1)!} p^{k-1} (1-p)^{n-1-(k-1)}$$

is the mean of $B(n-1, p)$

$$\begin{aligned}
 \therefore \sigma_X^2 &= \text{variance of } X \\
 &= n(n-1)p^2 + np - n^2p^2 \\
 &= np(1-p)
 \end{aligned}$$

2. Uniform Random Variable

A continuous random variable X is called uniformly distributed over the interval $[a, b]$,

$-\infty < a < b < \infty$, if its probability density function is given by

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

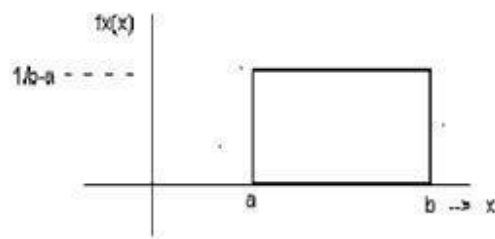


Figure 1

We use the notation $X \sim U(a, b)$ to denote a random variable X uniformly distributed over the interval $[a, b]$

Distribution function $F_X(x)$

For $x < a$

$$F_X(x) = 0$$

For $a \leq x \leq b$

$$\begin{aligned} & \int_{-\infty}^x f_X(u) du \\ &= \int_a^x \frac{1}{b-a} du \\ &= \frac{x-a}{b-a} \end{aligned}$$

For $x > b$,

$$F_X(x) = 1$$

Figure 2 illustrates the CDF of a uniform random variable.

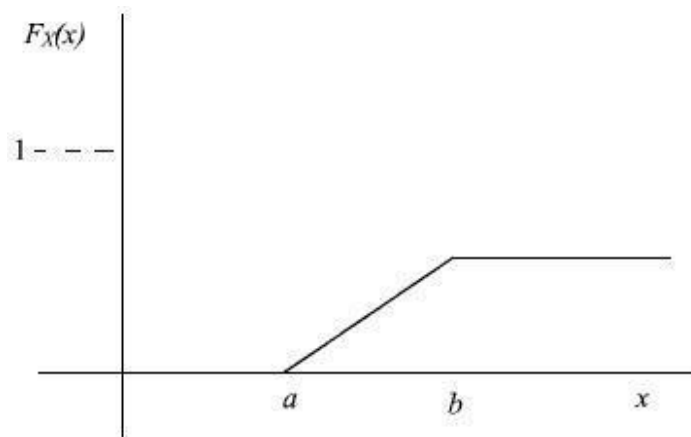


Figure 2: CDF of a uniform random variable

Mean and Variance of a Uniform Random Variable:

$$\begin{aligned}
 \mu_X = EX &= \int_{-\infty}^{\infty} xf_X(x)dx = \int_a^b \frac{x}{b-a} dx \\
 &= \frac{a+b}{2} \\
 EX^2 &= \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_a^b \frac{x^2}{b-a} dx \\
 &= \frac{b^2 + ab + a^2}{3} \\
 \therefore \sigma_X^2 = EX^2 - \mu_X^2 &= \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} \\
 &= \frac{(b-a)^2}{12}
 \end{aligned}$$

The characteristic function of the random variable $X \sim U(a, b)$ is given by

$$\begin{aligned}
 \phi_X(\omega) &= Ee^{j\omega X} = \int_a^b \frac{e^{j\omega x}}{b-a} dx \\
 &= \frac{e^{j\omega b} - e^{j\omega a}}{j\omega(b-a)}
 \end{aligned}$$

3. Normal or Gaussian Random Variable

The normal distribution is the most important distribution used to model natural and man made phenomena. Particularly, when the random variable is the result of the addition of large number of independent random variables, it can be modeled as a normal random variable.

continuous random variable X is called a *normal* or a *Gaussian random variable* with parameters μ_X and σ_X^2 if its probability density function is given by,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}, \quad -\infty < x < \infty$$

Where μ_X and $\sigma_X > 0$ are real numbers.

We write that X is $N(\mu_X, \sigma_X^2)$ distributed.

If $\mu_X = 0$ and $\sigma_X^2 = 1$, and the random variable X is called the *standard normal variable*.

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

Figure 3 illustrates two normal variables with the same mean but different variances.

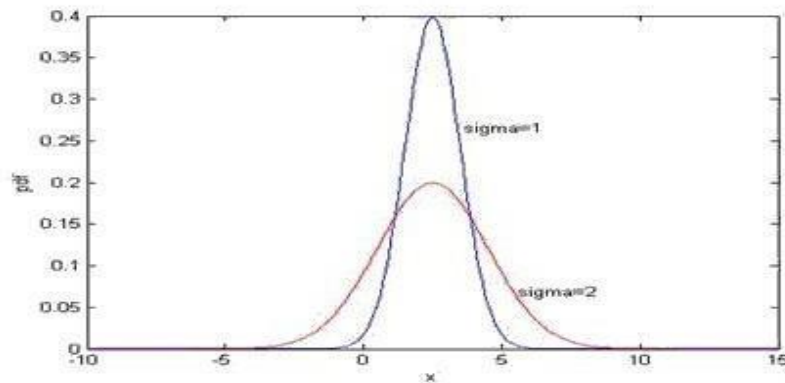


Figure 3

- $f_X(x)$ Is a bell-shaped function, symmetrical about $x = \mu_X$.
- σ_X^2 Determines the spread of the random variable X . If σ_X^2 is small X is more concentrated around the mean μ_X .
- Distribution function of a Gaussian variable is

$$F_X(x) = P(X \leq x) \\ = \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu_X}{\sigma_X}\right)^2} dt$$

Substituting $u = \frac{t - \mu_X}{\sigma_X}$, we get

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu_X}{\sigma_X}} e^{-\frac{1}{2}u^2} du \\ = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right)$$

where $\Phi(x)$ is the distribution function of the *standard normal variable*.

Thus $F_X(x)$ can be computed from tabulated values of $\Phi(x)$. The table $\Phi(x)$ was very useful in the pre-computer days.

In communication engineering, it is customary to work with the Q function defined by,

$$Q(x) = 1 - \Phi(x) \\ = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{2}} du$$

Note that $Q(0) = \frac{1}{2}$, $Q(-x) = Q(x)$ and

$$Q(x) = 1 - \phi(-x)$$

These results follow from the symmetry of the Gaussian pdf. The function $Q(x)$ is tabulated and the tabulated results are used to compute probability involving the Gaussian random variable.

Using the Error Function to compute Probabilities for Gaussian Random Variables

The function $Q(x)$ is closely related to the error function $erf(x)$ and the complementary error function $erfc(x)$.

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$$

Note that,

And the complementary error function $erfc(x)$ is given

$$\begin{aligned} erfc(x) &= \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2} du \\ &= 1 - erf(x) \\ \therefore Q(x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du \\ &= \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right) \\ &= \frac{1}{2} \left(1 - erf\left(\frac{x}{\sqrt{2}}\right)\right) \end{aligned}$$

Mean and Variance of a Gaussian Random Variable

If X is $N(\mu_X, \sigma_X^2)$ distributed, then

$$EX = \mu_X$$

$$\text{var}(X) = \sigma_X^2$$

Proof:

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} xf_X(x)dx = \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} xe^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} dx \\ &= \frac{\sigma_X}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (u\sigma_X + \mu_X) e^{-\frac{1}{2}u^2} \sigma_X du \\ &= \frac{1}{\sigma_X\sqrt{2\pi}} \int_{-\infty}^{\infty} udu + \frac{\mu_X}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du \\ &= 0 + \mu_X \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \\ &= \frac{\mu_X}{\sqrt{2\pi}} 2 \int_0^\infty e^{-\frac{u^2}{2}} du = \mu_X \end{aligned}$$

Substituting $\frac{x - \mu_X}{\sigma_X} = u$

so that $x = u\sigma_X + \mu_X$

-

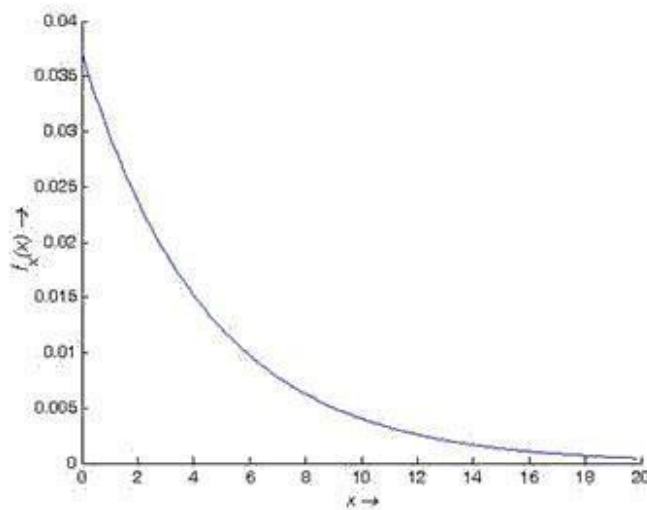
$$\begin{aligned}
 \text{Var}(X) &= E(X - \mu_x)^2 \\
 &= \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{\infty} (x - \mu_x)^2 e^{-\frac{1}{2}\left(\frac{x - \mu_x}{\sigma_x}\right)^2} dx \\
 &= \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{\infty} \sigma_x^2 u^2 e^{-\frac{1}{2}u^2} \sigma_x du \quad \left(\text{substituting } u = \frac{x - \mu_x}{\sigma_x}\right) \\
 &= 2 \times \frac{\sigma_x^2}{\sqrt{2\pi}} \int_0^{\infty} u^2 e^{-\frac{1}{2}u^2} du \\
 &= 2 \times \frac{\sigma_x^2}{\sqrt{2\pi}} \sqrt{2} \int_0^{\infty} t^{\frac{1}{2}} e^{-t} dt \quad \left(\text{substituting } t = \frac{u^2}{2}\right) \\
 &= 2 \times \frac{\sigma_x^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) \\
 &= 2 \times \frac{\sigma_x^2}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \\
 &= \frac{\sigma_x^2}{\sqrt{\pi}} \times \sqrt{\pi} \\
 &= \sigma_x^2
 \end{aligned}$$

4. Exponential Random Variable

A continuous random variable X is called exponentially distributed with the parameter

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

$\lambda > 0$ if the probability density function is of the PDF of Exponential Random Variable is



Example 1

Suppose the waiting time of packets in X in a computer network is an exponential RV with

$$f_X(x) = 0.5e^{-0.5x} \quad x \geq 0$$

Then,

$$\begin{aligned} P(\{0.1 < X \leq 0.5\}) &= \int_{0.1}^{0.5} 0.5e^{-0.5x} dx \\ &= e^{-0.5 \times 0.5} - e^{-0.5 \times 0.1} \\ &= 0.0241 \end{aligned}$$

Conditional Distribution and Density functions:

We discussed conditional probability in an earlier lecture. For two events A and B with $P(B) \neq 0$, the conditional probability $P(A/B)$ was defined as

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

Clearly, the conditional probability can be defined on events involving a Random Variable X

Conditional distribution function:

Consider the event $\{X \leq x\}$ and any event B involving the random variable X . The conditional distribution function of X given B is defined as

$$\begin{aligned} F_X(x/B) &= P[\{X \leq x\} / B] \\ &= \frac{P[\{X \leq x\} \cap B]}{P(B)} \quad P(B) \neq 0 \end{aligned}$$

We can verify that $F_X(x/B)$ satisfies all the properties of the distribution function. Particularly,

- $F_X(-\infty/B) = 0$ And $F_X(\infty/B) = 1$.
- $0 \leq F_X(x/B) \leq 1$.
- $F_X(x/B)$ Is a non-decreasing function of x .
- $P(\{x_1 < X \leq x_2\} / B) = P(\{X \leq x_2\} / B) - P(\{X \leq x_1\} / B)$
 $= F_X(x_2/B) - F_X(x_1/B)$

Conditional Probability Density Function

In a similar manner, we can define the conditional density function $f_X(x|B)$ of the random variable X given the event B as

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

All the properties of the pdf applies to the conditional pdf and we can easily show that

- $f_X(x|B) \geq 0$
- $\int_{-\infty}^{\infty} f_X(x|B) dx = F_X(\infty|B) = 1$
- $F_X(x|B) = \int_{-\infty}^x f_X(u|B) du$

$$\begin{aligned} P(\{x_1 < X \leq x_2\} | B) &= F_X(x_2|B) - F_X(x_1|B) \\ &= \int_{x_1}^{x_2} f_X(x|B) dx \end{aligned}$$

OPERATIONS ON RANDOM VARIABLE

Expected Value of a Random Variable:

- The *expectation* operation extracts a few parameters of a random variable and provides a summary description of the random variable in terms of these parameters.
- It is far easier to estimate these parameters from data than to estimate the distribution or density function of the random variable.
- Moments are some important parameters obtained through the expectation operation.

Expected value or mean of a random variable

The expected value of a random variable X is defined by

$$EX = \int_{-\infty}^{\infty} xf_X(x)dx$$

provided $\int_{-\infty}^{\infty} xf_X(x)dx$ exists.

EX is also called the mean or statistical average of the random variable X and denoted by μ_X .

Note that, for a discrete RV X defined by the *probability mass function* (pmf) $p_X(x_i), i = 1, 2, \dots, N$, the pdf $f_X(x)$ is given by

$$f_X(x) = \sum_{i=1}^N p_X(x_i) \delta(x - x_i)$$

$$\therefore \mu_X = EX = \int_{-\infty}^{\infty} x \sum_{i=1}^N p_X(x_i) \delta(x - x_i) dx$$

$$= \sum_{i=1}^N p_X(x_i) \int_{-\infty}^{\infty} x \delta(x - x_i) dx$$

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

Thus for a discrete random variable X with $p_X(x_i), i = 1, 2, \dots, N$,

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

Interpretation of the mean

- The mean gives an idea about the average value of the random value. The values of the random variable are spread about this value.
- Observe that

$$\begin{aligned} \mu_X &= \int_{-\infty}^{\infty} xf_X(x)dx \\ &= \frac{\int_{-\infty}^{\infty} xf_X(x)dx}{\int_{-\infty}^{\infty} f_X(x)dx} \quad \because \int_{-\infty}^{\infty} f_X(x)dx = 1 \end{aligned}$$

Therefore, the mean can be also interpreted as the *centre of gravity* of the pdf curve.

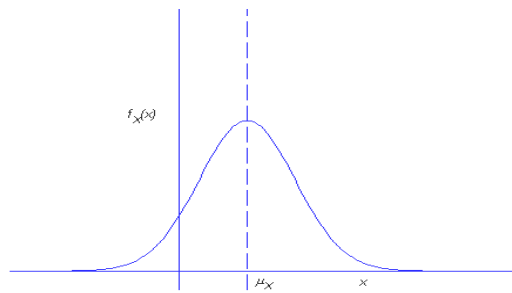


Fig. Mean of a random variable

Example 1 Suppose X is a random variable defined by the pdf

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

Then

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{a+b}{2} \end{aligned}$$

Example 2 Consider the random variable X with pmf as tabulated below

Value of the random variable x	0	1	2	3
$p_X(x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$

$$\begin{aligned} \therefore \mu_X &= \sum_{i=1}^N x_i p_X(x_i) \\ &= 0 \times \frac{1}{8} + 1 \times \frac{1}{8} + 2 \times \frac{1}{4} + 3 \times \frac{1}{2} \\ &= \frac{17}{8} \end{aligned}$$

Remark If $f_X(x)$ is an even function of x , then $\int_{-\infty}^{\infty} x f_X(x) dx = 0$. Thus the mean of a RV with an even symmetric pdf is 0.

Expected value of a function of a random variable

Suppose $Y = g(X)$ is a function of a random variable X as discussed in the last class. Then,

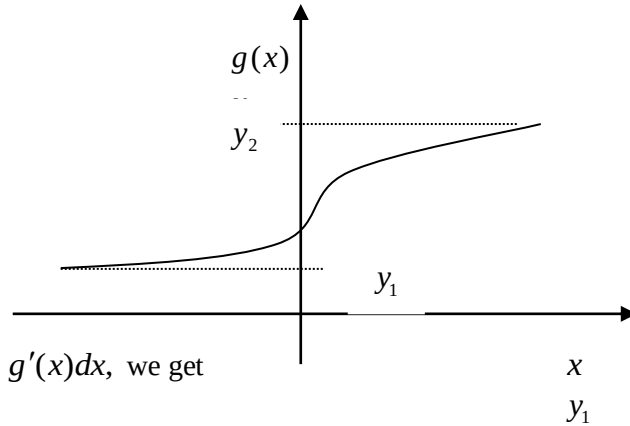
$$EY = Eg(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

We shall illustrate the theorem in the special case $g(X)$ when $y = g(x)$ is one-to-one and monotonically

increasing function of x . In this case,

$$f_Y(y) = \left. \frac{f_X(x)}{g'(x)} \right]_{x=g^{-1}(y)}$$

$$\begin{aligned} EY &= \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= \int_{y_1}^{y_2} y \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} dy \end{aligned}$$



where $y_1 = g(-\infty)$ and $y_2 = g(\infty)$.

Substituting $x = g^{-1}(y)$ so that $y = g(x)$ and $dy = g'(x)dx$, we get

$$EY = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

The following important properties of the expectation operation can be immediately derived:

(a) If c is a constant,

$$Ec = c$$

Clearly

$$Ec = \int_{-\infty}^{\infty} c f_X(x) dx = c \int_{-\infty}^{\infty} f_X(x) dx = c$$

(b) If $g_1(X)$ and $g_2(X)$ are two functions of the random variable X and c_1 and c_2 are constants,

$$E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E g_1(X) + c_2 E g_2(X)$$

$$\begin{aligned} E[c_1 g_1(X) + c_2 g_2(X)] &= \int_{-\infty}^{\infty} c_1 [g_1(x) + c_2 g_2(x)] f_X(x) dx \\ &= \int_{-\infty}^{\infty} c_1 g_1(x) f_X(x) dx + \int_{-\infty}^{\infty} c_2 g_2(x) f_X(x) dx \\ &= c_1 \int_{-\infty}^{\infty} g_1(x) f_X(x) dx + c_2 \int_{-\infty}^{\infty} g_2(x) f_X(x) dx \\ &= c_1 E g_1(X) + c_2 E g_2(X) \end{aligned}$$

The above property means that E is a linear operator.

Mean-square value

$$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

Variance

For a random variable X with the pdf $f_X(x)$ and mean μ_X , the variance of X is denoted by σ_X^2 and defined as

$$\sigma_X^2 = E(X - \mu_X)^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

Thus for a discrete random variable X with $p_X(x_i), i = 1, 2, \dots, N$,

$$\sigma_X^2 = \sum_{i=1}^N (x_i - \mu_X)^2 p_X(x_i)$$

The standard deviation of X is defined as

$$\sigma_X = \sqrt{E(X - \mu_X)^2}$$

Example3: Find the variance of the random variable discussed in Example 1.

$$\begin{aligned}\sigma_X^2 &= E(X - \mu_X)^2 \\ &= \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \left[\int_a^b x^2 dx - 2 \times \frac{a+b}{2} \int_a^b x dx + \left(\frac{a+b}{2}\right)^2 \int_a^b dx \right] \\ &= \frac{(b-a)^2}{12}\end{aligned}$$

Example4: Find the variance of the random variable discussed in Example 2.

As already computed

$$\begin{aligned}\mu_X &= \frac{17}{8} \\ \sigma_X^2 &= E(X - \mu_X)^2 \\ &= \left(0 - \frac{17}{8}\right)^2 \times \frac{1}{8} + \left(1 - \frac{17}{8}\right)^2 \times \frac{1}{8} + \left(2 - \frac{17}{8}\right)^2 \times \frac{1}{4} + \left(3 - \frac{17}{8}\right)^2 \times \frac{1}{2} \\ &= \frac{117}{128}\end{aligned}$$

Remark

- Variance is a central moment and measure of dispersion of the random variable about the mean.
- $E(X - \mu_X)^2$ is the average of the square deviation from the mean. It gives information about the deviation of the values of the RV about the mean. A smaller σ_X^2 implies that the random values are more clustered about the mean, Similarly, a bigger σ_X^2 means that the random values are more scattered.

For example, consider two random variables X_1 and X_2 with pmf as shown below. Note that each of

X_1 and X_2 has zero means. $\sigma_{X_1}^2 = \frac{1}{2}$ and $\sigma_{X_2}^2 = \frac{5}{3}$ implying that X_2 has more spread about the mean

Properties of variance:

$$(1) \quad \sigma_X^2 = EX^2 - \mu_X^2$$

$$\begin{aligned}\sigma_X^2 &= E(X - \mu_X)^2 \\ &= E(X^2 - 2\mu_X X + \mu_X^2) \\ &= EX^2 - 2\mu_X EX + E\mu_X^2 \\ &= EX^2 - 2\mu_X^2 + \mu_X^2 \\ &= EX^2 - \mu_X^2\end{aligned}$$

(2) If $Y = cX + b$, where c and b are constants, then $\sigma_Y^2 = c^2 \sigma_X^2$

$$\begin{aligned}\sigma_Y^2 &= E(cX + b - c\mu_X - b)^2 \\ &= Ec^2(X - \mu_X)^2 \\ &= c^2 \sigma_X^2\end{aligned}$$

(3) If c is a constant,

$$\text{var}(c) = 0.$$

Moments:

The ***n*th moment** of a distribution (or set of data) about a number is the expected value of the *n*th power of the deviations about that number. In statistics, moments are needed about the mean, and about the origin.

