PROBABILITY THEORY AND STOCHASTIC PROCESSES

UNIT I:

Probability and Random Variable

Probability: Set theory, Experiments and Sample Spaces, Discrete and Continuous Sample Spaces, Events, Probability Definitions and Axioms, Mathematical Model of Experiments, 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, Mixed Random Variable

UNIT II:

Distribution and density functions and Operations on One Random Variable

Distribution and density functions: Distribution and Density functions, Properties, Binomial, Poisson, Uniform, Exponential Gaussian, Rayleigh and Conditional Distribution, Methods of defining Conditioning Event, 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, transformations of a random variable, monotonic transformations for a continuous random variable, non monotonic transformations of continuous random variable, transformations of Discrete random variable

UNIT III:

Multiple Random Variables and Operations on Multiple Random Variables

Multiple Random Variables: Vector 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 and Sum of Several Random Variables, Central Limit Theorem - Unequal Distribution, Equal Distributions

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 and N Random Variable case, Properties, Transformations of Multiple Random Variables

UNIT VI:

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, Gaussian Random Processes.

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, Clave 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 – 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.

A brief history

Probability has an amazing history. A practical gambling problem faced by the French nobleman *Chevalier de Méré* sparked the idea of probability in the mind of *Blaise Pascal (1623-1662)*, the famous French mathematician. Pascal's correspondence with Pierre de Fermat (1601-1665), another French Mathematician in the form of seven letters in 1654 is regarded as the genesis of probability. Early mathematicians like Jacob Bernoulli (1654-1705), Abraham de Moivre (1667-1754), Thomas Bayes (1702-1761) and Pierre Simon De Laplace (1749-1827) contributed to the development of probability. Laplace's *Theory Analytique des Probabilities* gave comprehensive tools to calculate probabilities based on the principles of permutations and combinations. Laplace also said, *"Probability theory is nothing but common sense reduced to calculation."*

Later mathematicians like Chebyshev (1821-1894), Markov (1856-1922), von Mises (1883-1953), Norbert Wiener (1894-1964) and Kolmogorov (1903-1987) contributed to new developments. Over the last four centuries and a half, probability has grown to be one of the most essential mathematical tools applied in diverse fields like economics, commerce, physical sciences, biological sciences and engineering. It is particularly important for solving practical electrical-engineering problems in *communication*, *signal processing* and *computers*. Notwithstanding the above developments, a precise definition of probability eluded the mathematicians for centuries. Kolmogorov in 1933 gave the *axiomatic definition of probability* and resolved the problem.

Randomness arises because of

- random nature of the generation mechanism
- Limited understanding of the signal dynamics inherent imprecision in measurement, observation, etc.

For example, *thermal noise* appearing in an electronic device is generated due to random motion of electrons. We have deterministic model for weather prediction; it takes into account of the factors affecting weather. We can locally predict the temperature or the rainfall of a place on the basis of previous data. Probabilistic models are established from observation of a random phenomenon. While *probability* is concerned with analysis of a random phenomenon, *statistics* help in building such models from data.

Deterministic versus probabilistic models

A *deterministic model* can be used for a physical quantity and the process generating it provided sufficient information is available about the initial state and the dynamics of the process generating the physical quantity. For example,

- We can determine the position of a particle moving under a constant force if we know the initial position of the particle and the magnitude and the direction of the force.

- We can determine the current in a circuit consisting of resistance, inductance and capacitance for a known voltage source applying Kirchoff's laws.

Many of the physical quantities are *random* in the sense that these quantities cannot be predicted with *certainty* and can be described in terms of *probabilistic models* only. For example,

- The outcome of the tossing of a coin cannot be predicted with certainty. Thus the outcome of tossing a coin is random.

- The number of ones and zeros in a packet of binary data arriving through a communication channel cannot be precisely predicted is random.

- The ubiquitous *noise* corrupting the signal during acquisition, storage and transmission can be modelled only through statistical analysis.

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 BASIC CONCEPTS OF SET THEORY

Some of the basic concepts of set theory are:

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. We denote the phrase _belongs to‘ by the Greek symbol _∈_ (epsilon). Thus, we say that **a ∈ A**.

If _b‘ is an element which does not belong to A, we represent this as b ∉ 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,......}

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

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

The different types of sets are explained below with examples.

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 ∅ and is read as phi. In roster form, ∅ 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 ∈ N, 3 < x < 4}

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

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

• Let B = {x : x is a composite number less than 4}.

Here B is an empty set because there is no composite number less than 4.

Note:

∅ ≠ {0} ∴ has no element.

{0} is a set which has one element 0.

The cardinal number of an empty set, i.e., n(∅) = 0

2. **Singleton Set:**

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

For example:

• A = {x : x is neither prime nor composite}

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

• B = {x : x is a whole number, x < 1}

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

• Let A = {x : x ∈ N 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 is a 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 ∈ N, x < 7}

• P = {2, 3, 5, 7, 11, 13, 17, ... 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 ∈ N, x > 1}

• Set of all prime numbers

• B = {x : x ∈ W, x = 2n}

Note:

All infinite sets cannot be expressed in roster form.

For example:

The set of real numbers since the elements of this set do not follow any particular pattern.

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 ∈ N, x < 5}

A = {1, 2, 3, 4}

Therefore, n(A) = 4

• B = set of letters in the word ALGEBRA

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

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 ‗↔‘.

For example:

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

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

Therefore, A ↔ 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}

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 is a prime number}

B = {x : x 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;

• A = {a, b, c, d}

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

• X = {x : x ∈ N, x < 4}

Y = {x : x ∈ 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 ⊆ B** or **B ⊇ A**

The symbol **⊂** stands for _is a subset of‟ or _is contained in‘

• Every set is a subset of itself, i.e., A ⊂ A, B ⊂ B.

• Empty set is a subset of every set.

• Symbol _⊆‘ is used to denote _is a subset of‟ or _is contained in‘.

• A ⊆ B means A is a subset of B or A is contained in B.

• B ⊆ 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.

Notes:

If ACB and BCA, then A = B, i.e., they are equal sets.

Every set is a subset of itself.

Null set or ∅ is a subset of every set.

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

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

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

Here A ⊂ B and B ⊂ A.

Hence, we can say A = B

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

B = {4, 5, 6, 7}

Here A ⊄ B and also B ⊄ C

[⊄ denotes _not a subset of‟]

11. Super Set:

Whenever a set A is a subset of set B, we say the B is a superset of A and we write, B ⊇ A.

Symbol ⊇ is used to denote _is a super set of‟

For example;

A = {a, e, i, o, u}

B = {a, b, c,............., z}

Here A ⊆ B i.e., A is a subset of B but B ⊇ A i.e., B is a super set of A

12. Proper Subset:

If A and B are two sets, then A is called the proper subset of B if A ⊆ B but B ⊇ A i.e., A ≠ B. The symbol ‗⊂' is used to denote proper subset. Symbolically, we write A ⊂ 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 ⊂ B

Notes:

No set is a proper subset of itself.

Null set or ∅ 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 ≠ B.

Notes:

No set is a proper subset of itself.

Empty set is a proper subset of every set.

13. Power Set:

The collection of all subsets of set A is called the power set of A. It is denoted by P(A). In P(A), every element is a set.

For example;

If A = {p, q} then all the subsets of A will be

P(A) = {∅, {p}, {q}, {p, q}}

Number of elements of P(A) = n[P(A)] = 4 = 22

In general, n[P(A)] = 2m where m is the number of elements in set A.

14.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 ∪ 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 ⊆ U, B ⊆ U, C ⊆ U and U ⊇ A, U ⊇ B, U ⊇ 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 smallest 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 ‗∪‘.

Some properties of the operation of union:

(i) A∪B = B∪A **(Commutative law)**

(ii) A∪(B∪C) = (A∪B)∪C **(Associative law)**

(iii) A ∪ Φ = A **(Law of identity element, is the identity of ∪)**

(iv) A∪A = A **(Idempotent law)**

(v) U∪A = U **(Law of ∪)** ∪ is the universal set.

Notes:

A ∪ Φ = Φ∪ 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 ∪ **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** = {ϕ}. Find union of two given sets X and Y.