1. Moments about origin.
 2. Moments about mean or central moments
- The *n*th moment of a distribution about zero is given by $E(X^n)$
 - The *n*th moment of a distribution about the mean is given by $E[(X-\mu)^n]$

Then each type of measure includes a moment definition.

- The expected value, $E(X)$, is the first moment about zero.
- The variance, $\text{Var}(X)$, is the second moment about the mean, $E[(X-\mu)^2]$
- A common definition of skewness is the third moment about the mean, $E[(X-\mu)^3]$
- A common definition of kurtosis is the fourth moment about the mean, $E[(X-\mu)^4]$

Since moments about zero are typically much easier to compute than moments about the mean, alternative formulas are often provided.

- $\text{Var}(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$
- $\text{Skew}(X) = E[(X-\mu)^3] = E(X^3) - 3E(X)E(X^2) + 2[E(X)]^3$
- $\text{Kurt}(X) = E[(X-\mu)^4] = E(X^4) - 4E(X)E(X^3) + 6[E(X)]^2 E(X^2) - 3[E(X)]^4$

*n*th moment of a random variable

We can define the *n*th moment and the *n*th central-moment of a random variable X by the following relations

$$\text{nth-order moment } EX^n = \int_{-\infty}^{\infty} x^n f_X(x) dx \quad n=1, 2, \dots$$

$$\text{nth-order central moment } E(X - \mu_X)^n = \int_{-\infty}^{\infty} (x - \mu_X)^n f_X(x) dx \quad n=1, 2, \dots$$

Note that

- The mean $\mu_X = EX$ is the first moment and the mean-square value EX^2 is the second moment
- The first central moment is 0 and the variance $\sigma_X^2 = E(X - \mu_X)^2$ is the second central moment
- The third central moment measures lack of symmetry of the pdf of a random variable. $\frac{E(X - \mu_X)^3}{\sigma_X^3}$ is called the *coefficient of skewness* and If the pdf is symmetric this coefficient will be zero.
- The fourth central moment measures flatness of peakedness of the pdf of a random variable. $\frac{E(X - \mu_X)^4}{\sigma_X^4}$ is called *kurtosis*. If the peak of the pdf is sharper, then the random variable has a higher

kurtosi

Moment generating function:

Since each moment is an expected value, and the definition of expected value involves either a sum (in the discrete case) or an integral (in the continuous case), it would seem that the computation of moments could be tedious. However, there is a single expected value function whose derivatives can produce each of the required moments. This function is called a **moment generating function**.

In particular, if X is a random variable, and either $P(x)$ or $f(x)$ is the PDF of the distribution (the first is discrete, the second continuous), then the moment generating function is defined by the following formulas.

$$M_X(t) = E(e^{tX}) = \sum_{\text{all } x} e^{tx} P(x)$$

$$M_X(t) = E(e^{tX}) = \int_x e^{tx} f(x) dx$$

when the n th derivative (with respect to t) of the moment generating function is evaluated at $t=0$, the n th moment of the random variable X about zero will be obtained.

$$E(X^n) = \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0}$$

Properties of moment generating function

(a) The most significant property of moment generating function is that "the moment generating function uniquely determines the distribution."

(b) Let a and b be constants, and let $M_X(t)$ be the mgf of a random variable X . Then the mgf of the random variable $Y = a + bX$ can be given as follows.

$$M_Y(t) = E[e^{tY}] = E[e^{t(a+bX)}] = e^{at} E[e^{(bt)X}] = e^{at} M_X(bt).$$

(c) Let X and Y be independent random variables having the respective mgf's $M_X(t)$ and $M_Y(t)$. Recall that $E[g_1(X)g_2(Y)] = E[g_1(X)] \cdot E[g_2(Y)]$

For functions g_1 and g_2 .

We can obtain the mgf $M_Z(t)$ of the sum $Z = X + Y$ of random variables as follows.

$$M_Z(t) = E[e^{tZ}] = E[e^{t(X+Y)}] = E[e^{tX} e^{tY}] = E[e^{tX}] \cdot E[e^{tY}] = M_X(t) \cdot M_Y(t).$$

(d) When $t = 0$, it clearly follows that $M(0) = 1$. Now by differentiating $M(t)$ r times, we obtain

$$M^{(r)}(t) = \frac{d^r}{dt^r} E[e^{tX}] = E\left[\frac{d^r}{dt^r} e^{tX}\right] = E[X^r e^{tX}].$$

In particular when $t = 0$, $M^{(r)}(0)$ generates the r -th moment of X as follows.
 $M^{(r)}(0) = E[X^r], \quad r = 1, 2, 3, \dots$

Characteristic function

There are random variables for which the moment generating function does not exist on any real interval with positive length. For example, consider the random variable X that has a Cauchy distribution

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad \text{for all } x \in \mathbb{R}.$$

You can show that for any nonzero real number s

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} \frac{1}{\pi} \frac{1}{1+x^2} dx = \infty.$$

Therefore, the moment generating function does not exist for this random variable on any real interval with positive length. If a random variable does not have a well-defined MGF, we can use the characteristic function defined as

$$\phi_X(\omega) = E[e^{j\omega X}]$$

It is worth noting that $e^{j\omega X}$ is a complex-valued random variable. We have not discussed complex-valued random variables. Nevertheless, you can imagine that a complex random variable can be written as $X=Y+jZ$, where Y and Z are ordinary real-valued random variables. Thus, working with a complex random variable is like working with two real-valued random variables. The advantage of the characteristic function is that it is defined for all real-valued random variables. Specifically, if X is a real-valued random variable, we can write

$$|e^{j\omega X}| = 1.$$

The characteristic function has similar properties to the MGF.

1. If X and Y are independent

$$\begin{aligned} \phi_{X+Y}(\omega) &= E[e^{j\omega(X+Y)}] \\ &= E[e^{j\omega X} e^{j\omega Y}] \\ &= E[e^{j\omega X}] E[e^{j\omega Y}] \quad (\text{since } X \text{ and } Y \text{ are independent}) \\ &= \phi_X(\omega) \phi_Y(\omega) \end{aligned}$$

More generally, if $X_1, X_2, \dots, X_n$ are n independent random variables, then

$$\phi_{X_1+X_2+\dots+X_n}(\omega) = \phi_{X_1}(\omega) \phi_{X_2}(\omega) \cdots \phi_{X_n}(\omega).$$

2. Characteristic function and probability density function form Fourier transform pair. i.e

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(\omega) e^{-j\omega x} d\omega$$

Example 1

Consider the random variable X with pdf $f_X(x)$ given by

$$f_X(x) = \frac{1}{b-a} \quad a \leq x \leq b \quad = 0 \text{ otherwise. The characteristics function is given by}$$

$$\phi_X(\omega) = \frac{1}{j\omega(b-a)} (e^{j\omega b} - e^{j\omega a})$$

Solution:

$$\begin{aligned} \phi_X(\omega) &= \int_a^b \frac{1}{b-a} e^{j\omega x} dx \\ &= \frac{1}{b-a} \left[\frac{e^{j\omega x}}{j\omega} \right]_a^b \\ &= \frac{1}{j\omega(b-a)} (e^{j\omega b} - e^{j\omega a}) \end{aligned}$$

Example 2

The characteristic function of the random variable X with

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x} \quad \lambda > 0, x > 0 \text{ is} \\ \phi_X(\omega) &= \int_0^{\infty} e^{j\omega x} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{-(\lambda - j\omega)x} dx \\ &= \frac{\lambda}{\lambda - j\omega} \end{aligned}$$

Characteristic function of a discrete random variable

Suppose X is a random variable taking values from the discrete set $R_X = \{x_1, x_2, \dots\}$ with corresponding probability mass function $p_X(x_i)$ for the value x_i

Then,

$$\begin{aligned}\phi_X(\omega) &= E e^{j\omega X} \\ &= \sum_{X_i \in \mathcal{R}_X} p_X(X_i) e^{j\omega X_i}\end{aligned}$$

$$\begin{aligned}\phi_X(\omega) &= E e^{j\omega X} \\ &= \sum_{X_i \in \mathcal{R}_X} p_X(X_i) e^{j\omega X_i}\end{aligned}$$

If $\mathcal{R}_X$ is the set of integers, we can write

In this case, $\phi_X(\omega)$ can be interpreted as the discrete-time Fourier transform with $e^{j\omega X}$ substituting $e^{-j\omega X}$ in the original discrete-time Fourier transform. The inverse relation is

$$p_X(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j\omega k} \phi_X(\omega) d\omega$$

$p_X(k) = p(1-p)^k$, $k = 0, 1, \dots$, is given by

$$\begin{aligned}\phi_X(\omega) &= \sum_{k=0}^{\infty} e^{j\omega k} p(1-p)^k \\ &= p \sum_{k=0}^{\infty} e^{j\omega k} (1-p)^k \\ &= \frac{p}{1 - (1-p)e^{j\omega}}\end{aligned}$$

Moments and the characteristic function

$$\phi_X(\omega),$$

$$EX^n = \frac{1}{j^n} \frac{d^n}{d\omega^n} \phi_X(\omega) \Big|_{\omega=0}$$

$$e^{j\omega X}$$

$$e^{j\omega X} = 1 + j\omega X + \frac{(j\omega)^2 X^2}{2!} + \dots + \frac{(j\omega)^n X^n}{n!} + \dots$$

$$EX, EX^2, \dots, EX^n$$

$$\phi_X(\omega) = 1 + j\omega EX + \frac{(j\omega)^2 EX^2}{2!} + \dots + \frac{(j\omega)^n EX^n}{n!} + \dots$$

UNIT III

Multiple Random Variables and Operations on Multiple Random Variables

Multiple Random Variables:

- Joint Distribution Function and Properties
- Joint density Function and Properties
- Marginal Distribution and density Functions
- Conditional Distribution and density Functions
- Statistical Independence
- Distribution and density functions of Sum of Two Random Variables

Operations on Multiple Random Variables:

- Expected Value of a Function of Random Variables
- Joint Moments about the Origin
- Joint Central Moments
- Joint Characteristic Functions
- Jointly Gaussian Random Variables: Two Random Variables case

MULTIPLE RANDOM VARIABLES

Multiple Random Variables

In many applications we have to deal with more than two random variables. For example, in the navigation problem, the position of a space craft is represented by three random variables denoting the x, y and z coordinates. The noise affecting the R, G, B channels of color video may be represented by three random variables. In such situations, it is convenient to define the vector-valued random variables where each component of the vector is a random variable.

In this lecture, we extend the concepts of joint random variables to the case of multiple random variables. A generalized analysis will be presented for random variables defined on the same sample space.

Example1: Suppose we are interested in studying the height and weight of the students in a class. We can define the joint RV (X,Y) where X represents height and Y represents the weight.

Example 2 Suppose in a communication system X is the transmitted signal and Y is the corresponding noisy received signal. Then (X,Y) is a joint random variable.

Joint Probability Distribution Function:

Recall the definition of the distribution of a single random variable. The event $\{X \leq x\}$ was used to define the probability distribution function $F_X(x)$. Given $F_X(x)$, we can find the probability of any event involving the random variable. Similarly, for two random variables X and Y , the event $\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$ is considered as the representative event.

The probability $P\{X \leq x, Y \leq y\} \forall (x,y) \in \mathbb{R}^2$ is called the *joint distribution function of the random variables X and Y* and denoted by $F_{X,Y}(x,y)$.

Properties of Joint Probability Distribution Function:

The joint CDF satisfies the following properties:

1. $F_X(x) = F_{X,Y}(x, \infty)$, for any x (marginal CDF of X);

Proof:

$$\{X \leq x\} = \{X \leq x\} \cap \{Y \leq +\infty\}$$

$$\therefore F_X(x) = P(\{X \leq x\}) = P(\{X \leq x, Y \leq \infty\}) = F_{X,Y}(x, +\infty)$$

Similarly $F_Y(y) = F_{X,Y}(\infty, y)$.

2. $F_Y(y) = F_{X,Y}(\infty, y)$, for any y (marginal CDF of Y);
3. $F_{X,Y}(\infty, \infty) = 1$;
4. $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$;
5. $P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$;
6. if X and Y are independent, then $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
7. $F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2)$ if $x_1 \leq x_2$ and $y_1 \leq y_2$

Proof:

If $x_1 < x_2$ and $y_1 < y_2$,

$$\{X \leq x_1, Y \leq y_1\} \subseteq \{X \leq x_2, Y \leq y_2\}$$

$$\therefore P\{X \leq x_1, Y \leq y_1\} \leq P\{X \leq x_2, Y \leq y_2\}$$

$$\therefore F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2)$$

Example1:

Consider two jointly distributed random variables X and Y with the joint CDF

$$F_{X,Y}(x, y) = \begin{cases} (1 - e^{-2x})(1 - e^{-y}) & x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal CDFs

(b) Find the probability $P\{1 < X \leq 2, 1 < Y \leq 2\}$

Solution:

$$(a) \quad F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y) = \begin{cases} 1 - e^{-2x} & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y) = \begin{cases} 1 - e^{-y} & y \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

(b)

$$\begin{aligned} P\{1 < X \leq 2, 1 < Y \leq 2\} &= F_{X,Y}(2, 2) + F_{X,Y}(1, 1) - F_{X,Y}(1, 2) - F_{X,Y}(2, 1) \\ &= (1 - e^{-4})(1 - e^{-2}) + (1 - e^{-2})(1 - e^{-1}) - (1 - e^{-2})(1 - e^{-2}) - (1 - e^{-4})(1 - e^{-1}) \\ &= 0.0272 \end{aligned}$$

Jointly distributed discrete random variables

If X and Y are two discrete random variables defined on the same probability space (S, F, P) such that X takes values from the countable subset R_X and Y takes values from the countable subset R_Y . Then the joint random variable (X, Y) can take values from the countable subset in $R_X \times R_Y$. The joint random variable (X, Y) is completely specified by their *joint probability mass function*

$$p_{X,Y}(x, y) = P\{s \mid X(s) = x, Y(s) = y\}, \quad \forall (x, y) \in R_X \times R_Y.$$

Given $p_{X,Y}(x, y)$, we can determine other probabilities involving the random variables X and Y .

Remark

- $p_{X,Y}(x, y) = 0$ for $(x, y) \notin R_X \times R_Y$

- $\sum_{(x,y) \in R_X \times R_Y} p_{X,Y}(x, y) = 1$

This is because

$$\sum_{(x,y) \in R_X \times R_Y} p_{X,Y}(x, y) = P\left(\bigcup_{(x,y) \in R_X \times R_Y} \{x, y\}\right)$$

$$= P(R_X \times R_Y)$$

$$= P\{s \mid (X(s), Y(s)) \in (R_X \times R_Y)\}$$

$$= P(S) = 1$$

- **Marginal Probability Mass Functions:** The probability mass functions $p_X(x)$ and $p_Y(y)$ are obtained from the joint probability mass function as follows

$$p_X(x) = P\{X = x\} \cup R_Y$$

$$= \sum_{y \in R_Y} p_{X,Y}(x, y)$$

and similarly

$$p_Y(y) = \sum_{x \in R_X} p_{X,Y}(x, y)$$

These probability mass functions $p_X(x)$ and $p_Y(y)$ obtained from the joint probability mass functions are called *marginal probability mass functions*.

Example Consider the random variables X and Y with the joint probability mass function as tabulated in Table . The marginal probabilities are as shown in the last column and the last row

$X \backslash Y$	0	1	2	$p_Y(y)$
0	0.25	0.1	0.15	0.5
1	0.14	0.35	0.01	0.5
$p_X(x)$	0.39	0.45		

Joint Probability Density Function

If X and Y are two continuous random variables and their joint distribution function is continuous in both x and y , then we can define *joint probability density function* $f_{X,Y}(x, y)$ by

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y), \text{ provided it exists.}$$

$$\text{Clearly } F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du$$

Properties of Joint Probability Density Function:

- $f_{X,Y}(x, y)$ is always a non-negative quantity. That is,

$$f_{X,Y}(x, y) \geq 0 \quad \forall (x, y) \in \mathbb{R}^2$$

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

- Marginal probability density functions can be defined as

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy, \quad \text{for all } x,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx, \quad \text{for all } y.$$

- The probability of any Borel set B can be obtained by

$$P(B) = \iint_{(x,y) \in B} f_{X,Y}(x, y) dx dy$$

Marginal Distribution and density Functions:

The probability distribution functions of random variables X and Y obtained from joint distribution function is called as marginal distribution functions. i.e.

$$F_X(x) = F_{XY}(x, \infty), \text{ for any } x \text{ (marginal CDF of } X\text{);}$$

Proof:

$$\{X \leq x\} = \{X \leq x\} \cap \{Y \leq +\infty\}$$

$$\therefore F_X(x) = P(\{X \leq x\}) = P(\{X \leq x, Y \leq \infty\}) = F_{XY}(x, +\infty)$$

$$\text{Similarly } F_Y(y) = F_{XY}(\infty, y).$$

The marginal density functions $f_X(x)$ and $f_Y(y)$ of two joint RVs X and Y are given by the derivatives of the corresponding marginal distribution functions. Thus

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \frac{d}{dx} F_X(x, \infty) \\ &= \frac{d}{dx} \int_{-\infty}^x \left(\int_{-\infty}^{\infty} f_{X,Y}(u, y) dy \right) du \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \end{aligned}$$

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

The marginal CDF and pdf are same as the CDF and pdf of the concerned single random variable. The *marginal* term simply refers that it is derived from the corresponding joint distribution or density function of two or more jointly random variables.

Example2: The joint density function $f_{X,Y}(x, y)$ in the previous example is

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) \\ &= \frac{\partial^2}{\partial x \partial y} [(1 - e^{-2x})(1 - e^{-y})] \quad x \geq 0, y \geq 0 \\ &= 2e^{-2x} e^{-y} \quad x \geq 0, y \geq 0 \end{aligned}$$

Example3: The joint pdf of two random variables X and Y are given by

$$\begin{aligned} f_{X,Y}(x, y) &= cxy \quad 0 \leq x \leq 2, 0 \leq y \leq 2 \\ &= 0 \text{ otherwise} \end{aligned}$$

- (i) Find c .
- (ii) Find $F_{X,Y}(x, y)$
- (iii) Find $f_X(x)$ and $f_Y(y)$.
- (iv) What is the probability $P(0 < X \leq 1, 0 < Y \leq 1)$?

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

$$\therefore c = \frac{1}{4}$$

$$\begin{aligned} F_{X,Y}(x,y) &= \frac{1}{4} \int_0^y \int_0^x uv du dv \\ &= \frac{x^2 y^2}{16} \end{aligned}$$

$$\begin{aligned} f_X(x) &= \int_0^2 \frac{xy}{4} dy \quad 0 \leq y \leq 2 \\ &= \frac{x}{2} \quad 0 \leq y \leq 2 \end{aligned}$$

Similarly

$$f_Y(y) = \frac{y}{2} \quad 0 \leq y \leq 2$$

$$\begin{aligned} P(0 < X \leq 1, 0 < Y \leq 1) \\ &= F_{X,Y}(1,1) + F_{X,Y}(0,0) - F_{X,Y}(0,1) - F_{X,Y}(1,0) \\ &= \frac{1}{16} + 0 - 0 - 0 \\ &= \frac{1}{16} \end{aligned}$$

Conditional Distribution and Density functions

We discussed conditional probability in an earlier lecture. For two events A and B with $P(B) \neq 0$, the conditional probability $P(A/B)$ was defined as

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

Clearly, the conditional probability can be defined on events involving a random variable X .

Conditional distribution function

Consider the event $\{X \leq x\}$ and any event B involving the random variable X . The conditional distribution function of X given B is defined as

$$\begin{aligned} F_X(x/B) &= P[\{X \leq x\} / B] \\ &= \frac{P[\{X \leq x\} \cap B]}{P(B)} \quad P(B) \neq 0 \end{aligned}$$

Properties of Conditional distribution function

We can verify that $F_X(x/B)$ satisfies all the properties of the distribution function. Particularly.

- $F_X(-\infty/B) = 0$ and $F_X(\infty/B) = 1$.
- $0 \leq F_X(x/B) \leq 1$

- $F_X(x/B)$ is a non-decreasing function of x .
- $$P(\{x_1 < X \leq x_2\} / B) = P(\{X \leq x_2\} / B) - P(\{X \leq x_1\} / B)$$
$$= F_X(x_2 / B) - F_X(x_1 / B)$$

Conditional density function

In a similar manner, we can define the conditional density function $f_X(x/B)$ of the random variable X given the event B as

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

Properties of Conditional density function:

All the properties of the pdf applies to the conditional pdf and we can easily show that

- $f_X(x/B) \geq 0$
- $$\int_{-\infty}^{\infty} f_X(x/B) dx = F_X(\infty/B) = 1$$
- $$F_X(x/B) = \int_{-\infty}^x f_X(u/B) du$$
- $$P(\{x_1 < X \leq x_2\} / B) = F_X(x_2/B) - F_X(x_1/B)$$
$$= \int_{x_1}^{x_2} f_X(x/B) dx$$

Let (X, Y) be a discrete bivariate random vector with joint pmf $f(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $P(X = x) = f_X(x) > 0$, the conditional pmf of Y given that $X = x$ is the function of y denoted by $f(y|x)$ and defined by

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

For any y such that $P(Y = y) = f_Y(y) > 0$, the conditional pmf of X given that $Y = y$ is the function of x denoted by $f(x|y)$ and defined by

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

Example 1: Suppose X is a random variable with the distribution function $F_X(x)$. Define $B = \{X \leq b\}$.

Then

$$\begin{aligned} F_X(x/B) &= \frac{P(\{X \leq x\} \cap B)}{P(B)} \\ &= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{P\{X \leq b\}} \\ &= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)} \end{aligned}$$

Case 1: $x < b$

Then

$$\begin{aligned} F_X(x/B) &= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)} \\ &= \frac{P(\{X \leq x\})}{F_X(b)} = \frac{F_X(x)}{F_X(b)} \end{aligned}$$

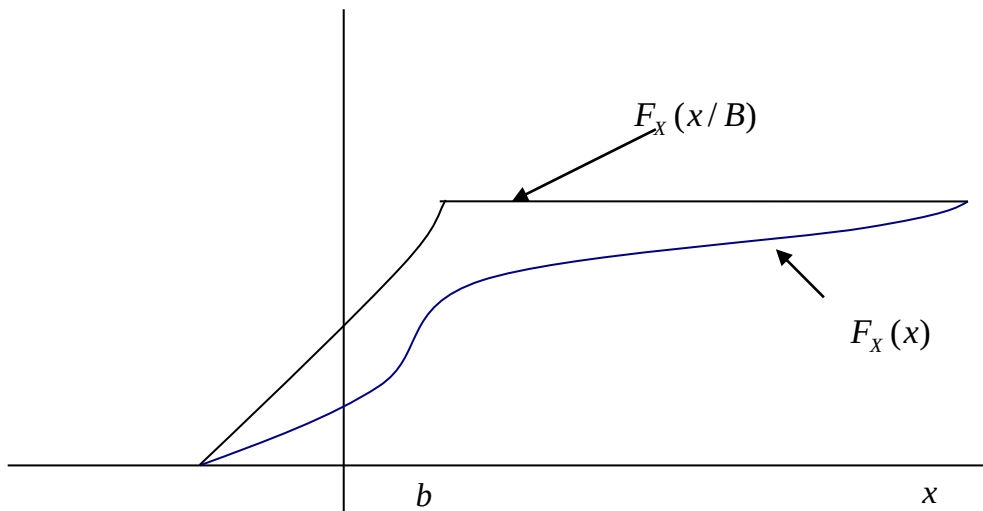
$$\text{And } f_X(x/B) = \frac{d F_X(x)}{dx F_X(b)} = \frac{f_X(x)}{f_X(b)}$$

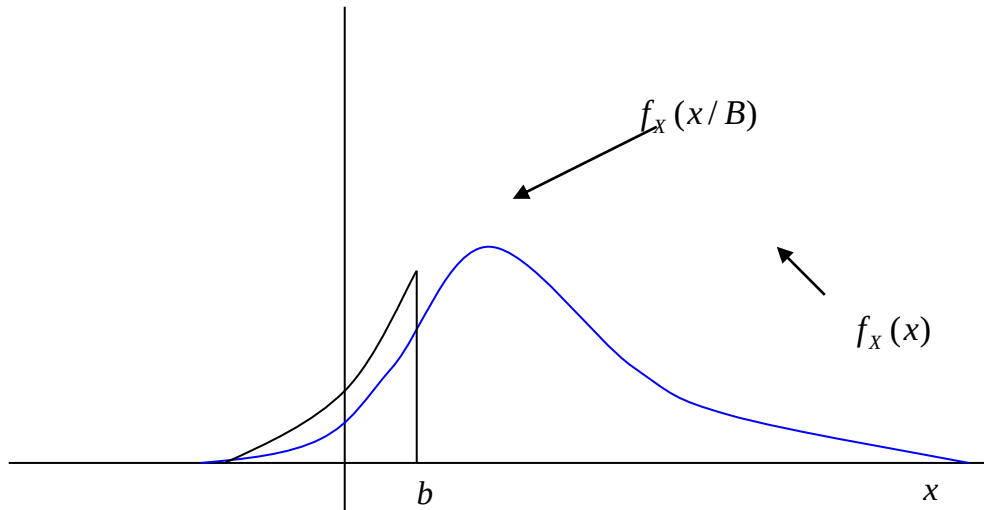
Case 2: $x \geq b$

$$\begin{aligned} F_X(x/B) &= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)} \\ &= \frac{P(\{X \leq b\})}{F_X(b)} = \frac{F_X(b)}{F_X(b)} = 1 \end{aligned}$$

$$\text{and } f_X(x/B) = \frac{d}{dx} F_X(x/B) = 0$$

$F_X(x/B)$ and $f_X(x/B)$ are plotted in the following figures.





Example 2 Suppose X is a random variable with the distribution function $F_X(x)$ and $B = \{X > b\}$.

Then

$$\begin{aligned} F_X(x/B) &= \frac{P(\{X \leq x\} \cap B)}{P(B)} \\ &= \frac{P(\{X \leq x\} \cap \{X > b\})}{P\{X > b\}} \\ &= \frac{P(\{X \leq x\} \cap \{X > b\})}{1 - F_X(b)} \end{aligned}$$

For $x \leq b$, $\{X \leq x\} \cap \{X > b\} = \emptyset$. Therefore,

$$F_X(x/B) = 0 \quad x \leq b$$

For $x > b$, $\{X \leq x\} \cap \{X > b\} = \{b < X \leq x\}$ Therefore,

$$\begin{aligned} F_X(x/B) &= \frac{P(\{b < X \leq x\})}{1 - F_X(b)} \\ &= \frac{F_X(x) - F_X(b)}{1 - F_X(b)} \end{aligned}$$

Thus,

$$F_X(x/B) = \begin{cases} 0 & x \leq b \\ \frac{F_X(x) - F_X(b)}{1 - F_X(b)} & \text{otherwise} \end{cases}$$

The corresponding pdf is given by

$$f_X(x|B) = \begin{cases} 0 & x \leq b \\ \frac{f_X(x)}{1 - F_X(b)} & \text{otherwise} \end{cases}$$

Example 3 The joint density of X and Y is given by

$$f(x, y) = \begin{cases} \frac{15}{2}x(2 - x - y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Compute the condition density of X , given that $Y = y$, where $0 < y < 1$.

Solution For $0 < x < 1, 0 < y < 1$, we have

$$\begin{aligned} f_X(x|y) &= \frac{f(x, y)}{f_Y(y)} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) \, dx} \\ &= \frac{x(2 - x - y)}{\int_0^1 x(2 - x - y) \, dx} = \frac{x(2 - x - y)}{\frac{2}{3} - \frac{y}{2}} = \frac{6x(2 - x - y)}{4 - 3y}. \end{aligned}$$

Example4:

Let the continuous random vector (X, Y) have joint pdf

$$f(x, y) = e^{-y}, \quad 0 < x < y < \infty.$$

The marginal of X is

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

Thus, marginally, X has an exponential distribution. The conditional distribution of Y is

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \begin{cases} \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}, & \text{if } y > x, \\ \frac{0}{e^{-x}} = 0, & \text{if } y \leq x \end{cases}$$

Conditional Probability Distribution Function

Consider two continuous jointly random variables X and Y with the joint probability distribution function $F_{X,Y}(x,y)$. We are interested to find the conditional distribution function of one of the random variables on the condition of a particular value of the other random variable.

We *cannot* define the conditional distribution function of the random variable Y on the condition of the event $\{X = x\}$ by the relation

$$\begin{aligned} F_{Y/X}(y/x) &= P(Y \leq y / X = x) \\ &= \frac{P(Y \leq y, X = x)}{P(X = x)} \end{aligned}$$

as $P(X = x) = 0$ in the above expression. The conditional distribution function is defined in the *limiting sense* as follows:

$$\begin{aligned} F_{Y/X}(y/x) &= \lim_{\Delta x \rightarrow 0} P(Y \leq y / x < X \leq x + \Delta x) \\ &= \lim_{\Delta x \rightarrow 0} \frac{P(Y \leq y, x < X \leq x + \Delta x)}{P(x < X \leq x + \Delta x)} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\int_0^y f_{X,Y}(x,u) \Delta x du}{f_X(x) \Delta x} \\ &= \frac{\int_0^y f_{X,Y}(x,u) du}{f_X(x)} \end{aligned}$$

$$\therefore F_{Y/X}(y/x) = \frac{\int_0^y f_{X,Y}(x,u) du}{f_X(x)}$$

Conditional Probability Density Function

given $f_{Y/X}(y/x) = f_{Y/X}(y/x)$ is called the *conditional probability density function* of Y on X

Let us define the conditional distribution function .

The conditional density is defined in the limiting sense as follows

$$f_{Y|X}(y/X=x) = \lim_{\Delta y \rightarrow 0} (F_{Y|X}(y + \Delta y/X=x) - F_{Y|X}(y/X=x)) / \Delta y$$

$$\therefore f_{Y|X}(y/X=x) = \lim_{\Delta y \rightarrow 0, \Delta x \rightarrow 0} (F_{Y|X}(y + \Delta y/x < X \leq x + \Delta x) - F_{Y|X}(y/x < X \leq x + \Delta x)) / \Delta y$$

$$\text{Because, } (X=x) = \lim_{\Delta x \rightarrow 0} (x < X \leq x + \Delta x)$$

The right hand side of the highlighted equation is

$$\begin{aligned} \lim_{\Delta y \rightarrow 0, \Delta x \rightarrow 0} (F_{Y|X}(y + \Delta y/x < X < x + \Delta x) - F_{Y|X}(y/x < X < x + \Delta x)) / \Delta y \\ &= \lim_{\Delta y \rightarrow 0, \Delta x \rightarrow 0} (P(y < Y \leq y + \Delta y/x < X \leq x + \Delta x)) / \Delta y \\ &= \lim_{\Delta y \rightarrow 0, \Delta x \rightarrow 0} (P(y < Y \leq y + \Delta y, x < X \leq x + \Delta x)) / P(x < X \leq x + \Delta x) \Delta y \\ &= \lim_{\Delta y \rightarrow 0, \Delta x \rightarrow 0} f_{X,Y}(x, y) \Delta x \Delta y / f_X(x) \Delta x \Delta y \\ &= f_{X,Y}(x, y) / f_X(x) \end{aligned}$$

$$\therefore f_{Y|X}(y/x) = f_{X,Y}(x, y) / f_X(x)$$

Similarly we have

$$\therefore f_{X|Y}(x/y) = f_{X,Y}(x, y) / f_Y(y)$$

Two random variables are *statistically independent* if for all $(x, y) \in \mathbb{R}^2$,

$$f_{Y|X}(y/x) = f_Y(y)$$

or equivalently

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

•

Example 2 X and Y are two jointly random variables with the joint pdf given by

$$\begin{aligned} f_{X,Y}(x, y) &= k \text{ for } 0 \leq x \leq 1 \\ &= 0 \text{ otherwise} \end{aligned}$$

find,

(a) k

(b) $f_X(x)$ and $f_Y(y)$

(a) $f_{X|Y}(x/y)$

Solution:

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

We get

$$k \cdot x \cdot \frac{1}{2} \cdot x \cdot 1 = 1$$

$$\Rightarrow k = 2$$

$$\therefore f_{X,Y}(x,y) = 2 \text{ for } 0 \leq x \leq 1 \text{ as } y \leq x \\ = 0 \text{ otherwise}$$

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

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx = 2 \int_y^1 dx = 2(1-y)$$

Independent Random Variables (or) Statistical Independence

Let X and Y be two random variables characterized by the joint distribution function

$$F_{X,Y}(x,y)$$

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

and the corresponding joint density function $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$

Then X and Y are independent if $\forall (x,y) \in \mathbb{R}^2$, $\{X \leq x\}$ and $\{Y \leq y\}$ are independent events.
Thus,

$$\begin{aligned}
 F_{X,Y}(x,y) &= P\{X \leq x, Y \leq y\} \\
 &= P\{X \leq x\}P\{Y \leq y\} \\
 &= F_X(x)F_Y(y)
 \end{aligned}$$

$$\begin{aligned}
 \therefore f_{X,Y}(x,y) &= \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \\
 &= \frac{dF_X(x)}{dx} \frac{dF_Y(y)}{dy} \\
 &= f_X(x)f_Y(y) \\
 \therefore f_{X,Y}(x,y) &= f_X(x)f_Y(y)
 \end{aligned}$$

and equivalently $f_{Y/X}(y) = f_Y(y)$

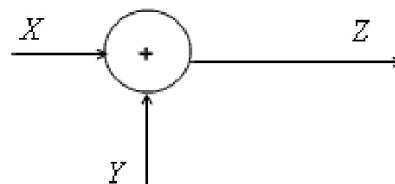
Density function of Sum of Two Random Variables:

We are often interested in finding out the probability density function of a function of two or more RVs. Following are a few examples.

- The received signal by a communication receiver is given by

$$Z = X + Y$$

where Z is received signal which is the superposition of the message signal X and the noise Y .



- The frequently applied operations on communication signals like modulation, demodulation, correlation etc. involve multiplication of two signals in the form $Z = XY$.

We have to know about the probability distribution of Z in any analysis of Z . More formally, given two random variables X and Y with joint probability density function $f_{X,Y}(x,y)$ and a function $Z = g(X,Y)$, we have to find $f_Z(z)$.

In this lecture, we shall address this problem.

We consider the transformation $g : \mathbb{R}^2 \rightarrow \mathbb{R}$.

Consider the event $\{Z \leq z\}$ corresponding to each z . We can find a variable subset $D_z \subseteq \mathbb{R}^2$ such that $D_z = \{(x, y) \mid g(x, y) \leq z\}$.

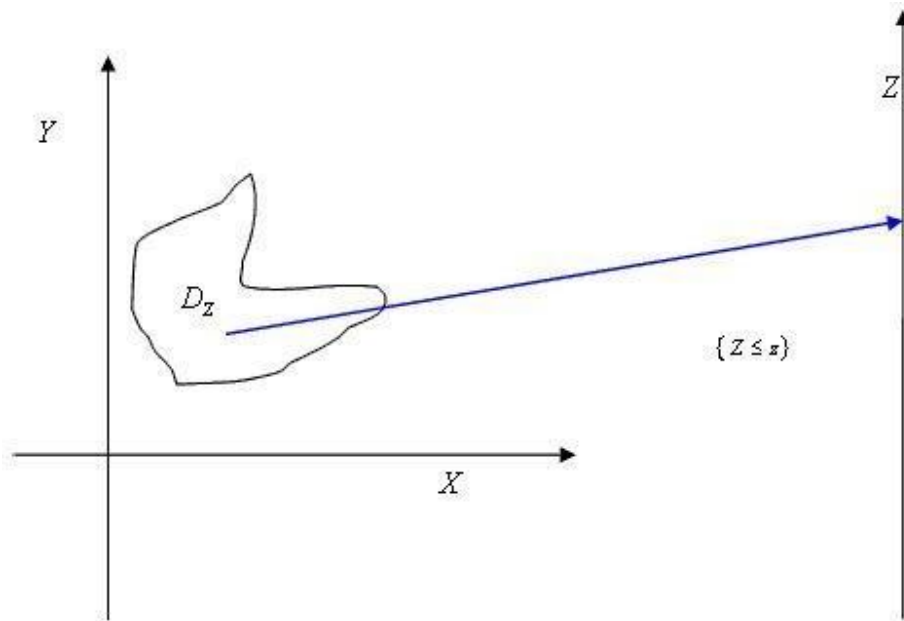


Figure 1

$$\begin{aligned} \therefore F_Z(z) &= P(\{Z \leq z\}) \\ &= P(\{(x, y) \mid (x, y) \in D_z\}) \\ &= \iint_{(x, y) \in D_z} f_{X,Y}(x, y) \, dy \, dx \end{aligned}$$

$$\text{and } f_Z(z) = \frac{dF_Z(z)}{dz}$$

Probability density function of $Z = X + Y$.

Consider Figure 2

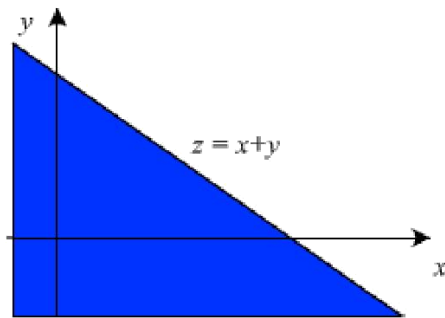


Figure 2

We have

$$Z \leq z$$

$$\Rightarrow X + Y \leq z$$

Therefore, D_Z is the colored region in the Figure

$$\begin{aligned} \therefore F_Z(z) &= \iint_{(x,y) \in D_z} f_{X,Y}(x,y) dx dy \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f_{X,Y}(x,y) dy \right] dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^z f_{X,Y}(x, u-x) du \right] dx \quad \text{substituting } y = u - x \\ &= \int_{-\infty}^z \left[\int_{-\infty}^{\infty} f_{X,Y}(x, u-x) dx \right] du \quad \text{interchanging the order of integration} \\ \therefore f_Z(z) &= \frac{d}{dz} \int_{-\infty}^z \left[\int_{-\infty}^{\infty} f_{X,Y}(x, u-x) dx \right] du \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) dx \end{aligned}$$

$$\therefore f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) dx$$

OPERATIONS ON MULTIPLE RANDOM VARIABLES

Expected Values of Functions of Random Variables

Introduction:

In this Part of Unit we will see the concepts of expectation such as mean, variance, moments, characteristic function, Moment generating function on Multiple Random variables. We are already familiar with same operations on Single Random variable. This can be used as basic for our topics we are going to see on multiple random variables.

Function of joint random variables:

If $g(x,y)$ is a function of two random variables X and Y with joint density function $f_{x,y}(x,y)$ then the expected value of the function $g(x,y)$ is given as

$$\overline{g} = E[g(x,y)] \text{ or}$$

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

Similarly, for N Random variables $X_1, X_2, \dots, X_N$ With joint density function $f_{x_1,x_2,\dots,x_N}(x_1,x_2,\dots,x_n)$, the expected value of the function $g(x_1,x_2,\dots,x_n)$ is given as

$$\overline{g} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1,x_2,\dots,x_N) f_{x_1,x_2,\dots,x_N}(x_1,x_2,\dots,x_N) dx_1 dx_2 \dots dx_N$$

Properties :

The properties of $E(X)$ for continuous random variables are the same as for discrete ones:

1. If X and Y are random variables on a sample space Ω then $E(X + Y) = E(X) + E(Y)$. (linearity I)
2. If a and b are constants then $E(aX + b) = aE(X) + b$.

If $Y = g(X)$ is a function of a discrete random variable X , then $EY = Eg(X) = \sum_{x \in \mathcal{R}_X} g(x)p_X(x)$

Suppose $Z = g(X, Y)$ is a function of continuous random variables X and Y then the expected value of Z is given by

$$\begin{aligned} EZ = Eg(X, Y) &= \int_{-\infty}^{\infty} zf_Z(z)dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dxdy \end{aligned}$$

Thus EZ can be computed without explicitly determining $f_Z(z)$.

We can establish the above result as follows.

Suppose $Z = g(X, Y)$ has n roots (x_i, y_i) , $i = 1, 2, \dots, n$ at $Z = z$. Then

$$\{z < Z \leq z + \Delta z\} = \bigcup_{i=1}^n \{(x_i, y_i) \in \Delta D_i\}$$

Where

ΔD_i is the differential region containing (x_i, y_i) . The mapping is illustrated in Figure 1 for $n = 3$.

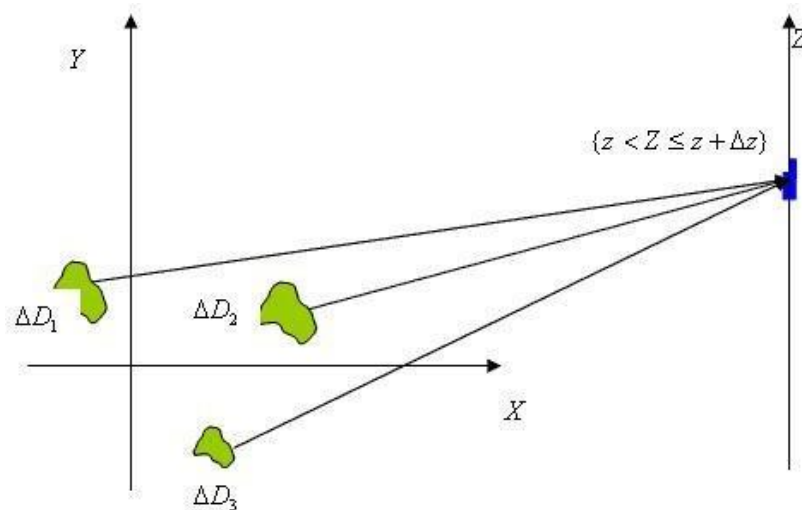


Figure 1

Note that

$$\begin{aligned}
 P(\{z < Z \leq z + \Delta z\}) &= f_Z(z) \Delta z = \sum_{(x_i, y_i) \in D_i} f_{X,Y}(x_i, y_i) \Delta x_i \Delta y_i \\
 \therefore z f_Z(z) \Delta z &= \sum_{(x_i, y_i) \in D_i} z f_{X,Y}(x_i, y_i) \Delta x_i \Delta y_i \\
 &= \sum_{(x_i, y_i) \in D_i} g(x_i, y_i) f_{X,Y}(x_i, y_i) \Delta x_i \Delta y_i
 \end{aligned}$$

As Z is varied over the entire Z axis, the corresponding (non-overlapping) differential regions in $X - Y$ plane cover the entire plane.

$$\therefore \int_{-\infty}^{\infty} z f_Z(z) dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy$$

Thus,

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

If $Z = g(X, Y)$ is a function of discrete random variables X and Y , we can similarly show that

$$EZ = E g(X, Y) = \sum_{x, y \in \mathbb{R}_X \times \mathbb{R}_Y} g(x, y) p_{X,Y}(x, y)$$

Example 1 The joint pdf of two random variables X and Y is given by

$$\begin{aligned}
 f_{X,Y}(x, y) &= \frac{1}{4} xy \quad 0 \leq x \leq 2, 0 \leq y \leq 2 \\
 &= 0 \quad \text{otherwise}
 \end{aligned}$$

Find the joint expectation of $g(X, Y) = X^2Y$

$$\begin{aligned}
 Eg(X, Y) &= EX^2Y \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy \\
 &= \int_0^2 \int_0^2 x^2 y \frac{1}{4} xy dx dy \\
 &= \frac{1}{4} \int_0^2 x^3 dx \int_0^2 y^2 dy \\
 &= \frac{1}{4} \times \frac{2^4}{4} \times \frac{2^3}{3} \\
 &= \frac{8}{3}
 \end{aligned}$$

Example 2 If $Z = aX + bY$, where a and b are constants, then

$$EZ = aEX + bEY$$

Proof:

$$\begin{aligned}
 EZ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by) f_{X,Y}(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ax f_{X,Y}(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} by f_{X,Y}(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} ax \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy dx + \int_{-\infty}^{\infty} by \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy \\
 &= a \int_{-\infty}^{\infty} x f_X(x) dx + b \int_{-\infty}^{\infty} y f_Y(y) dy \\
 &= aEX + bEY
 \end{aligned}$$

Thus, expectation is a linear operator.