Solution:

X ∪ 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 largest 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 ‗∩‗.

Some properties of the operation of intersection

(i) $A \cap B = B \cap A$ (Commutative law)

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

(iii) $\Phi \cap A = \Phi$ (Law of Φ)

(iv) $U \cap A = A$ (Law of ∪)

(v) $A \cap A = A$ (Idempotent law)

(vi) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributive law) Here ∩ distributes over ∪

Also, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (Distributive law) Here ∪ distributes over ∩

Notes:

$A \cap \Phi = \Phi \cap A = \Phi$ 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 ∩ 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** = {ϕ}. Find intersection of two given sets X and Y.

Solution:

X ∩ 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}

In general, $B - A = \{x : x \in B, \text{ and } x \notin A\}$

• 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‘.

Some properties of complement sets

(i) $A \cup A' = A' \cup A = \cup$ (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) $\Phi' = \cup$ (Law of empty set - The complement of an empty set is a universal set.

(vii) $\cup' = \Phi$ 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 U B = B U A

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

2. Associative Laws:

For any three finite sets A, B and C;

(i) (A U B) U C = A U (B U 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 U A = A

(ii) A ∩ A = A

4. Distributive Laws:

For any three finite sets A, B and C;

(i) A U (B ∩ C) = (A U B) ∩ (A U C)

(ii) A ∩ (B U C) = (A ∩ B) U (A ∩ 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 U C) = (A – B) ∩ (A – C)

(ii) A - (B ∩ C) = (A – B) U (A – C)

De Morgan's Laws can also we written as:

(i) (A U B)' = A' ∩ B'

(ii) (A ∩ B)' = A' U B'

More laws of algebra of sets:

6. For any two finite sets A and B;

(i) A – B = A ∩ B'

(ii) B – A = B ∩ A'

(iii) A – B = A ⇔ A ∩ B = ∅

(iv) (A – B) U B = A U B

(v) (A – B) ∩ B = ∅

(vi) (A – B) U (B – A) = (A U B) – (A ∩ 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 U B)' = A' ∩ B' (which is a De Morgan's law of union).

(ii) (A ∩ B)' = A' U B' (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.

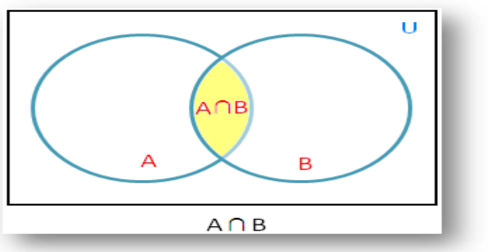
In this,

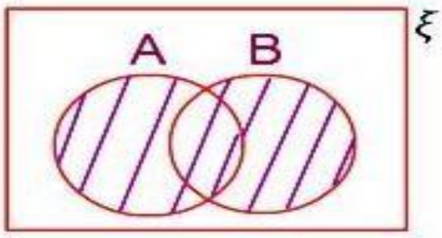
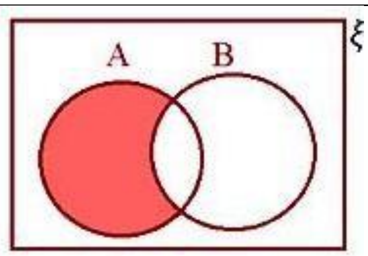
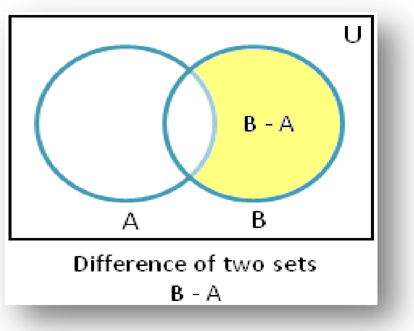
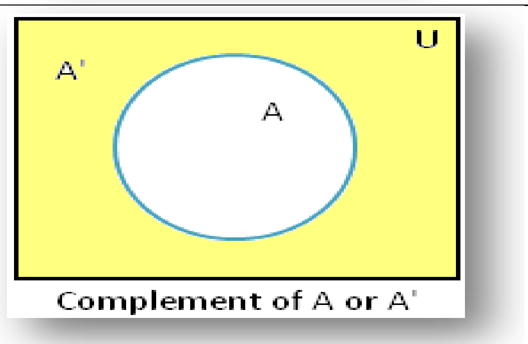
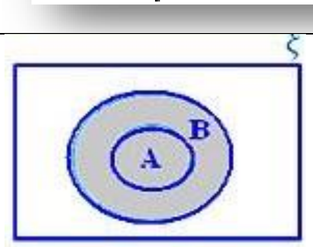
•A rectangle is used to represent a universal set.