Example 3

Consider the discrete random variables X and Y discussed in Example 4 in lecture 18. The joint probability mass function of the random variables are tabulated in Table . Find the joint expectation of $g(X, Y) = XY$.

$\begin{matrix} X \\ Y \end{matrix}$	0	1	2	$p_Y(y)$
0	0.25	0.1	0.15	0.5
1	0.14	0.35	0.01	0.5
$p_X(x)$	0.39	0.45	0.16	

$$\begin{aligned}
 \text{Clearly, } EXY &= \sum_{x,y \in \mathbb{R}_+ \times \mathbb{R}_+} \sum g(x,y) p_{X,Y}(x,y) \\
 &= 1 \times 1 \times 0.35 + 1 \times 2 \times 0.01 \\
 &= 0.37
 \end{aligned}$$

Remark

(1) We have earlier shown that expectation is a linear operator. We can generally write

$$E[a_1 g_1(X, Y) + a_2 g_2(X, Y)] = a_1 E g_1(X, Y) + a_2 E g_2(X, Y)$$

Thus
$$E(XY + 5 \log_e XY) = EXY + 5E \log_e XY$$

(2) If X and Y are independent random variables and $g(X, Y) = g_1(X)g_2(Y)$, then

$$\begin{aligned}
 E g(X, Y) &= E g_1(X) g_2(Y) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(X) g_2(Y) f_{X,Y}(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(X) g_2(Y) f_X(x) f_Y(y) dx dy \\
 &= \int_{-\infty}^{\infty} g_1(X) f_X(x) dx \int_{-\infty}^{\infty} g_2(Y) f_Y(y) dy \\
 &= E g_1(X) E g_2(Y)
 \end{aligned}$$

Joint Moments of Random Variables

Just like the moments of a random variable provide a summary description of the random variable, so also the *joint moments* provide summary description of two random variables. For two continuous random variables X and Y , the *joint moment of order $m+n$* is defined as

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

And the joint central moment of order $m+n$ is defined as

$$E(X - \mu_X)^m (Y - \mu_Y)^n = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^m (y - \mu_Y)^n f_{X,Y}(x, y) dx dy$$

where $\mu_X = EX$ and $\mu_Y = EY$

Remark

(1) If X and Y are discrete random variables, the joint expectation of order m and n is defined as

$$E(X^m Y^n) = \sum_{(x,y) \in \mathbb{R}_{X,Y}} x^m y^n p_{X,Y}(x, y)$$

$$E(X - \mu_X)^m (Y - \mu_Y)^n = \sum_{(x,y) \in \mathbb{R}_{X,Y}} (x - \mu_X)^m (y - \mu_Y)^n p_{X,Y}(x, y)$$

(2) If $m = 1$ and $n = 1$, we have the second-order moment of the random variables X and Y given by

$$E(XY) = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy & \text{if } X \text{ and } Y \text{ are continuous} \\ \sum_{(x,y) \in \mathbb{R}_{X,Y}} xy p_{X,Y}(x, y) & \text{if } X \text{ and } Y \text{ are discrete} \end{cases}$$

(3) If X and Y are independent, $E(XY) = EXEY$

Covariance of two random variables

The *covariance* of two random variables X and Y is defined as

$$\text{Cov}(X, Y) = E(X - \mu_X)(Y - \mu_Y)$$

$\text{Cov}(X, Y)$ is also denoted as $\sigma_{X,Y}$.

Expanding the right-hand side, we get

$$\begin{aligned} \text{Cov}(X, Y) &= E(X - \mu_X)(Y - \mu_Y) \\ &= E(XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y) \\ &= EXY - \mu_Y EX - \mu_X EY + \mu_X \mu_Y \\ &= EXY - \mu_X \mu_Y \end{aligned}$$

The ratio $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$ is called the **correlation coefficient**.

If $\rho_{X,Y} > 0$ then X and Y are called positively correlated.

If $\rho_{X,Y} < 0$ then X and Y are called negatively correlated

If $\rho_{X,Y} = 0$ then X and Y are uncorrelated.

We will also show that $|\rho(X,Y)| \leq 1$. To establish the relation, we prove the following result:

For two random variables X and Y $E^2(XY) \leq EX^2 EY^2$

Proof:

Consider the random variable $Z = aX + Y$

$$\begin{aligned} E(aX + Y)^2 &\geq 0 \\ \Rightarrow a^2 EX^2 + EY^2 + 2aEXY &\geq 0. \end{aligned}$$

Non-negativity of the left-hand side implies that its minimum also must be nonnegative.

For the minimum value,

$$\frac{dEZ^2}{da} = 0 \Rightarrow a = -\frac{EXY}{EX^2}$$

so the corresponding minimum is

$$\begin{aligned} \frac{E^2 XY}{EX^2} + EY^2 - 2 \frac{E^2 XY}{EX^2} \\ = EY^2 - \frac{E^2 XY}{EX^2} \end{aligned}$$

Since the minimum is nonnegative,

$$\begin{aligned} EY^2 - \frac{E^2 XY}{EX^2} &\geq 0 \\ \Rightarrow E^2 XY &\leq EX^2 EY^2 \\ \Rightarrow |EXY| &\leq \sqrt{EX^2} \sqrt{EY^2} \end{aligned}$$

Now

$$\begin{aligned}
 \rho(X, Y) &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \\
 &= \frac{E(X - \mu_X)(Y - \mu_Y)}{\sqrt{E(X - \mu_X)^2} \sqrt{E(Y - \mu_Y)^2}} \\
 \therefore |\rho(X, Y)| &= \frac{|E(X - \mu_X)(Y - \mu_Y)|}{\sqrt{E(X - \mu_X)^2} \sqrt{E(Y - \mu_Y)^2}} \\
 &\leq \frac{\sqrt{E(X - \mu_X)^2} \sqrt{E(Y - \mu_Y)^2}}{\sqrt{E(X - \mu_X)^2} \sqrt{E(Y - \mu_Y)^2}} \\
 &= 1
 \end{aligned}$$

Thus $|\rho(X, Y)| \leq 1$

Uncorrelated random variables

Two random variables X and Y are called *uncorrelated* if

$$\text{Cov}(X, Y) = 0$$

which also means

$$E(XY) = \mu_X \mu_Y$$

Recall that if X and Y are independent random variables, then $f_{X,Y}(x, y) = f_X(x) f_Y(y)$.

$$\begin{aligned}
 EXY &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy && \text{assuming } X \text{ and } Y \text{ are continuous} \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\
 &= \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy \\
 &= EX EY
 \end{aligned}$$

then

Thus two independent random variables are always uncorrelated.

Note that independence implies uncorrelated. But uncorrelated generally does not imply independence (except for jointly Gaussian random variables).

Joint Characteristic Functions of Two Random Variables

The *joint characteristic function* of two random variables X and Y is defined by

$$\phi_{X,Y}(\omega_1, \omega_2) = E e^{j\omega_1 X + j\omega_2 Y}$$

If X and Y are jointly continuous random variables, then

$$\phi_{X,Y}(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) e^{j\omega_1 x + j\omega_2 y} dy dx$$

Note that $\phi_{X,Y}(\omega_1, \omega_2)$ is same as the two-dimensional Fourier transform with the basis function $e^{j\omega_1 x + j\omega_2 y}$ instead of $e^{-(j\omega_1 x + j\omega_2 y)}$.

$f_{X,Y}(x, y)$ is related to the joint characteristic function by the Fourier inversion formula

$$f_{X,Y}(x, y) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{X,Y}(\omega_1, \omega_2) e^{-j\omega_1 x - j\omega_2 y} d\omega_1 d\omega_2$$

If X and Y are discrete random variables, we can define the joint characteristic function in terms of the joint probability mass function as follows:

$$\phi_{X,Y}(\omega_1, \omega_2) = \sum_{(x,y) \in \mathbf{R}_X \times \mathbf{R}_Y} p_{X,Y}(x, y) e^{j\omega_1 x + j\omega_2 y}$$

Properties of the Joint Characteristic Function

The joint characteristic function has properties similar to the properties of the characteristic function of a single random variable. We can easily establish the following properties:

1. $\phi_X(\omega) = \phi_{X,Y}(\omega, 0)$
2. $\phi_Y(\omega) = \phi_{X,Y}(0, \omega)$
3. If X and Y are independent random variables, then

$$\begin{aligned} \phi_{X,Y}(\omega_1, \omega_2) &= E e^{j\omega_1 X + j\omega_2 Y} \\ &= E(e^{j\omega_1 X} e^{j\omega_2 Y}) \\ &= E e^{j\omega_1 X} E e^{j\omega_2 Y} \\ &= \phi_X(\omega_1) \phi_Y(\omega_2) \end{aligned}$$

4. We have,

$$\begin{aligned}
\phi_{X,Y}(\omega_1, \omega_2) &= Ee^{j\omega_1 X + j\omega_2 Y} \\
&= E\left(1 + j\omega_1 X + j\omega_2 Y + \frac{j^2(\omega_1 X + j\omega_2 Y)^2}{2} + \dots\right) \\
&= 1 + j\omega_1 EX + j\omega_2 EY + \frac{j^2 \omega_1^2 EX^2}{2} + \frac{j^2 \omega_2^2 EY^2}{2} + \omega_1 \omega_2 EXY + \dots
\end{aligned}$$

Hence,

$$\begin{aligned}
\phi_{X,Y}(0, 0) &= 1 \\
EX &= \frac{1}{j} \frac{\partial}{\partial \omega_1} \phi_{X,Y}(\omega_1, \omega_2) \Big|_{\omega_1=0} \\
EY &= \frac{1}{j} \frac{\partial}{\partial \omega_2} \phi_{X,Y}(\omega_1, \omega_2) \Big|_{\omega_2=0} \\
EXY &= \frac{1}{j^2} \frac{\partial^2 \phi_{X,Y}(\omega_1, \omega_2)}{\partial \omega_1 \partial \omega_2} \Big|_{\omega_1=0, \omega_2=0}
\end{aligned}$$

In general, the $(m+n)th$ order joint moment is given by

$$EX^m Y^n = \frac{1}{j^{m+n}} \frac{\partial^m \partial^n \phi_{X,Y}(\omega_1, \omega_2)}{\partial \omega_1^m \partial \omega_2^n} \Big|_{\omega_1=0, \omega_2=0}$$

Example 2 The joint characteristic function of the jointly Gaussian random variables X and Y with the joint pdf

$$f_{X,Y}(x, y) = \frac{e^{-\frac{1}{2(1-\rho_{X,Y}^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{X,Y} \left(\frac{x-\mu_X}{\sigma_X} \right) \left(\frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right]}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}}$$

Let us recall the characteristic function of a Gaussian random variable

$$X \sim N(\mu_X, \sigma_X^2)$$

$$\begin{aligned}
\phi_X(\omega) &= E e^{j\omega X} \\
&= \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} e^{j\omega x} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \frac{x^2 - 2(\mu_X - \sigma_X^2 j\omega)x + (\mu_X - \sigma_X^2 j\omega)^2 - (\mu_X - \sigma_X^2 j\omega)^2 + \mu_X^2}{\sigma_X^2}} dx \\
&= e^{\frac{1}{2} \frac{(-\sigma_X^2 \omega^2 + 2\mu_X \sigma_X^2 j\omega)}{\sigma_X^2}} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu_X - \sigma_X^2 j\omega}{\sigma_X}\right)^2} dx}_{\text{Area under a Gaussian}} \\
&= e^{\mu_X j\omega - \sigma_X^2 \omega^2 / 2} \times 1 \\
&= e^{\mu_X j\omega - \sigma_X^2 \omega^2 / 2}
\end{aligned}$$

If X and Y is jointly Gaussian,

$$f_{X,Y}(x,y) = \frac{e^{-\frac{1}{2(1-\rho_{X,Y}^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{X,Y} \left(\frac{x-\mu_X}{\sigma_X} \right) \left(\frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right]}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}}$$

we can similarly show that

$$\begin{aligned}
\phi_{X,Y}(\omega_1, \omega_2) &= E e^{j(X\omega_1 + Y\omega_2)} \\
&= e^{j\mu_X\omega_1 + j\mu_Y\omega_2 - \frac{1}{2}(\sigma_X^2\omega_1^2 + 2\rho_{X,Y}\sigma_X\sigma_Y\omega_1\omega_2 + \sigma_Y^2\omega_2^2)}
\end{aligned}$$

We can use the joint characteristic functions to simplify the probabilistic analysis as illustrated on next page:

Jointly Gaussian Random Variables

Many practically occurring random variables are modeled as jointly Gaussian random variables. For example, noise samples at different instants in the communication system are modeled as *jointly Gaussian random variables*.

Two random variables X and Y are called jointly Gaussian if their joint probability density

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{X,Y} \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right]}, \quad -\infty < x < \infty, -\infty < y < \infty$$

The joint pdf is determined by 5 parameters

- means μ_X and μ_Y
- variances σ_X^2 and σ_Y^2
- correlation coefficient $\rho_{X,Y}$

We denote the jointly Gaussian random variables X and Y with these parameters as

$$(X, Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{X,Y})$$

The joint pdf has a bell shape centered at (μ_X, μ_Y) as shown in the Figure 1 below. The variances σ_X^2 and σ_Y^2 determine the spread of the pdf surface and $\rho_{X,Y}$ determines the orientation of the surface in the $X - Y$ plane.

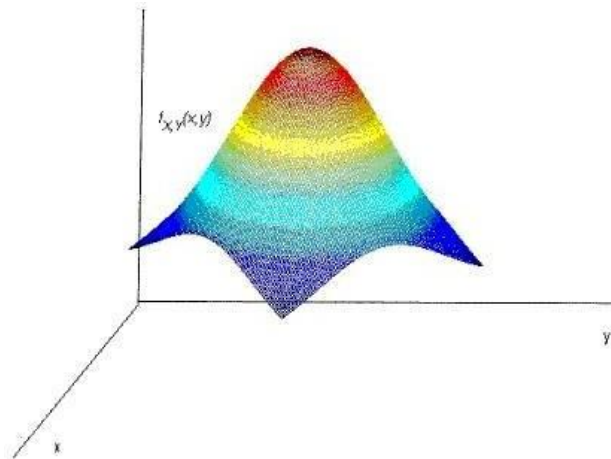


Figure 1 Jointly Gaussian PDF surface

Properties of jointly Gaussian random variables

- (1) If X and Y are jointly Gaussian, then X and Y are both Gaussian.

We have

$$\begin{aligned}
 f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \\
 &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} e^{-\frac{1}{2(1-\rho_{XY}^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{XY} \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right]} dy \\
 &= \frac{e^{-\frac{1}{2} \left(\frac{x-\mu_X}{\sigma_X} \right)^2}}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{XY}^2}} e^{-\frac{1}{2(1-\rho_{XY}^2)} \left[\frac{\rho_{XY}^2 (x-\mu_X)^2}{\sigma_X^2} - 2\rho_{XY} \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right]} dy \\
 &= \frac{e^{-\frac{1}{2} \left(\frac{x-\mu_X}{\sigma_X} \right)^2}}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{XY}^2}} e^{-\frac{1}{2\sigma_Y^2(1-\rho_{XY}^2)} \left[(y-\mu_Y - \frac{\rho_{XY}\sigma_Y}{\sigma_X} (x-\mu_X))^2 \right]} dy \\
 &= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2} \left(\frac{x-\mu_X}{\sigma_X} \right)^2}
 \end{aligned}$$

Similarly

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2} \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2}$$

(2) The converse of the above result is not true. If each of X and Y is Gaussian, X and Y are not necessarily jointly Gaussian. Suppose

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right]} (1 + \sin x \sin y)$$

$f_{X,Y}(x,y)$ in this example is non-Gaussian and qualifies to be a joint pdf. Because,

$f_{X,Y}(x,y) \geq 0$ And

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} (1 + \sin x \sin y) dy dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} dy dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} \sin x \sin y dy dx \\
&= 1 + \underbrace{\frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}} \sin x dx \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}} \sin y dy}_{\text{integration of an odd function}} \\
&= 1 + 0 \\
&= 1
\end{aligned}$$

The marginal density $f_X(x)$ is given by

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} (1 + \sin x \sin y) dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} dy + \underbrace{\int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} \sin x \sin y dy}_{\text{integration of an odd function}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} + 0 \\
&= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}
\end{aligned}$$

Similarly, $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}$

Thus X and Y are both Gaussian, but not jointly Gaussian.

(3) If X and Y are jointly Gaussian, then for any constants a and b , the random variable Z given by $Z = aX + bY$ is Gaussian with mean $\mu_Z = a\mu_X + b\mu_Y$ and variance $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_X\sigma_Y\rho_{X,Y}$

(4) Two jointly Gaussian RVs X and Y are independent if and only if X and Y are uncorrelated ($\rho_{X,Y} = 0$). Observe that if X and Y are uncorrelated, then

$$\begin{aligned}
 f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} \\
 &= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \\
 &= f_X(x)f_Y(y)
 \end{aligned}$$

Example 1 Suppose X and Y are two jointly-Gaussian 0-mean random variables with variances of 1 and 4 respectively and a covariance of 1. Find the joint PDF $f_{X,Y}(x,y)$

$$\mu_X = \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 4 \text{ and } \text{cov}(X,Y) = 1.$$

$$\therefore \rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{1}{1 \times 2} = \frac{1}{2}$$

and

$$\begin{aligned}
 f_{X,Y}(x,y) &= \frac{1}{2\pi \times 1 \times 2 \sqrt{1 - \frac{1}{4}}} e^{-\frac{1}{2 \times \frac{1}{4}} \left[x^2 - 2 \times \frac{1}{2} \times \frac{xy}{1 \times 2} + \frac{y^2}{4} \right]} \\
 &= \frac{1}{2\sqrt{3}\pi} e^{-\frac{1}{3} \left[x^2 - \frac{xy}{1} + \frac{y^2}{4} \right]}
 \end{aligned}$$

We have

Example 2 Linear transformation of two random variables

Suppose $Z = aX + bY$. then

$$\phi_Z(\omega) = Ee^{j\omega Z} = Ee^{j(aX+bY)\omega} = \phi_{X,Y}(a\omega, b\omega)$$

If X and Y are jointly Gaussian, then

$$\begin{aligned}
 \phi_Z(\omega) &= \phi_{X,Y}(a\omega, b\omega) \\
 &= e^{j(\mu_X + \mu_Y)\omega - \frac{1}{2}(a^2\sigma_X^2 + 2\rho_{X,Y}a\omega b\sigma_X\sigma_Y + b^2\sigma_Y^2)\omega^2}
 \end{aligned}$$

Which is the characteristic function of a Gaussian random variable

with mean $\mu_Z = \mu_X + \mu_Y$ and variance $\sigma_Z^2 = \sigma_X^2 + 2\rho_{X,Y}\sigma_X\sigma_Y + \sigma_Y^2$

thus the linear transformation of two Gaussian random variables is a Gaussian random variable

UNIT IV

Stochastic Processes-Temporal Characteristics

- The Stochastic process Concept
- Classification of Processes, Deterministic and Nondeterministic Processes
- Distribution and Density Functions
- Statistical Independence
- Concept of Stationarity: First-Order Stationary Processes, Second-Order and Wide-Sense Stationarity, Nth-Order and Strict-Sense Stationarity
- Time Averages and Ergodicity
- Mean-Ergodic Processes
- Correlation-Ergodic Processes
- Autocorrelation Function and Its Properties
- Cross-Correlation Function and Its Properties
- Covariance Functions and its properties

Linear system Response:

- Mean
- Mean-squared value
- Autocorrelation
- Cross-Correlation Functions

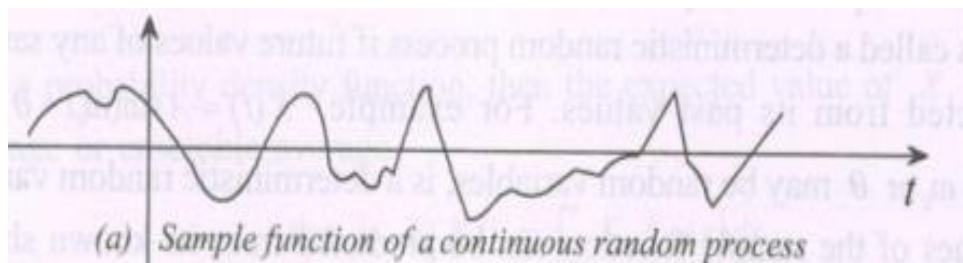
Stochastic process Concept:

The random processes are also called as stochastic processes which deal with randomly varying time wave forms such as any message signals and noise. They are described statistically since the complete knowledge about their origin is not known. So statistical measures are used. Probability distribution and probability density functions give the complete statistical characteristics of random signals. A random process is a function of both sample space and time variables. And can be represented as $\{X x(s,t)\}$.

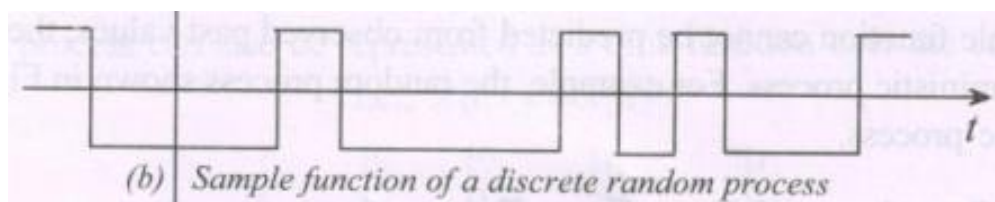
Deterministic and Non-deterministic processes: In general a random process may be deterministic or non deterministic. A process is called as deterministic random process if future values of any sample function can be predicted from its past values. For example, $X(t) = A \sin (\omega_0 t + \Theta)$, where the parameters A , ω_0 and Θ may be random variables, is deterministic random process because the future values of the sample function can be detected from its known shape. If future values of a sample function cannot be detected from observed past values, the process is called non-deterministic process.

Classification of random process: Random processes are mainly classified into four types based on the time and random variable X as follows.

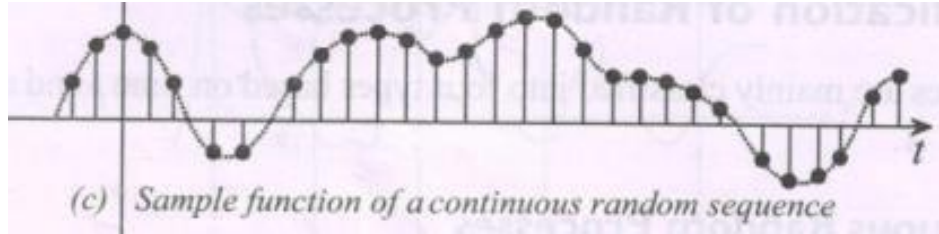
1. Continuous Random Process: A random process is said to be continuous if both the random variable X and time t are continuous. The below figure shows a continuous random process. The fluctuations of noise voltage in any network is a continuous random process.



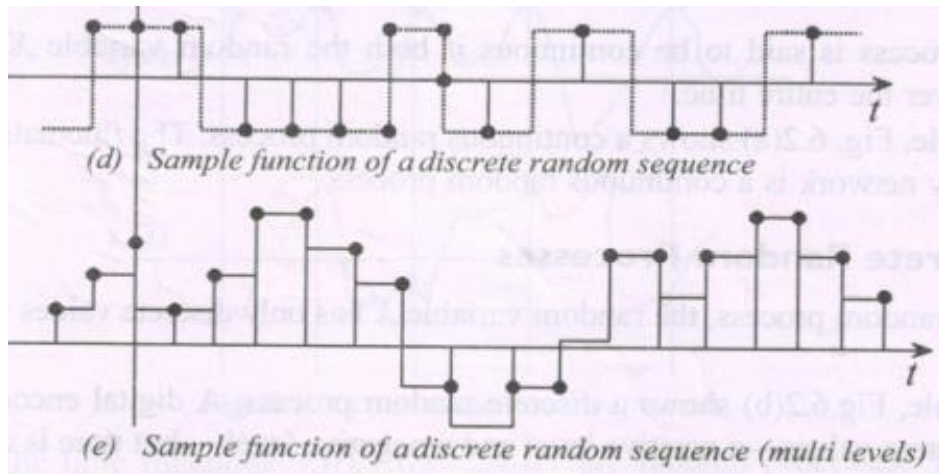
2. Discrete Random Process: In discrete random process, the random variable X has only discrete values while time, t is continuous. The below figure shows a discrete random process. A digital encoded signal has only two discrete values a positive level and a negative level but time is continuous. So it is a discrete random process.



3. Continuous Random Sequence: A random process for which the random variable X is continuous but t has discrete values is called continuous random sequence. A continuous random signal is defined only at discrete (sample) time intervals. It is also called as a discrete time random process and can be represented as a set of random variables $\{X(t_k)\}$ for samples $t_k, k=0, 1, 2, \dots$



4. Discrete Random Sequence: In discrete random sequence both random variable X and time t are discrete. It can be obtained by sampling and quantizing a random signal. This is called the random process and is mostly used in digital signal processing applications. The amplitude of the sequence can be quantized into two levels or multi levels as shown in below figure s (d) and (e).



Joint distribution functions of random process: Consider a random process $X(t)$. For a single random variable at time t_1 , $X_1=X(t_1)$, The cumulative distribution function is defined as $F_X(x_1; t_1) = P \{(X(t_1) \leq x_1)\}$ where x_1 is any real number. The function $F_X(x_1; t_1)$ is known as the first order distribution function of $X(t)$. For two random variables at time instants t_1 and t_2 $X(t_1) = X_1$ and $X(t_2) = X_2$, the joint distribution is called the second order joint distribution function of the random process $X(t)$ and is given by

$F_X(x_1, x_2 ; t_1, t_2) = P \{(X(t_1) \leq x_1, X(t_2) \leq x_2)\}$. In general for N random variables at N time intervals $X(t_i) = X_i, i=1,2,\dots,N$, the N^{th} order joint distribution function of $X(t)$ is defined as

$$F_X(x_1, x_2, \dots, x_N ; t_1, t_2, \dots, t_N) = P \{(X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_N) \leq x_N)\}.$$

Joint density functions of random process: Joint density functions of a random process can be obtained from the derivatives of the distribution functions.

- 1 First order density function:

$$f_X(x_1; t_1) = \frac{dF_X(x_1; t_1)}{dx_1}$$

- 2 second order density function

$$f_X(x_1, x_2; t_1, t_2) = \frac{\partial^2 F_X(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2}$$

- 3 Nth order density function:

$$f_X(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) = \frac{\partial^N F_X(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N)}{\partial x_1 \partial x_2 \dots \partial x_N}$$

Independent random processes: Consider a random process $X(t)$. Let $X(t_i) = x_i$, $i = 1, 2, \dots, N$ be N Random variables defined at time constants $t_1, t_2, \dots, t_N$ with density functions $f_X(x_1; t_1)$, $f_X(x_2; t_2)$, ... $f_X(x_N; t_N)$. If the random process $X(t)$ is statistically independent, then the N^{th} order joint density function is equal to the product of individual joint functions of $X(t)$ i.e.

$f_X(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) = f_X(x_1; t_1) f_X(x_2; t_2) \dots f_X(x_N; t_N)$. Similarly the joint distribution will be the product of the individual distribution functions.

Statistical properties of Random Processes: The following are the statistical properties of random processes.

1. **Mean:** The mean value of a random process $X(t)$ is equal to the expected value of the random process i.e

$$\overline{X(t)} = E[X(t)] = \int_{-\infty}^{\infty} x f_X(x; t) dx$$

2. **Autocorrelation:** Consider random process $X(t)$. Let X_1 and X_2 be two random variables defined at times t_1 and t_2 respectively with joint density function $f_X(x_1, x_2; t_1, t_2)$. The correlation of X_1 and X_2 , $E[X_1 X_2] = E[X(t_1) X(t_2)]$ is called the autocorrelation function of the random process $X(t)$ defined as $R_{XX}(t_1, t_2) = E[X_1 X_2] = E[X(t_1) X(t_2)]$ or

$$R_{XX}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_X(x_1, x_2; t_1, t_2) dx_1 dx_2$$

3. **Cross correlation:** Consider two random processes $X(t)$ and $Y(t)$ defined with random variables X and Y at time instants t_1 and t_2 respectively. The joint density function is $f_{XY}(x, y; t_1, t_2)$. Then the correlation of X and Y , $E[XY] = E[X(t_1) Y(t_2)]$ is called the cross correlation function of the random processes $X(t)$ and $Y(t)$ which is defined as

$$R_{XY}(t_1, t_2) = E[X Y] = E[X(t_1) Y(t_2)] \text{ or}$$

$$R_{XY}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y; t_1, t_2) dx dy$$

Stationary Processes: A random process is said to be stationary if all its statistical properties such as mean, moments, variances etc... do not change with time. The stationarity which depends on the density functions has different levels or orders.

1. **First order stationary process:** A random process is said to be stationary to order one or first order stationary if its first order density function does not change with time or shift in time value. If $X(t)$ is a first order stationary process then

$$f_X(x_1; t_1) = f_X(x_1; t_1 + \Delta t) \text{ for any time } t_1.$$

Where Δt is shift in time value. Therefore the condition for a process to be a first order stationary random process is that its mean value must be constant at any time instant. i.e.

$$E[X(t)] = \bar{X} = \text{constant}$$

2. **Second order stationary process:** A random process is said to be stationary to order two or second order stationary if its second order joint density function does not change with time or shift in time value i.e. $f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; t_1 + \Delta t, t_2 + \Delta t)$ for all t_1, t_2 and Δt . It is a function of time difference (t_2, t_1) and not absolute time t . Note that a second order stationary process is also a first order stationary process. The condition for a process to be a second order stationary is that its autocorrelation should depend only on time differences and not on absolute time. i.e. If

$R_{XX}(t_1, t_2) = E[X(t_1) X(t_2)]$ is autocorrelation function and $\tau = t_2 - t_1$ then

$R_{XX}(t_1, t_1 + \tau) = E[X(t_1) X(t_1 + \tau)] = R_{XX}(\tau)$. $R_{XX}(\tau)$ should be independent of time t .

3. **Wide sense stationary (WSS) process:** If a random process $X(t)$ is a second order stationary process, then it is called a wide sense stationary (WSS) or a weak sense stationary process. However the converse is not true.

The conditions for a wide sense stationary process are

1. $E[X(t)] = \bar{X} = \text{constant}$

2. $E[X(t)X(t+\tau)] = R_{XX}(\tau)$ is independent of absolute time t .

Joint wide sense stationary process: Consider two random processes $X(t)$ and $Y(t)$. If they are jointly WSS, then the cross correlation function of $X(t)$ and $Y(t)$ is a function of time difference $\tau = t_2 - t_1$ only and not absolute time. i.e. $R_{XY}(t_1, t_2) = E[X(t_1) Y(t_2)]$. If $\tau = t_2 - t_1$ then

$$R_{XY}(t, t + \tau) = E[X(t) Y(t + \tau)] = R_{XY}(\tau).$$

Therefore the conditions for a process to be joint wide sense stationary are

1. $E[X(t)] = \bar{X} = \text{constant}.$

2. $E[Y(t)] = \bar{Y} = \text{constant}$

3. $E[X(t) Y(t + \tau)] = R_{XY}(\tau)$ is independent of time t .

Strict sense stationary (SSS) processes: A random process $X(t)$ is said to be strict Sense stationary if its Nth order joint density function does not change with time or shift in time value. i.e.

$$f_X(x_1, x_2, \dots, x_N; t_1, t_2, \dots, t_N) = f_X(x_1, x_2, \dots, x_N; t_1 + \Delta t, t_2 + \Delta t, \dots, t_N + \Delta t) \text{ for all } t_1, t_2, \dots, t_N \text{ and } \Delta t.$$

A process that is stationary to all orders $n=1, 2, \dots, N$ is called strict sense stationary process. Note that SSS process is also a WSS process. But the reverse is not true.

Time Average Function: Consider a random process $X(t)$. Let $x(t)$ be a sample function which exists for all time at a fixed value in the given sample space S . The average value of $x(t)$ taken over all times is called the time average of $x(t)$. It is also called mean value of $x(t)$. It can be expressed as

$$\bar{x} = A[X(t)] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt$$

Time autocorrelation function: Consider a random process $X(t)$. The time average of the product $X(t)$ and $X(t + \tau)$ is called time average autocorrelation function of $x(t)$ and is denoted

$$R_{XX}(\tau) = E[X(t) X(t + \tau)] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) x(t + \tau) dt$$

Time mean square function: If $\tau = 0$, the time average of $x(t)$ is called time mean square value of $x(t)$ defined as

$$A[X^2(t)] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x^2(t) dt .$$

Time cross correlation function: Let $X(t)$ and $Y(t)$ be two random processes with sample functions $x(t)$ and $y(t)$ respectively. The time average of the product of $x(t)$ $y(t + \tau)$ is called time cross correlation function of $x(t)$ and $y(t)$. Denoted as

$$R_{XY}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) y(t + \tau) dt$$

Ergodic Theorem and Ergodic Process: The Ergodic theorem states that for any random process $X(t)$, all time averages of sample functions of $x(t)$ are equal to the corresponding statistical or ensemble averages of $X(t)$. Random processes that satisfy the Ergodic theorem are called Ergodic processes.

Joint Ergodic Process: Let $X(t)$ and $Y(t)$ be two random processes with sample functions $x(t)$ and $y(t)$ respectively. The two random processes are said to be jointly Ergodic if they are individually Ergodic and their time cross correlation functions are equal to their respective statistical cross correlation functions. i.e.

$$1. \bar{x} = \bar{X}, \bar{y} = \bar{Y} \quad 2. R_{xx}(\tau) = R_{XX}(\tau), R_{xy}(\tau) = R_{XY}(\tau)$$

$$\text{and } R_{yy}(\tau) = R_{YY}(\tau).$$

Mean Ergodic Random Process: A random process $X(t)$ is said to be mean Ergodic if time average of any sample function $x(t)$ is equal to its statistical average, $\bar{x}$ which is constant and the probability of all other sample functions is equal to one. i.e.

$$E[X(t)] = \bar{X} = A[X(t)] = \bar{x}$$

with probability one for all $x(t)$.

Autocorrelation Ergodic Process: A stationary random process $X(t)$ is said to be Autocorrelation Ergodic if and only if the time autocorrelation function of any sample function $x(t)$ is equal to the statistical autocorrelation function of $X(t)$. i.e. $A[x(t) x(t+\tau)] = E[X(t) X(t+\tau)]$ or $R_{xx}(\tau) = R_{XX}(\tau)$.

Cross Correlation Ergodic Process: Two stationary random processes $X(t)$ and $Y(t)$ are said to be cross correlation Ergodic if and only if its time cross correlation function of sample functions $x(t)$ and $y(t)$ is equal to the statistical cross correlation function of $X(t)$ and $Y(t)$. i.e. $A[x(t) y(t+\tau)] = E[X(t) Y(t+\tau)]$ or $R_{xy}(\tau) = R_{XY}(\tau)$.

Properties of Autocorrelation function: Consider that a random process $X(t)$ is at least WSS and is a function of time difference $\tau = t_2 - t_1$. Then the following are the properties of the autocorrelation function of $X(t)$.

1. Mean square value of $X(t)$ is $E[X^2(t)] = R_{xx}(0)$. It is equal to the power (average) of the process, $X(t)$.

Proof: We know that for $X(t)$, $R_{xx}(\tau) = E[X(t) X(t+\tau)]$. If $\tau = 0$, then $R_{xx}(0) = E[X(t) X(t)] = E[X^2(t)]$ hence proved.

2. Autocorrelation function is maximum at the origin i.e.

$$|R_{xx}(\tau)| \leq R_{xx}(0).$$

Proof: Consider two random variables $X(t_1)$ and $X(t_2)$ of $X(t)$ defined at time intervals t_1 and t_2 respectively. Consider a positive quantity $[X(t_1) \pm X(t_2)]^2 \geq 0$

Taking Expectation on both sides, we get $E\{[X(t_1) \pm X(t_2)]^2\} \geq 0$

$$E[X^2(t_1) + X^2(t_2) \pm 2X(t_1) X(t_2)] \geq 0$$

$$E[X^2(t_1)] + E[X^2(t_2)] \pm 2E[X(t_1) X(t_2)] \geq 0$$

$$R_{xx}(0) + R_{xx}(0) \pm 2 R_{xx}(t_1, t_2) \geq 0 \quad [\text{Since } E[X^2(t)] = R_{xx}(0)]$$

Given $X(t)$ is WSS and $\tau = t_2 - t_1$.

$$\text{Therefore } 2 R_{xx}(0) \pm 2 R_{xx}(\tau) \geq 0$$

$$R_{xx}(0) \pm R_{xx}(\tau) \geq 0 \text{ or}$$

$$|R_{xx}(\tau)| \leq R_{xx}(0).$$

3. $R_{XX}(\tau)$ is an even function of τ i.e. $R_{XX}(-\tau) = R_{XX}(\tau)$.

Proof: We know that $R_{XX}(\tau) = E[X(t) X(t+\tau)]$

Let $\tau = -\tau$ then

$$R_{XX}(-\tau) = E[X(t) X(t-\tau)]$$

Let $u = t - \tau$ or $t = u + \tau$

$$\text{Therefore } R_{XX}(-\tau) = E[X(u+\tau) X(u)] = E[X(u) X(u+\tau)]$$

$$R_{XX}(-\tau) = R_{XX}(\tau) \text{ hence proved.}$$

4. If a random process $X(t)$ has a non zero mean value, $E[X(t)] \neq 0$ and Ergodic with no periodic components, then

$$\lim_{|\tau| \rightarrow \infty} R_{XX}(\tau) = \bar{X}^2.$$

Proof: Consider a random variable $X(t)$ with random variables $X(t_1)$ and $X(t_2)$. Given the mean value is

$$E[X(t)] = \bar{X} \neq 0$$

We know that

$R_{XX}(\tau) = E[X(t)X(t+\tau)] = E[X(t_1) X(t_2)]$. Since the process has no periodic components, as $|\tau|$ tends to ∞ , the random variable becomes independent, i.e.

$$\lim_{|\tau| \rightarrow \infty} R_{XX}(\tau) = E[X(t_1) X(t_2)] = E[X(t_1)] E[X(t_2)]$$

Since $X(t)$ is Ergodic $E[X(t_1)] = E[X(t_2)] = \bar{X}$ so it satisfies above property.

5. If $X(t)$ is periodic then its autocorrelation function is also periodic.

Proof: Consider a Random process $X(t)$ which is periodic with period T_0 Then

$$X(t) = X(t \pm T_0) \text{ or}$$

$$X(t+\tau) = X(t+\tau \pm T_0). \text{ Now we have } R_{XX}(\tau) = E[X(t)X(t+\tau)] \text{ then}$$

$$R_{XX}(\tau \pm T_0) = E[X(t)X(t+\tau \pm T_0)]$$

$$\text{Given } X(t) \text{ is WSS, } R_{XX}(\tau \pm T_0) = E[X(t)X(t+\tau)]$$

$$R_{XX}(\tau \pm T_0) = R_{XX}(\tau)$$

Therefore $R_{XX}(\tau)$ is periodic hence proved.

6. The autocorrelation function of random process $R_{XX}(\tau)$ cannot have any arbitrary shape.

Proof:

The autocorrelation function $R_{XX}(\tau)$ is an even function of τ and has maximum value at the origin. Hence the autocorrelation function cannot have an arbitrary shape hence proved.

7. If a random process $X(t)$ with zero mean has the DC component A as $Y(t) = A + X(t)$, Then $R_{YY}(\tau) = A^2 + R_{XX}(\tau)$.

Proof: Given a random process $Y(t) = A + X(t)$.

We know that $R_{YY}(\tau) = E[Y(t)Y(t+\tau)] = E[(A + X(t))(A + X(t+\tau))]$

$$= E[(A^2 + AX(t) + AX(t+\tau) + X(t)X(t+\tau))]$$

$$= E[A^2] + AE[X(t)] + E[AX(t+\tau)] + E[X(t)X(t+\tau)]$$

$$= A^2 + 0 + 0 + R_{XX}(\tau).$$

Therefore $R_{YY}(\tau) = A^2 + R_{XX}(\tau)$ hence proved.

8. If a random process $Z(t)$ is sum of two random processes $X(t)$ and $Y(t)$ i.e., $Z(t) = X(t) + Y(t)$. Then $R_{ZZ}(\tau) = R_{XX}(\tau) + R_{XY}(\tau) + R_{YX}(\tau) + R_{YY}(\tau)$ Proof: Given $Z(t) = X(t) + Y(t)$.

We know that $R_{ZZ}(\tau) = E[Z(t)Z(t+\tau)]$

$$= E[(X(t) + Y(t))(X(t+\tau) + Y(t+\tau))]$$

$$= E[X(t)X(t+\tau) + X(t)Y(t+\tau) + Y(t)X(t+\tau) + Y(t)Y(t+\tau)]$$

$$= E[X(t)X(t+\tau)] + E[X(t)Y(t+\tau)] + E[Y(t)X(t+\tau)] + E[Y(t)Y(t+\tau)]$$

Therefore $R_{ZZ}(\tau) = R_{XX}(\tau) + R_{XY}(\tau) + R_{YX}(\tau) + R_{YY}(\tau)$ hence proved.