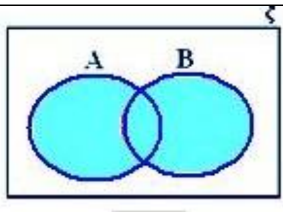
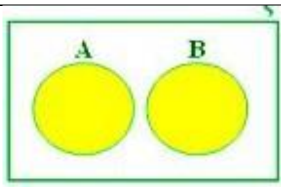
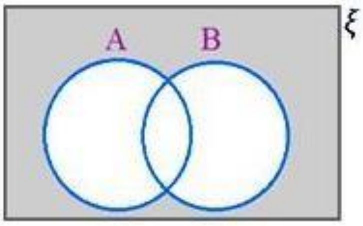
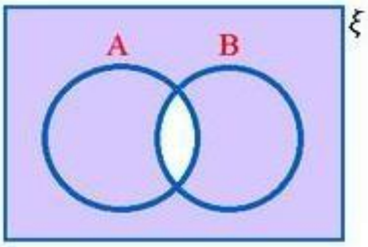
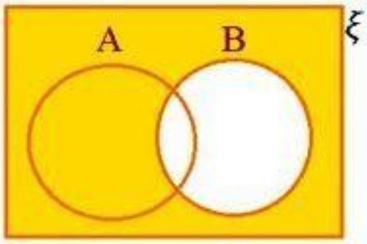
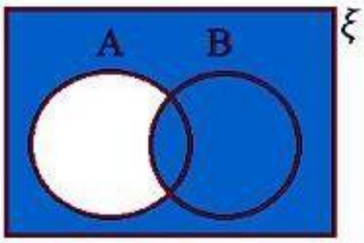
• Circles or ovals are used to represent other subsets of the universal set.

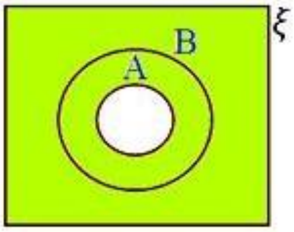
Venn diagrams in different situations

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 ∪ B when A ⊂ B	

7	A ∪ B when neither A ⊂ B nor B ⊂ A	
8	A ∪ B when A and B are disjoint sets	
9	**(A ∪ B)’** (A union B dash)	
10	**(A ∩ B)’** (A intersection B dash)	
11	**B’** (B dash)	
12	**(A - B)’** (Dash of sets A minus B)	

13	**(A ⊂ 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 ∪ B) = 36, find n(A ∩ B).

Solution:

Using the formula n(A ∪ B) = n(A) + n(B) - n(A ∩ B).

then n(A ∩ B) = n(A) + n(B) - n(A ∪ B)

= 20 + 28 - 36

= 48 - 36

= 12

2. If n(A - B) = 18, n(A ∪ B) = 70 and n(A ∩ B) = 25, then find n(B).

Solution:

Using the formula n(A∪B) = n(A - B) + n(A ∩ 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 ∩ 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

$(A \cup B) = 60$ $n(A) = 27$ $n(B) = 42$ 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:

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

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

(i) 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,

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

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

and n(B - A) = n(B) - n(A ∩ 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 **p**redicted 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 = {odd, 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, T H , 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 count ably 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={Heads} and B={Tails} would be mutually exclusive.

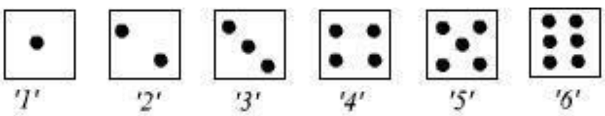
An event consisting of a single point of the sample space 'S' is called a simple event or elementary event.