Properties of Cross Correlation Function: Consider two random processes $X(t)$ and $Y(t)$ are at least jointly WSS. And the cross correlation function is a function of the time difference $\tau = t_2 - t_1$. Then the following are the properties of cross correlation function.

- i) $R_{XY}(\tau) = R_{YX}(-\tau)$ is a Symmetrical property.

Proof: We know that $R_{XY}(\tau) = E[X(t)Y(t+\tau)]$

$$\text{also } R_{YX}(\tau) = E[Y(t)X(t+\tau)]$$

Let $\tau = -\tau$ then

$$R_{YX}(-\tau) = E[Y(t)X(t-\tau)]$$

Let $u = t - \tau$ or $t = u + \tau$. then

$$R_{YX}(-\tau) = E[Y(u+\tau)X(u)] = E[X(u)Y(u+\tau)]$$

Therefore $R_{YX}(-\tau) = R_{XY}(\tau)$ hence proved.

$$(ii) |R_{XY}(\tau)| \leq \sqrt{R_X(0)R_Y(0)}$$

We have

$$|R_{XY}(\tau)|^2 = |EX(t+\tau)Y(t)|^2$$

$$\leq EX^2(t+\tau)EY^2(t) \quad \text{using Cauch-Schwartz Inequality}$$

$$= R_X(0)R_Y(0)$$

$$\therefore |R_{XY}(\tau)| \leq \sqrt{R_X(0)R_Y(0)}$$

Further,

$$\sqrt{R_X(0)R_Y(0)} \leq \frac{1}{2}(R_X(0) + R_Y(0)) \quad \because \text{Geometric mean} \leq \text{Arithmetic mean}$$

$$\therefore |R_{XY}(\tau)| \leq \sqrt{R_X(0)R_Y(0)} \leq \frac{1}{2}(R_X(0) + R_Y(0))$$

Hence the absolute value of the cross correlation function is always less than or equal to the geometrical mean of the autocorrelation functions.

- (iii) If $X(t)$ and $Y(t)$ are uncorrelated, then $R_{XY}(\tau) = EX(t+\tau)EY(t) = \mu_X \mu_Y$

- (iv) If $X(t)$ and $Y(t)$ is orthogonal process, $R_{XY}(\tau) = EX(t+\tau)Y(t) = 0$

Example Suppose $X(t) = A\cos(w_0 t + \phi_1)$ and $Y(t) = A\sin(w_0 t + \phi_2)$ where A and w_0 are constants and ϕ_1 and ϕ_2 are independent random variables each uniformly distributed between 0 and 2π . Then

$$R_{XY}(t_1, t_2) = EX(t_1)X(t_2)$$

$$= EA\cos(w_0 t_1 + \phi_1)A\sin(w_0 t + \phi_2)$$

$$= EA\cos(w_0 t_1 + \phi_1)EA\sin(w_0 t + \phi_2)$$

$$= 0 \times 0 = 0$$

$\because \phi_1$ and ϕ_2 are independent

Therefore, random processes $\{X(t), t \in \Gamma\}$ and $\{Y(t), t \in \Gamma\}$ are orthogonal.

Covariance functions for random processes:

Auto Covariance function:

Consider two random processes $X(t)$ and $X(t + \tau)$ at two time intervals t and $t + \tau$. The auto covariance function can be expressed as

$$C_{XX}(t, t + \tau) = E[(X(t) - E[X(t)])((X(t + \tau) - E[X(t + \tau)])] \text{ or}$$

$$C_{XX}(t, t + \tau) = R_{XX}(t, t + \tau) - E[X(t)]E[X(t + \tau)]$$

Properties:

1. If $X(t)$ is WSS, then $C_{XX}(\tau) = R_{XX}(\tau) - \bar{X}^2$ since $E[X(t)] = E[X(t + \tau)] = \bar{X}$
2. $C_{XX}(0) = R_{XX}(0) - \bar{X}^2 = E[X^2(t)] - \{E[X(t)]\}^2 = \sigma_X^2$

Therefore at $\tau = 0$, the auto covariance function becomes the Variance of the random process.

3. The autocorrelation coefficient of the random process, $X(t)$ is defined as

$$\rho_{XX}(t, t + \tau) = \frac{C_{XX}(t, t + \tau)}{\sqrt{C_{XX}(t, t)C_{XX}(t + \tau, t + \tau)}} \text{ if } \tau \neq 0,$$

$$\rho_{XX}(0) = \frac{C_{XX}(t, t)}{C_{XX}(t, t)} = 1. \text{ Also } |\rho_{XX}(t, t + \tau)| \leq 1.$$

Cross Covariance Function:

If two random processes $X(t)$ and $Y(t)$ have random variables $X(t)$ and $Y(t + \tau)$, then the cross covariance function can be defined as

$$C_{XY}(t, t + \tau) = E[(X(t) - E[X(t)])((Y(t + \tau) - E[Y(t + \tau)])] \text{ or}$$

$$C_{XY}(t, t + \tau) = R_{XY}(t, t + \tau) - E[X(t)]E[Y(t + \tau)].$$

Properties:

1. If $X(t)$ and $Y(t)$ are jointly WSS, then

$$C_{XY}(\tau) = R_{XY}(\tau) - \bar{X}\bar{Y}$$

2. If $X(t)$ and $Y(t)$ are Uncorrelated then $C_{XY}(t, t+\tau) = 0$.

3. The cross correlation coefficient of random processes $X(t)$ and $Y(t)$ is defined as

$$\rho_{XY}(t, t+\tau) = \frac{C_{XY}(t, t+\tau)}{\sqrt{C_{XX}(t, t)C_{YY}(t+\tau, t+\tau)}} \text{ if } \tau = 0,$$

$$\rho_{XY}(0) = \frac{C_{XY}(0)}{\sqrt{C_{XX}(0)C_{YY}(0)}} = \frac{C_{XY}(0)}{\sigma_X \sigma_Y}.$$

Response of Linear time-invariant system to WSS input

In many applications, physical systems are modeled as linear time invariant (LTI) systems. The dynamic behavior of an LTI system to deterministic inputs is described by linear differential equations. We are familiar with time and transform domain (such as Laplace transform and Fourier transform) techniques to solve these equations. In this lecture, we develop the technique to analyze the response of an LTI system to WSS random process.

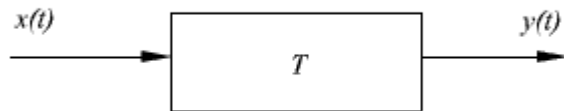
The purpose of this study is two-folds:

- (1) Analysis of the response of a system
- (2) Finding an LTI system that can optionally estimate an unobserved random process from an observed process. The observed random process is statistically related to the unobserved random process. For example, we may have to find LTI system (also called a filter) to estimate the signal from the noisy observations.

Basics of Linear Time Invariant Systems:

A system is modeled by a transformation T that maps an input signal $x(t)$ to an output signal $y(t)$. We can thus write,

$$y(t) = T[x(t)]$$



Linear system

The system is called linear if superposition applies: **the weighted sum of inputs results in the weighted sum of the corresponding outputs.** Thus for a linear system

$$T[a_1x_1(t) + a_2x_2(t)] = a_1T[x_1(t)] + a_2T[x_2(t)]$$

Example : Consider the output of a linear differentiator, given by

$$y(t) = \frac{d x(t)}{dt}$$

Then, $\frac{d}{dt} (a_1x_1(t) + a_2x_2(t))$

$$= a_1 \frac{d}{dt} x_1(t) + a_2 \frac{d}{dt} x_2(t)$$

Hence the linear differentiator is a linear system.

Linear time-invariant system

Consider a linear system with $y(t) = T x(t)$. The system is called time-invariant if

$$T x(t - t_0) = y(t - t_0) \quad \forall t_0$$

It is easy to check that the differentiator in the above example is a linear time-invariant system.

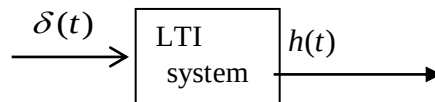
Causal system

The system is called causal if the output of the system at $t = t_0$ depends only on the present and past values of input. Thus for a causal system

$$y(t_0) = T(x(t), t \leq t_0)$$

Response of a linear time-invariant system to deterministic input

A linear system can be characterised by its impulse response $h(t) = T \delta(t)$ where $\delta(t)$ is the *Dirac delta function*.



Recall that any function $x(t)$ can be represented in terms of the Dirac delta function as follows

$$x(t) = \int_{-\infty}^{\infty} x(\tau) \delta(t - s) ds$$

If $x(t)$ is input to the linear system $y(t) = T x(t)$, then

$$\begin{aligned} y(t) &= T \int_{-\infty}^{\infty} x(s) \delta(t - s) ds \\ &= \int_{-\infty}^{\infty} x(s) T \delta(t - s) ds \quad [\text{Using the linearity property}] \\ &= \int_{-\infty}^{\infty} x(s) h(t, s) ds \end{aligned}$$

Where $h(t, s) = T \delta(t - s)$ is the response at time t due to the shifted impulse $\delta(t - s)$.

If the system is time invariant,

$$h(t, s) = h(t - s)$$

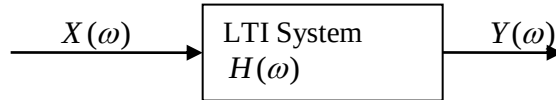
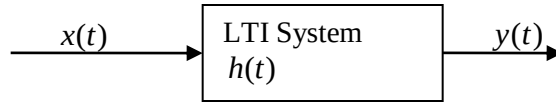
Therefore for a linear time invariant system,

$$y(t) = \int_{-\infty}^{\infty} x(s) h(t - s) ds = x(t) * h(t)$$

where $*$ denotes the convolution operation.

Thus for a LTI System,

$$\boxed{y(t) = x(t) * h(t) = h(t) * x(t)}$$



Taking the Fourier transform, we get

$$Y(\omega) = H(\omega) X(\omega)$$

where $H(\omega) = FT h(t) = \int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt$ is the frequency response of the system

Response of an LTI System to WSS input

Consider an LTI system with impulse response $h(t)$. Suppose $\{X(t)\}$ is a WSS process input to the system. The output $\{Y(t)\}$ of the system is given by

$$Y(t) = \int_{-\infty}^{\infty} h(s) X(t-s) ds = \int_{-\infty}^{\infty} h(t-s) X(s) ds$$

where we have assumed that the integrals exist in the mean square (m.s.) sense.

Mean and autocorrelation of the output process $\{Y(t)\}$

The mean of the output process is given by,

$$\begin{aligned} EY(t) &= E \int_{-\infty}^{\infty} h(s) X(t-s) ds \\ &= \int_{-\infty}^{\infty} h(s) EX(t-s) ds \\ &= \int_{-\infty}^{\infty} h(s) \mu_X ds \\ &= \mu_X \int_{-\infty}^{\infty} h(s) ds \\ &= \mu_X H(0) \end{aligned}$$

where $H(0)$ is the frequency response $H(\omega)$ at 0 frequency ($\omega = 0$) given by

$$H(\omega)|_{\omega=0} = \int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt \Big|_{\omega=0} = \int_{-\infty}^{\infty} h(t) dt$$

Therefore, the mean of the output process $\{Y(t)\}$ is a constant

Cross correlation function of input and output of LTI system:

The Cross correlation of the input $X(t)$ and the out put $Y(t)$ is given by

$$\begin{aligned} E(X(t+\tau)Y(t)) &= E X(t+\tau) \int_{-\infty}^{\infty} h(s) X(t-s) ds \\ &= \int_{-\infty}^{\infty} h(s) E X(t+\tau) X(t-s) ds \\ &= \int_{-\infty}^{\infty} h(s) R_X(\tau+s) ds \\ &= \int_{-\infty}^{\infty} h(-u) R_X(\tau-u) du \quad [\text{ Put } s = -u] \\ &= h(-\tau) * R_X(\tau) \end{aligned}$$

$$\begin{aligned} \therefore R_{XY}(\tau) &= h(-\tau) * R_X(\tau) \\ \text{also } R_{YX}(\tau) &= R_{XY}(-\tau) = h(\tau) * R_X(-\tau) \\ &= h(\tau) * R_X(\tau) \end{aligned}$$

Thus we see that $R_{XY}(\tau)$ is a function of lag τ only. Therefore, $X(t)$ and $Y(t)$ are jointly wide-sense stationary.

The autocorrelation function of the output process $Y(t)$ is given by,

$$\begin{aligned} \therefore E(Y(t+\tau)Y(t)) &= E \int_{-\infty}^{\infty} h(s) X(t+\tau-s) ds Y(t) \\ &= \int_{-\infty}^{\infty} h(s) E X(t+\tau-s) Y(t) ds \\ &= \int_{-\infty}^{\infty} h(s) R_{XY}(\tau-s) ds \\ &= h(\tau) * R_{XY}(\tau) = h(\tau) * h(-\tau) * R_X(\tau) \end{aligned}$$

Thus the autocorrelation of the output process $\{Y(t)\}$ depends on the time-lag τ , i.e., $EY(t)Y(t+\tau) = R_Y(\tau)$.

Thus

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$$

The above analysis indicates that for an LTI system with WSS input

- (1) the output is WSS and
- (2) the input and output are jointly WSS.

The average power of the output process $\{Y(t)\}$ is given by

$$\begin{aligned} P_Y &= R_Y(0) \\ &= R_X(0) * h(0) * h(0) \end{aligned}$$

UNIT V

Stochastic Processes-Spectral Characteristics

- The Power Spectrum and its Properties
- Relationship between Power Spectrum and Autocorrelation Function
- The Cross-Power Density Spectrum and Properties
- Relationship between Cross-Power Spectrum and Cross-Correlation Function.

Spectral characteristics of system response:

- Power density spectrum of response
- Cross power spectral density of input and output of a linear system

Introduction:

In this unit we will study the characteristics of random processes regarding correlation and covariance functions which are defined in time domain. This unit explores the important concept of characterizing random processes in the frequency domain. These characteristics are called spectral characteristics. All the concepts in this unit can be easily learnt from the theory of Fourier transforms.

Consider a random process $X(t)$. The amplitude of the random process, when it varies randomly with time, does not satisfy Dirichlet's conditions. Therefore it is not possible to apply the Fourier transform directly on the random process for a frequency domain analysis. Thus the autocorrelation function of a WSS random process is used to study spectral characteristics such as power density spectrum or power spectral density (psd).

Power Density Spectrum: The power spectrum of a WSS random process $X(t)$ is defined as the Fourier transform of the autocorrelation function $R_{XX}(\tau)$ of $X(t)$. It can be expressed as

$$S_{XX}(\omega) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-j\omega\tau} d\tau$$

We can obtain the auto correlation function from power spectral density by taking the inverse

Fourier transform i.e

$$R_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XX}(\omega) e^{j\omega\tau} d\omega$$

Therefore, the power density spectrum $S_{XX}(\omega)$ and the autocorrelation function $R_{XX}(\tau)$ are Fourier transform pairs.

Definition of Power Spectral Density of a WSS Process

Let us define

$$\begin{aligned} X_T(t) &= X(t) & -T < t < T \\ &= 0 & \text{otherwise} \\ &= X(t) \text{rect}\left(\frac{t}{2T}\right) \end{aligned}$$

where $\text{rect}\left(\frac{t}{2T}\right)$ is the unity-amplitude rectangular pulse of width $2T$ centering the origin. As

$t \rightarrow \infty$, $X_T(t)$ will represent the random process $X(t)$.

Define the mean-square integral

$$FTX_T(\omega) = \int_{-T}^T X_T(t) e^{-j\omega t} dt$$

Applying the Parseval's theorem we find the energy of the signal

$$\int_{-T}^T X_T^2(t) dt = \int_{-\infty}^{\infty} |FTX_T(\omega)|^2 d\omega$$

Therefore, the power associated with $X_T(t)$ is

$$\frac{1}{2T} \int_{-T}^T X_T^2(t) dt = \frac{1}{2T} \int_{-\infty}^{\infty} |FTX_T(\omega)|^2 d\omega \text{ and}$$

The average power is given by

$$\frac{1}{2T} E \int_{-T}^T X_T^2(t) dt = \frac{1}{2T} E \int_{-\infty}^{\infty} |FTX_T(\omega)|^2 d\omega = E \int_{-\infty}^{\infty} \frac{|FTX_T(\omega)|^2}{2T} d\omega$$

where $\frac{E |FTX_T(\omega)|^2}{2T}$ is the contribution to the average power at frequency ω and represents the power spectral density for $X_T(t)$. As $T \rightarrow \infty$, the left-hand side in the above expression represents the average power of $X(t)$. Therefore, the PSD $S_X(\omega)$ of the process $X(t)$ is defined in the limiting sense by

$$S_X(\omega) = \lim_{T \rightarrow \infty} \frac{E |FTX_T(\omega)|^2}{2T}$$

The power spectral density can also be defined as

$$S_{XX}(\omega) = \lim_{T \rightarrow \infty} \frac{E[|X_{T(\omega)}|^2]}{2T}$$

Where $X_T(\omega)$ is a Fourier transform of $X(t)$ in interval $[-T, T]$

Average power of the random process: The average power P_{XX} of a WSS random process $X(t)$ is defined as the time average of its second order moment or autocorrelation function at $\tau = 0$.

Mathematically, $P_{XX} = A \{E[X^2(t)]\}$

$$P_{XX} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E[X^2(t)] dt$$

$$P_{XX} = R_{XX}(\tau) \text{ at } \tau = 0$$

We know that from the power density spectrum,

$$R_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XX}(\omega) e^{j\omega\tau} d\omega$$

$$\text{at } \tau = 0 \quad P_{XX} = R_{XX}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XX}(\omega) d\omega$$

Therefore average power of $X(t)$ is

$$P_{XX} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XX}(\omega) d\omega$$

Properties of power density spectrum: The properties of the power density spectrum $S_{XX}(\omega)$ for a WSS random process $X(t)$ are given as

1. The average power of a random process $X(t)$ is

$$\begin{aligned} EX^2(t) &= R_X(0) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega \end{aligned}$$

The average power in the band $[\omega_1, \omega_2]$ is $2 \int_{\omega_1}^{\omega_2} S_X(\omega) d\omega$

2. If $\{X(t)\}$ is real, $R_X(\tau)$ is a real and even function of τ . Therefore,

$$\begin{aligned} S_X(\omega) &= \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} R_X(\tau) (\cos \omega\tau + j \sin \omega\tau) d\tau \\ &= \int_{-\infty}^{\infty} R_X(\tau) \cos \omega\tau d\tau \\ &= 2 \int_0^{\infty} R_X(\tau) \cos \omega\tau d\tau \end{aligned}$$

Average Power in
the frequency band
 $[\omega_1, \omega_2]$

Thus $S_X(\omega)$ is a real and even function of ω .

3. From the definition $S_X(\omega) = \lim_{T \rightarrow \infty} \frac{E|X_T(\omega)|^2}{2T}$ is always non-negative. Thus $S_X(\omega) \geq 0$.
4. If $\{X(t)\}$ has a periodic component, $R_X(\tau)$ is periodic and so $S_X(\omega)$ will have impulses.
5. If $X(t)$ is a WSS random process with psd $S_{XX}(\omega)$, then the psd of the derivative of $X(t)$ is equal to ω^2 times the psd $S_{XX}(\omega)$. That is

$$S_{\dot{X}\dot{X}}(\omega) = \omega^2 S_{XX}(\omega)$$

Proof: We know that $S_{XX}(\omega) = \lim_{T \rightarrow \infty} \frac{E[|X_T(\omega)|^2]}{2T}$ and $X_T(\omega) = \int_{-T}^T X(t) e^{-j\omega t} dt$

$$= \dot{X}_T(\omega) = \frac{d X_T(\omega)}{dt} = \frac{d}{dt} \int_{-T}^T X(t) e^{-j\omega t} dt$$

$$= \int_{-T}^T X(t) \frac{d}{dt} e^{-j\omega t} dt$$

$$= \int_{-T}^T X(t) (-j\omega) e^{-j\omega t} dt$$

$$= (-j\omega) \int_{-T}^T X(t) e^{-j\omega t} dt$$

$$\text{Therefore } S \dot{X} \dot{X}(\omega) = \lim_{T \rightarrow \infty} \frac{E[|\dot{X}_T(\omega)|^2]}{2T}$$

$$= \lim_{T \rightarrow \infty} \frac{E[|(-j\omega) X_T(\omega)|^2]}{2T}$$

$$= \lim_{T \rightarrow \infty} \frac{\omega^2 E[|(-j\omega) X_T(\omega)|^2]}{2T}$$

$$S \dot{X} \dot{X}(\omega) = \omega^2 \lim_{T \rightarrow \infty} \frac{E[|X_T(\omega)|^2]}{2T}$$

Therefore $S \dot{X} \dot{X}(\omega) = \omega^2 S_{XX}(\omega)$ hence proved.

Remark

- The function $S_X(\omega)$ is the PSD of a WSS process $\{X(t)\}$ if and only if $S_X(\omega)$ is a non-negative, real and even function of ω and $\int_{-\infty}^{\infty} S_X(\omega) d\omega < \infty$
- The above condition on $S_X(\omega)$ also ensures that the corresponding autocorrelation function $R_X(\tau)$ is non-negative definite. Thus the non-negative definite property of an autocorrelation function can be tested through its power spectrum.
- Recall that a periodic function has the *Fourier series* expansion. If $\{X(t)\}$ is M.S. periodic we can have an equivalent *Fourier series* expansion $\{X(t)\}$.

Relation between the autocorrelation function and PSD: Wiener-Khinchin-Einstein theorem:

Statement:

It states that PSD and Auto correlation function of a WSS random process form Fourier transform pair. i.e

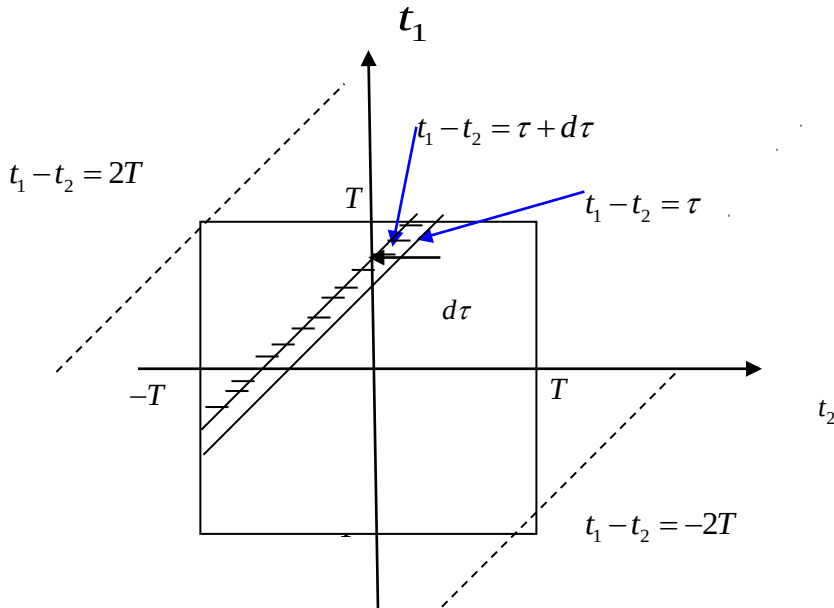
$$S_{XX}(\omega) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-j\omega\tau} d\tau \quad \text{And}$$

$$R_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XX}(\omega) e^{j\omega\tau} d\omega$$

Proof:

We have

$$\begin{aligned} E \frac{|FTX_T(\omega)|^2}{2T} &= E \frac{FTX_T(\omega) FTX_T^*(\omega)}{2T} \\ &= \frac{1}{2T} \int_{-T}^T \int_{-T}^T EX_T(t_1) X_T(t_2) e^{-j\omega t_1} e^{+j\omega t_2} dt_1 dt_2 \\ &= \frac{1}{2T} \int_{-T}^T \int_{-T}^T R_X(t_1 - t_2) e^{-j\omega(t_1 - t_2)} dt_1 dt_2 \end{aligned}$$



Note that the above integral is to be performed on a square region bounded by $t_1 = \pm T$ and $t_2 = \pm T$.

Substitute $t_1 - t_2 = \tau$ so that $t_2 = t_1 - \tau$ is a family of straight lines parallel to $t_1 - t_2 = 0$. The differential area in terms of τ is given by the shaded area and equal to $(2T - |\tau|)d\tau$. The double integral is now replaced by a single integral in τ . Therefore,

$$\begin{aligned} E \frac{FTX_T(\omega) X_T^*(\omega)}{2T} &= \frac{1}{2T} \int_{-2T}^{2T} R_X(\tau) e^{-j\omega\tau} (2T - |\tau|) d\tau \\ &= \int_{-2T}^{2T} R_X(\tau) e^{-j\omega\tau} \left(1 - \frac{|\tau|}{2T}\right) d\tau \end{aligned}$$

If $R_X(\tau)$ is integrable then the right hand integral converges to $\int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau$ as $T \rightarrow \infty$,

$$\therefore \lim_{T \rightarrow \infty} \frac{E|FTX_T(\omega)|^2}{2T} = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau$$

As we have noted earlier, the power spectral density $S_X(\omega) = \lim_{T \rightarrow \infty} \frac{E|FTX_T(\omega)|^2}{2T}$ is the contribution to the average power at frequency ω and is called the power spectral density $X(t)$. Thus

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau$$

and using the inverse Fourier transform

$$R_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) e^{j\omega\tau} d\omega$$

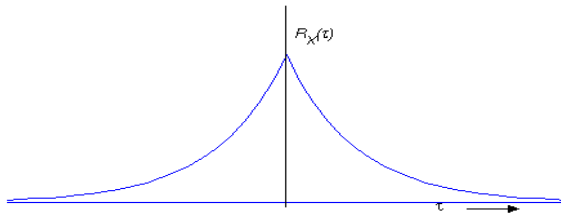
Example 1 The autocorrelation function of a WSS process $X(t)$ is given by

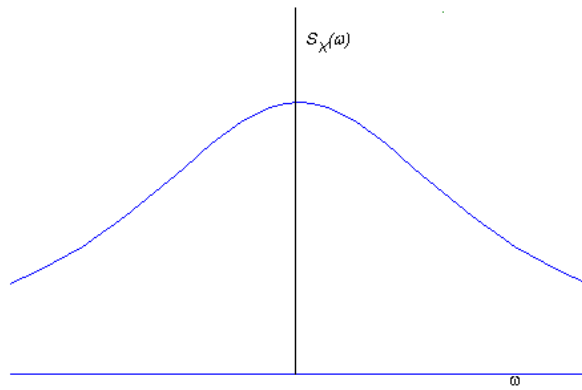
$$R_X(\tau) = a^2 e^{-b|\tau|} \quad b > 0$$

Find the power spectral density of the process.

$$\begin{aligned} S_X(\omega) &= \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} a^2 e^{-b|\tau|} e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^0 a^2 e^{b\tau} e^{-j\omega\tau} d\tau + \int_0^{\infty} a^2 e^{-b\tau} e^{-j\omega\tau} d\tau \\ &= \frac{a^2}{b - j\omega} + \frac{a^2}{b + j\omega} \\ &= \frac{2a^2 b}{b^2 + \omega^2} \end{aligned}$$

The autocorrelation function and the PSD are shown in Fig.





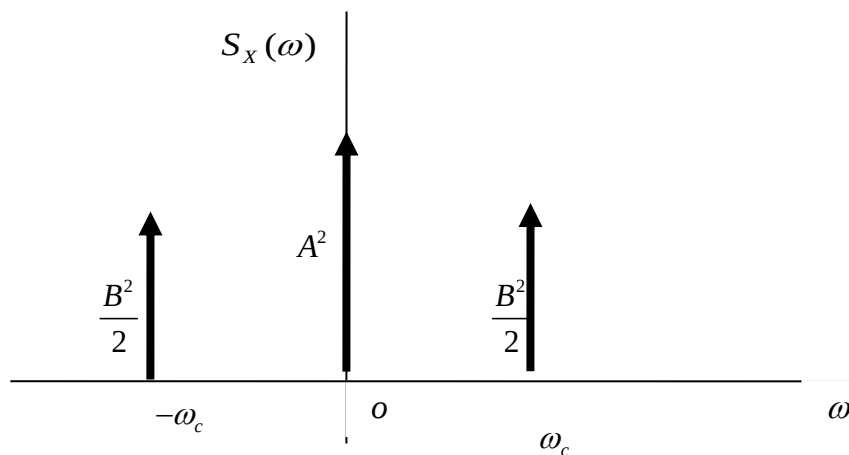
Example2: Suppose $X(t) = A + B \sin(\omega_c t + \Phi)$ where A is a constant bias and $\Phi \sim U[0, 2\pi]$.

Find $R_X(\tau)$ and $S_X(\omega)$.

$$\begin{aligned} R_X(\tau) &= E[X(t+\tau)X(t)] \\ &= E[(A + B \sin(\omega_c(t+\tau) + \Phi))(A + B \sin(\omega_c t + \Phi))] \\ &= A^2 + \frac{B^2}{2} \cos \omega_c \tau \end{aligned}$$

$$\therefore S_X(\omega) = A^2 \delta(\omega) + \frac{B^2}{4} (\delta(\omega + \omega_c) + \delta(\omega - \omega_c))$$

where $\delta(\omega)$ is the Dirac Delta function.



Example 3. Find PSD of the amplitude-modulated random-phase sinusoid

$$X(t) = M(t) \cos(\omega_c t + \Phi), \quad \Phi \sim U(0, 2\pi)$$

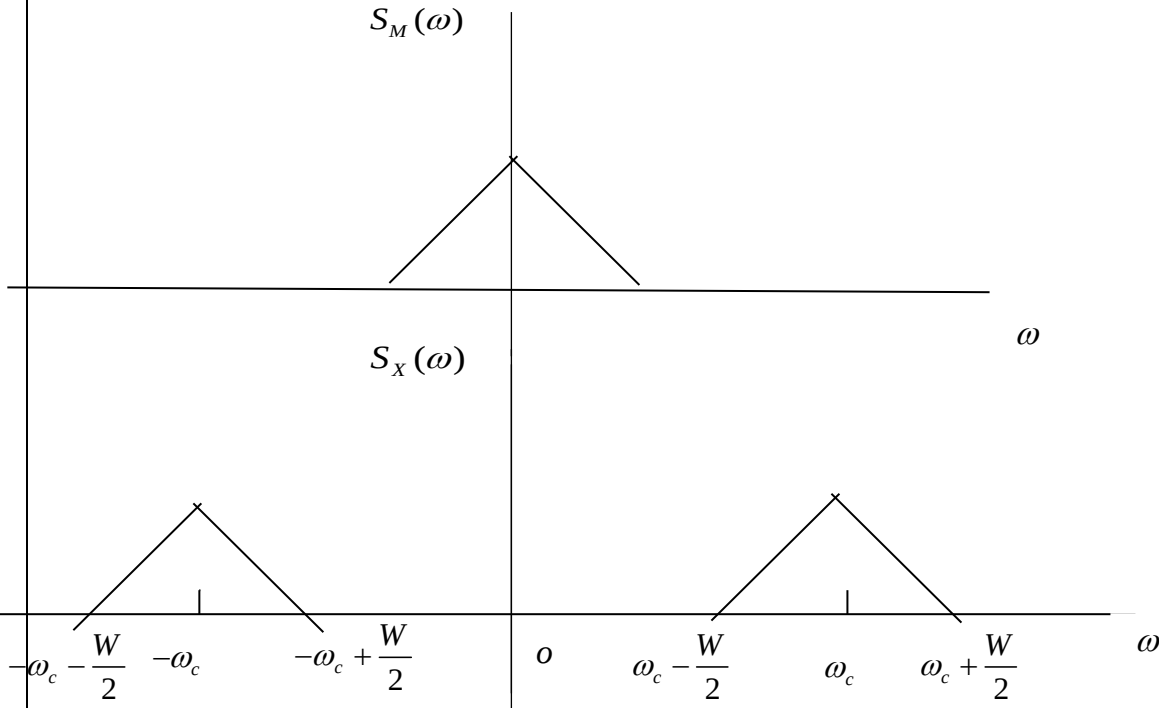
where $M(t)$ is a WSS process independent of Φ .

$$\begin{aligned}
\Rightarrow R_X(\tau) &= E M(t+\tau) \cos(\omega_c(t+\tau) + \Phi) M(t) \cos(\omega_c \tau + \Phi) \\
&= E M(t+\tau) M(t) E \cos(\omega_c(t+\tau) + \Phi) \cos(\omega_c \tau + \Phi) \\
&\quad \text{(Using the independence of } M(t) \text{ and the sinusoid)}
\end{aligned}$$

$$= R_M(\tau) \frac{A^2}{2} \cos \omega_c \tau$$

$$\therefore S_X(\omega) = \frac{A^2}{4} (S_M(\omega + \omega_c) + S_M(\omega - \omega_c))$$

where $S_M(\omega)$ is the PSD of $M(t)$



Example 4 The PSD of a noise process is given by

$$\begin{aligned}
S_N(\omega) &= \frac{N_0}{2} \quad |\omega \pm \omega_c| \leq \frac{W}{2} \\
&= 0 \quad \text{Otherwise}
\end{aligned}$$

Find the autocorrelation of the process.

$$\begin{aligned}
\therefore R_N(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_N(\omega) e^{j\omega\tau} d\omega \\
&= \frac{1}{2\pi} \times 2 \times \int_{\omega_c - \frac{W}{2}}^{\omega_c + \frac{W}{2}} \frac{N_0}{2} \cos \omega\tau d\omega \\
&= \frac{N_0}{2\pi} \left[\frac{\sin\left(\omega_c + \frac{W}{2}\right)\tau - \sin\left(\omega_c - \frac{W}{2}\right)\tau}{\tau} \right] \\
&= \frac{N_0 W}{2\pi} \frac{\sin \frac{W\tau}{2}}{\frac{W\tau}{2}} \cos \omega_c \tau
\end{aligned}$$

Example 5: The autocorrelation function of a WSS process $X(t)$ is given by

$$R_X(\tau) = a^2 e^{-b|\tau|} \quad b > 0$$

Find PSD of the random process $X(t)$.

Solution:

$$\begin{aligned} S_X(\omega) &= \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} a^2 e^{-b|\tau|} e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^0 a^2 e^{b\tau} e^{-j\omega\tau} d\tau + \int_0^{\infty} a^2 e^{-b\tau} e^{-j\omega\tau} d\tau \\ &= \frac{a^2}{b - j\omega} + \frac{a^2}{b + j\omega} \\ &= \frac{2a^2 b}{b^2 + \omega^2} \end{aligned}$$

Cross power spectral density:

Consider a random process $Z(t)$ which is sum of two real jointly WSS random processes $X(t)$ and $Y(t)$.

As we have seen earlier,

$$R_Z(\tau) = R_X(\tau) + R_Y(\tau) + R_{XY}(\tau) + R_{YX}(\tau)$$

If we take the Fourier transform of both sides,

$$S_Z(\omega) = S_X(\omega) + S_Y(\omega) + FT(R_{XY}(\tau)) + FT(R_{YX}(\tau))$$

where $FT(.)$ stands for the Fourier transform.

Thus we see that $S_Z(\omega)$ includes contribution from the Fourier transform of the cross-correlation functions $R_{XY}(\tau)$ and $R_{YX}(\tau)$. These Fourier transforms represent cross power spectral densities.

Definition of Cross Power Spectral Density

Given two real jointly WSS random processes $X(t)$ and $Y(t)$, the cross power spectral density (CPSD)

$S_{XY}(\omega)$ is defined as

$$S_{XY}(\omega) = \lim_{T \rightarrow \infty} E \frac{FTX_T^*(\omega) FTY_T(\omega)}{2T}$$

where $FTX_T(\omega)$ and $FTY_T(\omega)$ are the Fourier transform of the truncated processes

$X_T(t) = X(t) \text{rect}(\frac{t}{2T})$ and $Y_T(t) = Y(t) \text{rect}(\frac{t}{2T})$ respectively and $*$ denotes the complex conjugate

operation.

We can similarly define $S_{YX}(\omega)$ by

$$S_{YX}(\omega) = \lim_{T \rightarrow \infty} E \frac{FTY_T^*(\omega)FTX_T(\omega)}{2T}$$

Proceeding in the same way as the derivation of the Wiener-Khinchin-Einstein theorem for the WSS process, it can be shown that

$$S_{XY}(\omega) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j\omega\tau} d\tau$$

and

$$S_{YX}(\omega) = \int_{-\infty}^{\infty} R_{YX}(\tau) e^{-j\omega\tau} d\tau$$

The cross-correlation function and the cross-power spectral density form a Fourier transform pair and we can write

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} S_{XY}(\omega) e^{j\omega\tau} d\omega$$

and

$$R_{YX}(\tau) = \int_{-\infty}^{\infty} S_{YX}(\omega) e^{j\omega\tau} d\omega$$

Properties of the CPSD

The CPSD is a complex function of the frequency ω . Some properties of the CPSD of two jointly WSS processes $X(t)$ and $Y(t)$ are listed below:

$$(1) \quad S_{XY}(\omega) = S_{YX}^*(\omega)$$

Note that $R_{XY}(\tau) = R_{YX}(-\tau)$

$$\begin{aligned} \therefore S_{XY}(\omega) &= \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} R_{YX}(-\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} R_{YX}(\tau) e^{j\omega\tau} d\tau \\ &= S_{YX}^*(\omega) \end{aligned}$$

(2) $\text{Re}(S_{XY}(\omega))$ is an even function of ω and $\text{Im}(S_{XY}(\omega))$ is an odd function of ω

We have

$$\begin{aligned} S_{XY}(\omega) &= \int_{-\infty}^{\infty} R_{XY}(\tau) (\cos \omega\tau + j \sin \omega\tau) d\tau \\ &= \int_{-\infty}^{\infty} R_{XY}(\tau) \cos \omega\tau d\tau + j \int_{-\infty}^{\infty} R_{XY}(\tau) \sin \omega\tau d\tau \\ &= \text{Re}(S_{XY}(\omega)) + j \text{Im}(S_{XY}(\omega)) \end{aligned}$$

where

$$\text{Re}(S_{XY}(\omega)) = \int_{-\infty}^{\infty} R_{XY}(\tau) \cos \omega\tau d\tau \text{ is an even function of } \omega \text{ and}$$

$$\text{Im}(S_{XY}(\omega)) = \int_{-\infty}^{\infty} R_{XY}(\tau) \sin \omega\tau d\tau \text{ is an odd function of } \omega \text{ and}$$

(3) $X(t)$ and $Y(t)$ are uncorrelated and have constant means, then

$$S_{XY}(\omega) = S_{YX}(\omega) = \mu_X \mu_Y \delta(\omega)$$

Observe that

$$\begin{aligned} R_{XY}(\tau) &= EX(t+\tau)Y(t) \\ &= EX(t+\tau)EY(t) \\ &= \mu_X \mu_Y \\ &= \mu_Y \mu_X \\ &= R_{XY}(\tau) \end{aligned}$$

$$\therefore S_{XY}(\omega) = S_{YX}(\omega) = \mu_X \mu_Y \delta(\omega)$$

(4) If $X(t)$ and $Y(t)$ are orthogonal, then

$$S_{XY}(\omega) = S_{YX}(\omega) = 0$$

If $X(t)$ and $Y(t)$ are orthogonal,

$$\begin{aligned} R_{XY}(\tau) &= EX(t+\tau)Y(t) \\ &= 0 \\ &= R_{XY}(\tau) \end{aligned}$$

$$\therefore S_{XY}(\omega) = S_{YX}(\omega) = 0$$

(5) The cross power P_{XY} between $X(t)$ and $Y(t)$ is defined by

$$P_{XY} = \lim_{T \rightarrow \infty} \frac{1}{2T} E \int_{-T}^T X(t)Y(t)dt$$

Applying Parseval's theorem, we get

$$\begin{aligned} P_{XY} &= \lim_{T \rightarrow \infty} \frac{1}{2T} E \int_{-T}^T X(t)Y(t)dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} E \int_{-\infty}^{\infty} X_T(t)Y_T(t)dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} E \frac{1}{2\pi} \int_{-\infty}^{\infty} FTX_T^*(\omega)FTY_T(\omega)d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{EFTX_T^*(\omega)FTY_T(\omega)}{2T} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}(\omega)d\omega \\ \therefore P_{XY} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}(\omega)d\omega \end{aligned}$$

Similarly,

$$\begin{aligned} P_{YX} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{YX}(\omega)d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}^*(\omega)d\omega \\ &= P_{XY}^* \end{aligned}$$

Example5: Consider the random process $Z(t) = X(t) + Y(t)$ discussed in the beginning of the lecture. Here $Z(t)$ is the sum of two jointly WSS orthogonal random processes $X(t)$ and $Y(t)$.

We have,

$$R_Z(\tau) = R_X(\tau) + R_Y(\tau) + R_{XY}(\tau) + R_{YX}(\tau)$$

Taking the Fourier transform of both sides,

$$S_Z(\omega) = S_X(\omega) + S_Y(\omega) + S_{XY}(\omega) + S_{YX}(\omega)$$

$$\therefore \frac{1}{2\pi} \int_{-\infty}^{\infty} S_Z(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} S_Y(\omega) d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}(\omega) d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{YX}(\omega) d\omega$$

Therefore,

$$\boxed{P_Z(\omega) = P_X(\omega) + P_Y(\omega) + P_{XY}(\omega) + P_{YX}(\omega)}$$

Remark

- $P_{XY}(\omega) + P_{YX}(\omega)$ is the additional power contributed by $X(t)$ and $Y(t)$ to the resulting power of $X(t) + Y(t)$
- If $X(t)$ and $Y(t)$ are orthogonal, then

$$\begin{aligned} S_Z(\omega) &= S_X(\omega) + S_Y(\omega) + 0 + 0 \\ &= S_X(\omega) + S_Y(\omega) \end{aligned}$$

Consequently

$$P_Z(\omega) = P_X(\omega) + P_Y(\omega)$$

Thus in the case of two jointly WSS orthogonal processes, the power of the sum of the processes is equal to the sum of respective powers.

Spectral characteristics of system response:

Power spectrum of the output process

Consider that a random process $X(t)$ is applied on an LTI system having a transfer function $H(\omega)$. The output response is $Y(t)$. If the power spectrum of the input process is $S_{XX}(\omega)$, then the power spectrum of the output response is given by $S_{YY}(\omega) = S_{XX}(\omega)|H(\omega)|^2$.