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 ϕ.

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

$$A = \text{The event of getting an odd face} = \{'1', \ '3', \ '5'\}.$$
$$B = \text{The event of getting a six} = \{'6'\}$$

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 $S = \{H, \ TH, TTH, ...\}$.

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 occuring.
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 w i t h 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:

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

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

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

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 $A = \{ '6' \}$

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

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

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*)≥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

$$0 \le P(E_i) \le 1$$

Axiom2: The probability of the sure event is 1. Namely *P*(Ω)=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*∪*B*∪*C*)=*P*(*A*)+*P*(*B*)+*P*(*C*). i.e Additivity:

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

$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$ for $n = 1, 2, \ldots, N$ where $E_1, E_2, \ldots$ 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. $P(A)+P(\bar{A})=1$. Or $P(\bar{A})=1-P(A)$
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(Ø)=0
4. If $A \subset B$, then $P(A) \le P(B)$.
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:

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

Rules of Probability:

Rule of Subtraction:

Rule of Subtraction The probability that event A will occur is equal to 1 minus the probability that event A will not occur.

$$P(A) = 1 - P(A')$$

Rule of Multiplication:

Rule of Multiplication The probability that Events A and B both occur is equal to the probability that Event A occurs times the probability that Event B occurs, given that A has occurred.

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

Rule of addition:

Rule of Addition The probability that Event A or Event B occurs is equal to the probability that Event A occurs plus the probability that Event B occurs minus the probability that both Events A and B occur.

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

Note: Invoking the fact that $P(A \cap B) = P(A)P(B | A)$, the Addition Rule can also be expressed as

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

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	$nP_r = r!/nC_r = n!/(n-r)!$	n C r = P (n , r) / r ! = n ! / r ! (n - r) !

Example 1: An urn contains 6 red balls, 5 green balls and 4 blue balls. 9 balls were picked at random from the urn without replacement. What is the probability that out of the balls 4 are red, 3 are green and 2 are blue?

Sol:

9 balls can be picked from a population of 15 balls in $^{15}C_9 = \frac{15!}{9!6!}$.

Therefore the required probability is $\frac{^6C_4 \times {}^5C_3 \times {}^4C_2}{^{15}C_9}$

Example2: What is the probability that in a throw of 12 dice each face occurs twice.

Solution: The total number of elements in the sample space of the outcomes of a single throw of 12 dice is $= 6^{12}$

The number of favourable outcomes is the number of ways in which 12 dice can be arranged in six groups of size 2 each – group 1 consisting of two dice each showing 1, group 2

consisting of two dice each showing 2 and so on.
Therefore, the total number distinct groups

$$= \frac{12!}{2!2!2!2!2!2!}$$

Hence the required probability is $= \frac{12!}{(2)^6 6^{12}}$

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∩B

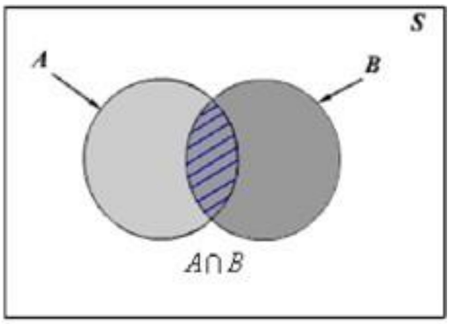


Figure 1

$$P(B/A) = \frac{\text{Number of outcomes favourable to A and B}}{\text{Number of outcomes in A}}$$

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

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

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

Problems:

Example 1 Consider the example tossing the fair die. Suppose

$$A = \text{event of getting an even number} = \{2, 4, 6\}$$
$$B = \text{event of getting a number less than } 4 = \{1, 2, 3\}$$
$$\therefore A \cap B = \{2\}$$
$$\therefore P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/6}{3/6} = \frac{1}{3}$$

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 girls

$$S = \{gg,\ gb,\ bg,\ bb\},\ A = \{gg,\ gb,\ bg\} \text{ and } B = \{gg\}$$
$$A \cap B = \{gg\}$$
$$\therefore P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