Proof: Let $R_{YY}(\tau)$ be the autocorrelation of the output response $Y(t)$. Then the power spectrum of the response is the Fourier transform of $R_{YY}(\tau)$.

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$$

Using the property of Fourier transform, we get the power spectral density of the output process given by

$$\begin{aligned} S_Y(\omega) &= S_X(\omega) H(\omega) H^*(\omega) \\ &= S_X(\omega) |H(\omega)|^2 \end{aligned}$$

Also note that

$$R_{XY}(\tau) = h(-\tau) * R_X(\tau)$$

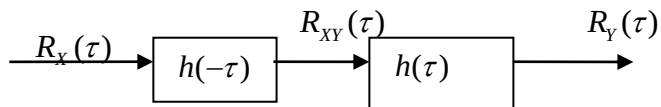
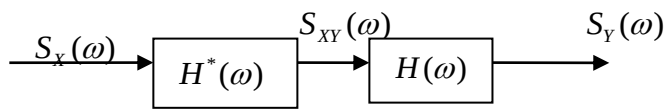
$$\text{and } R_{YX}(\tau) = h(\tau) * R_X(\tau)$$

Taking the Fourier transform of $R_{XY}(\tau)$ we get the cross power spectral density $S_{XY}(\omega)$ given by

$$S_{XY}(\omega) = H^*(\omega) S_X(\omega)$$

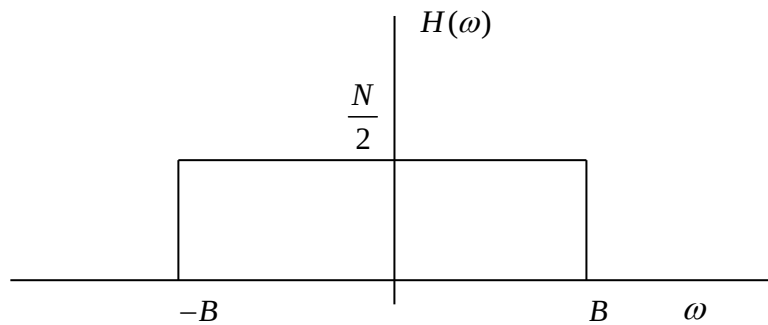
and

$$S_{YX}(\omega) = H(\omega) S_X(\omega)$$



Example:

(a) White noise process $X(t)$ with power spectral density $\frac{N_0}{2}$ is input to an ideal low pass filter of band-width B . Find the PSD and autocorrelation function of the output process.



The input process $X(t)$ is white noise with power spectral density $S_x(\omega) = \frac{N_0}{2}$.

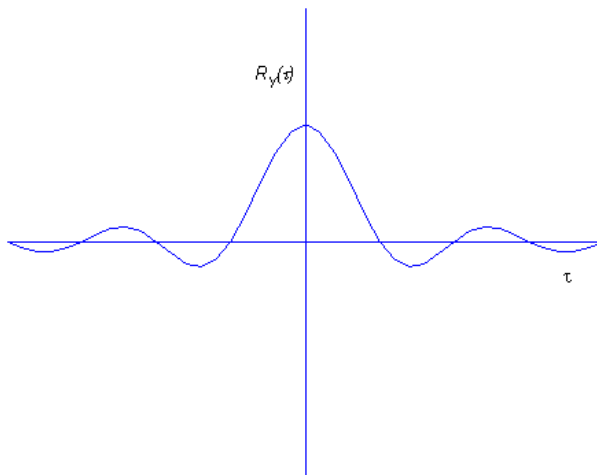
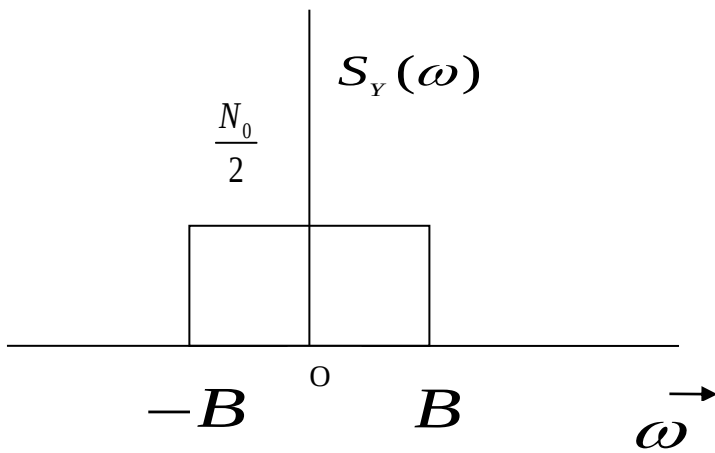
The output power spectral density $S_y(\omega)$ is given by,

$$\begin{aligned}
 S_Y(\omega) &= |H(\omega)|^2 S_X(\omega) \\
 &= 1 \times \frac{N_0}{2} \quad -B \leq \omega \leq B \\
 &= \frac{N_0}{2} \quad -B \leq \omega \leq B
 \end{aligned}$$

$\therefore R_Y(\tau)$ = Inverse Fourier transform of $S_Y(\omega)$

$$= \frac{1}{2\pi} \int_{-B}^B \frac{N_0}{2} e^{j\omega\tau} d\omega = \frac{N_0}{2} \frac{\sin B\tau}{\pi\tau}$$

The output PSD $S_Y(\omega)$ and the output autocorrelation function $R_Y(\tau)$ are illustrated in Fig. below.



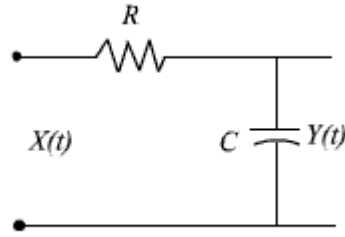
Example 2:

A random voltage modeled by a white noise process $X(t)$ with power spectral density $\frac{N_0}{2}$ is input to an RC network shown in the fig.

Find (a) output PSD $S_Y(\omega)$

(b) output auto correlation function $R_Y(\tau)$

(c) average output power $EY^2(t)R$



The frequency response of the system is given by

$$\begin{aligned} H(\omega) &= \frac{1}{R + \frac{1}{jC\omega}} \\ &= \frac{1}{jRC\omega + 1} \end{aligned}$$

Therefore,

(a)

$$\begin{aligned} S_Y(\omega) &= |H(\omega)|^2 S_X(\omega) \\ &= \frac{1}{R^2 C^2 \omega^2 + 1} S_X(\omega) \\ &= \frac{1}{R^2 C^2 \omega^2 + 1} \frac{N_0}{2} \end{aligned}$$

(b) Taking inverse Fourier transform

$$R_Y(\tau) = \frac{N_0}{4RC} e^{-\frac{|\tau|}{RC}}$$

(c) Average output power

$$EY^2(t) = R_Y(0) = \frac{N_0}{4RC}$$

PROBABILITY THEORY AND STOCHASTIC PROCESSES

Unit wise Important Questions

UNIT-I

1. State and Prove Bayes' theorem.
2. Differentiate Joint and Conditional Probability
3. Show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. Define Probability axioms and properties.
5. Define probability based on Relative frequency and classical definition
6. Define equally likely events, Exhaustive events and mutually exclusive events and sample space.
7. Define Random variable and explain the classifications of Random variable with example.
8. Explain about conditional probability and its properties (or) show that conditional probability satisfies the axioms of probability.
9. State and prove the total probability theorem?
10. Define independent events and Show that if A and B are independent, then so the following also independent.
a) A and $\bar{B}$ b) $\bar{A}$ and B c) $\bar{A}$ and $\bar{B}$

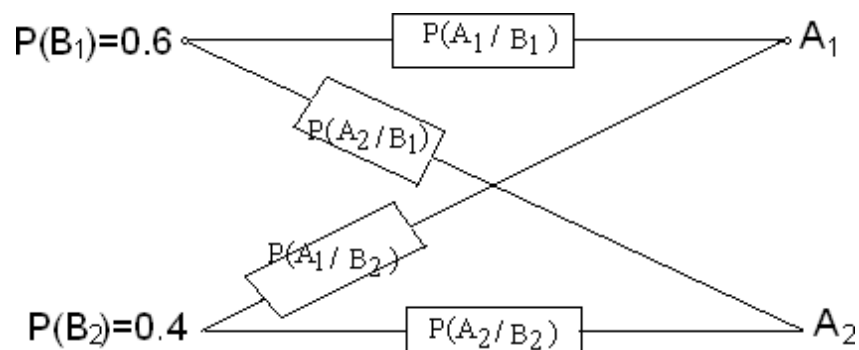
Problems:

1. In a box there are 100 resistors having resistance and tolerance values given in table. Let a resistor be selected from the box and assume that each resistor has the same likelihood of being chosen. Event A: Draw a 47Ω resistor, Event B: Draw a resistor with 5% tolerance, Event C: Draw a 100Ω resistor. Find the individual, joint and conditional probabilities.

Resistance (Ω)	Tolerance		Total
	5%	10%	
22	10	14	24
47	28	16	44
100	24	8	32
Total	62	38	100

2. Two boxes are selected randomly. The first box contains 2 white balls and 3 black balls. The second box contains 3 white and 4 black balls. What is the probability of drawing a white ball.
b) An aircraft is used to fire at a target. It will be successful if 2 or more bombs hit the target. If the aircraft fires 3 bombs and the probability of the bomb hitting the target is 0.4, then what is the probability that the target is hit?
3. An experiment consists of observing the sum of the outcomes when two fair dice are thrown. Find the probability that the sum is 7 and find the probability that the sum is greater than 10.

4. In a factory there are 4 machines produce 10%,20%,30%,40% of an items respectively. The defective items produced by each machine are 5%,4%,3% and 2% respectively. Now an item is selected which is to be defective, what is the probability it being from the 2nd machine. And also write the statement of total probability theorem?
5. Determine probabilities of system error and correct system transmission of symbols for an elementary binary communication system shown in below figure consisting of a transmitter that sends one of two possible symbols (a 1 or a 0) over a channel to a receiver. The channel occasionally causes errors to occur so that a '1' show up at the receiver as a '0' and vice versa. Assume the symbols '1' and '0' are selected for a transmission as 0.6 and 0.4 respectively.



6. If $P(A)=0.4$, $P(B)=0.7$ and $P(A \cap B)=0.3$ find $P(A' \cap B')$.
7. A pack contains 6 white and 4 green pencils; another pack contains 5 white and 7 green pencils. If one pencil is drawn from each pack, find the probability that i) both are white
ii) One is white and another is green.
8. Two fair dice are thrown independently. Three events A,B,C are respectively defined as follows:
i) Odd face with first die ii) Odd face with second die iii) The sum of two numbers in the two dice is odd. Are the events A, B, C mutually independent or pair wise independent

UNIT-II

1. Define probability Distribution function and state and prove its properties.
2. Define probability density function and state and prove its properties.
3. Derive the Binomial density function and find mean & variance.
4. Write short notes on Gaussian distribution and also find its mean and variance.
5. Find the variance and Moment generating function of exponential distribution?
6. Find the mean, variance and Moment generating function of uniform distribution?
7. Define Moment Generating Function and state and prove its properties.
8. Define variance and state and prove its properties.
9. Define and State the properties of Conditional Distribution and Conditional Density functions.
10. Explain the moments of the random variable in detail.

Problems:

1. Let X be a continuous random variable with pdf $f(x) = \begin{cases} 3x, & 0 < x < 1 \\ 0 & \text{else} \end{cases}$ find $P(X \leq 0.6)$?
2. a) If X is uniformly distributed in $[-2, 2]$, find (i) $P(x < 0)$ (ii) $P(|X - 1| \geq \frac{1}{2})$
 b) If X is uniformly distributed with $E(x) = 1$ and $\text{Var}(x) = 4/3$, find $P(X < 0)$
 c) If X is uniformly distributed over $(0, 10)$ calculate the probability that $3 < x < 8$.
3. A random variable X has the following probability density function

$$f_X(x) = \begin{cases} x & 0 < x < 1 \\ k(2-x) & 1 \leq x \leq 2 \\ 0 & \text{else} \end{cases}$$
 - i) Find the value of k?
 - ii) Find $P(0.2 < x < 1.2)$?
 - iii) Find the distribution function of X.
4. Consider that a fair coin is tossed 3 times, Let X be a random variable, defined as X = number of tails appeared, find the expected value of X.
5. If the probability density of X is given by $f(x) = \begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & \text{else} \end{cases}$
 Find its r^{th} moment and evaluate $E[(2X+1)^2]$
6. If the random variable X has the following probability distribution

X:	-2	-1	0	1
P(x):	0.4	K	0.2	0.3

 Find k and the mean and variance of x.
7. If X is a discrete random variable with a Moment generating function of $M_X(v)$, find the Moment generating function of
 - i) $Y = aX + b$
 - ii) $Y = KX$
 - iii) $Y = \frac{X+a}{b}$
8. A random variable X has the distribution function

$$F_X(x) = \sum_{n=1}^{12} \frac{n^2}{650} u(x - n)$$
 Find the probability of a) $P\{-\infty < X \leq 6.5\}$ b) $p\{X > 4\}$ c) $p\{6 < X \leq 9\}$

UNIT-III

1. State and prove the density function of sum of two random variables.
2. State and explain the properties of joint density function.
3. Define and State the properties of joint cumulative distribution function of two random variables X and Y.
4. Define and state and prove the properties of Joint Characteristic Function of two random variables.
5. Explain jointly Gaussian random variables and also state its properties.
6. Explain the joint moments of two random variables in detail.
7. Define covariance and derive its properties.
8. Explain the statistical independence of random variables with example.

Problems:

1. The joint density function of two random variables X and Y is

$$f_{XY}(x,y) = \begin{cases} \frac{(x+y)^2}{40} & ; -1 < x < 1 \text{ and } -3 < y < 3 \\ 0; & \text{otherwise} \end{cases}$$

Find the variances of X and Y.

2. The joint probability density function of X&Y is $f_{X,Y}(x,y) = c(2x+y); 0 \leq x \leq 2, 0 \leq y \leq 3$
0; else

Then find the value of constant c and also find joint distribution function.

3. Let $Z=X+Y-C$, where X and Y are independent random variables with variance σ^2_X , σ^2_Y and C is constant. Find the variance of Z in terms of σ^2_X , σ^2_Y and C.
4. If X is a random variable for which $E(X) = 10$ and $\text{var}(X) = 25$, for what possible value of 'a' and 'b' does $Y=aX-b$ have expectation 0 and variance 1 ?

5. The joint density function of random variables X and Y is

$$f_{X,Y}(x,y) = 8xy, \quad 0 \leq x < 1, 0 \leq y < 1$$

Find $f(y/x)$ and $f(x/y)$

6. The input to a binary communication system is a random variable X, takes on one of two values 0 and 1, with probabilities $\frac{3}{4}$ and $\frac{1}{4}$ respectively. Due to the errors caused by the channel noise, the output random variable Y, differs from the Input X occasionally. The behavior of the communication system is modeled by the conditional probabilities as

$$P\left\{\frac{(Y=1)}{(X=1)}\right\} = \frac{3}{4} \text{ and } P\left\{\frac{(Y=0)}{(X=0)}\right\} = \frac{7}{8}$$

Find a) the probability for a transmitted message to be received as 0.

b) the probability that transmitted message is a 1 if the received is 1.

7. Let X and Y be the random variables defined as $X=\cos\theta$ and $Y=\sin\theta$ where θ is a uniform random variable over $(0, 2\pi)$
 - a. Are X and Y Uncorrelated?
 - b. Are X and Y Independent?

8. Two random variables X and Y have zero mean and variance $\sigma_X^2 = 16$ and $\sigma_Y^2 = 36$ correlation coefficient is 0.5 determine the following

- i) The variance of the sum of X and Y
- ii) The variance of the difference of X and Y

9. Two random variables X and Y have the joint pdf is

$$f_{x,y}(x,y) = \begin{cases} Ae^{-(2x+y)} & x, y \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

- i. Evaluate A
- ii. Find the marginal pdf's
- iii. Find the marginal pdf's
- iv. Find the joint cdf
- v. Find the distribution functions and conditional cdf's.

10. a) Given is the joint distribution of X and Y

		X		
		0	1	2
Y	0	0.02	0.08	0.10
	1	0.05	0.20	0.25
	2	0.03	0.12	0.15

- Obtain (i) Marginal Distributions and (ii) The Conditional Distribution of X given Y=0.
- b) If X and Y are uncorrelated random variables with variances 16 and 9. Find the Correlation coefficient between X+Y and X-Y.

UNIT-IV

1. Define Stationary random Process and explain various levels of stationary random processes.
2. State and prove the properties of auto correlation function.
3. Briefly explain the distribution and density functions in the context of stationary and independent random processes.
4. Differentiate random variable and random process with example.
5. State and prove the properties of cross correlation function.
6. Explain types of random process in detail.
7. Define covariance of the random process and derive its properties.
8. Find mean square value and auto correlation function of response of LTI system.
9. Explain Ergodic random processes in detail.
10. If the input to the LTI system is the WSS random process then show that output is also a WSS random process.

Problems:

1. A random process is given as $X(t) = At$, where A is a uniformly distributed random variable on $(0,2)$. Find whether $X(t)$ is wide sense stationary or not.
2. $X(t)$ is a stationary random process with a mean of 3 and an auto correlation function of $R_{XX}(\tau) = \exp(-0.2|\tau|)$. Find the second central Moment of the random variable $Y=Z-W$, where 'Z' and 'W' are the samples of the random process at $t=4$ sec and $t=8$ sec respectively.
3. Given the RP $X(t) = A \cos(\omega_0 t) + B \sin(\omega_0 t)$ where ω_0 is a constant, and A and B are uncorrelated Zero mean random variables having different density functions but the same variance σ^2 . Show that $X(t)$ is wide sense stationary.
4. The function of time $Z(t) = X_1 \cos \omega_0 t - X_2 \sin \omega_0 t$ is a random process. If X_1 and X_2 are independent Gaussian random variables, each with zero mean and variance σ^2 , find $E[Z]$, $E[Z^2]$ and $\text{var}(z)$.
5. Examine whether the random process $X(t) = A \cos(\omega t + \theta)$ is a wide sense stationary if A and ω are constants and θ is uniformly distributed random variable in $(0, 2\pi)$

6. A random process $X(t)$ defined by $X(t) = A \cos(t) + B \sin(t)$, where A and B are independent random variables each of which takes a value 2 with Probability $1/3$ and a value 1 with probability $2/3$. Show that $X(t)$ is wide-sense stationary

7. A random process has sample functions of the form $X(t) = A \cos(\omega t + \theta)$ where ω is constant, A is a random variable with mean zero and variance one and θ is a random variable that is uniformly distributed between 0 and 2π . Assume that the random variables A and θ are independent. Is $X(t)$ mean ergodic-process?

8. Find the variance of the stationary process $\{X(t)\}$ whose ACF is given by

$$R_{XX}(\tau) = 16 + \frac{9}{1+6\tau^2}$$

UNIT-V

1. State and prove the properties of power spectral density
2. Derive the relation between input PSD and output PSD of an LTI system
3. Derive the relation between PSD and auto correlation function of the random processes.
4. State and prove the properties of cross power spectral density
5. Derive the expression for cross power spectral density between input and output of LTI system.
6. Write short notes on cross power density spectrum.

Problems:

1. Check the following power spectral density functions are valid or not

i) $\frac{\cos 8(\omega)}{2 + \omega^4}$ ii) $e^{-(\omega-1)^2}$

2. A stationary random process $X(t)$ has spectral density $S_{XX}(\omega) = 25/(\omega^2 + 25)$ and an independent stationary process $Y(t)$ has the spectral density $S_{YY}(\omega) = \omega^2/(\omega^2 + 25)$. If $X(t)$ and $Y(t)$ are of zero mean, find the:

- a) PSD of $Z(t) = X(t) + Y(t)$
- b) Cross spectral density of $X(t)$ and $Z(t)$

3. The input to an LTI system with impulse response $h(t) = \delta(t) + t^2 e^{-at}$. $U(t)$ is a WSS process with mean of 3. Find the mean of the output of the system.

4. A random process $Y(t)$ has the power spectral density $S_{YY}(\omega) = \frac{9}{\omega^2 + 64}$

Find i) The average power of the process ii) The Auto correlation function

5. a) A random process has the power density spectrum $S_{YY}(\omega) = \frac{6\omega^2}{1 + \omega^4}$. Find the average power in the process.

6. Find the cross correlation function corresponding to the cross power spectrum

$$S_{XY}(\omega) = \frac{6}{(9 + \omega^2)(3 + j\omega)^2}$$

7. Consider a random process $X(t) = \cos(\omega t + \theta)$ where ω is a real constant and θ is a uniform random variable in $(0, \pi/2)$. Find the average power in the process.

8. $X(t)$ and $Y(t)$ are zero mean and stochastically independent random processes having autocorrelation functions $R_{XX}(\tau) = e^{-|\tau|}$ and $R_{YY}(\tau) = \cos 2\pi\tau$ respectively.

- a) Find autocorrelation functions of $W(t) = X(t) + Y(t)$ and $Z(t) = X(t) - Y(t)$.
- b) Find cross correlation function between $W(t)$ and $Z(t)$

9. A random process is defined as $Y(t) = X(t) - X(t-a)$, where $X(t)$ is a WSS process and $a > 0$ is a constant. Find the PSD of $Y(t)$ in terms of the corresponding quantities of $X(t)$