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:

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

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*

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

Joint probability

Joint probability is defined as the probability of both A and B taking place, and is denoted by P (AB) or P(A∩B)

Joint probability is not the same as conditional probability, though the two concepts are often confused. Conditional probability assumes that one event has taken place or will take place, and then asks for the probability of the other (A, given B). Joint probability does not have such conditions; it simply asks for the chances of both happening (A and B). In a problem, to help distinguish between the two, look for qualifiers that one event is conditional on the other (conditional) or whether they will happen concurrently (joint).

Probability definitions can find their way into CFA exam questions. Naturally, there may also be questions that test the ability to calculate joint probabilities. Such computations require use of the multiplication rule, which states that the joint probability of A and B is the product of the conditional probability of A given B, times the probability of B. In probability notation:

$$P(AB) = P(A \mid B) * P(B)$$

Given a conditional probability P(A | B) = 40%, and a probability of B = 60%, the joint probability P(AB) = 0.6*0.4 or 24%, found by applying the multiplication rule.

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

For independent events: P(AB) = P(A) * P(B)

Moreover, the rule generalizes for more than two events provided they are all independent of one another, so the joint probability of three events P(ABC) = P(A) * (P(B) *P(C), again assuming independence.

Summary of probabilities

Event	Probability
A	$P(A) \in [0,1]$
not A	$P(A^c) = 1 - P(A)$
A or B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $P(A \cup B) = P(A) + P(B)$ **if A and B are mutually exclusive**
A and B	$P(A \cap B) = P(A\|B)P(B) = P(B\|A)P(A)$ $P(A \cap B) = P(A)P(B)$ **if A and B are independent**
A given B	$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B\|A)P(A)}{P(B)}$

Total Probability theorem:

Let $A_1, A_2, \ldots, A_n$ be n events such that $S = A_1 \cup A_2 \ldots \cup A_n$ and $A_i \cap A_j = \phi$ for $i \neq j$. Then for any event B,

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B / A_i)$$

Proof: We have

$$\therefore P(B) = P(\bigcup_{i=1}^{n} B \cap A_i)$$

$$= \sum_{i=1}^{n} P(B \cap A_i)$$

$$= \sum_{i=1}^{n} P(A_i) P(B / A_i)$$

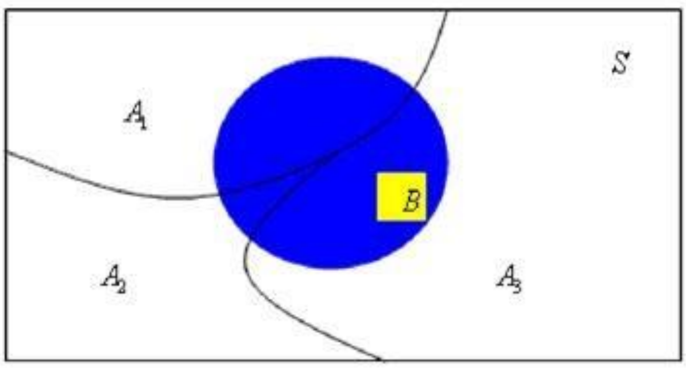


Figure 3

Remark

(1) A decomposition of a set S into 2 or more disjoint nonempty subsets is called *a partition* of S.The subsets $A_1, A_2, \ldots, A_n$ form a partition of S if $S = A_1 \cup A_2 \ldots \cup A_n$ and $A_i \cap A_j = \phi$ for i ≠ j.

(2) The theorem of total probability can be used to determine the probability of a complex event in terms of related simpler events. This result will be used in Bays' theorem to be discussed to the end of the lecture**.**

Bayes' Theorem:

Suppose $A_1, A_2, \ldots A_n$ *are partitions on S such that* $S = A_1 \cup A_2 \ldots\ldots \cup A_n$ and $A_i \cap A_j = \phi$ for $i \neq j$.

Suppose the event B occurs if one of the events $A_1, A_2, \ldots A_n$ occurs. Thus we have the information of the probabilities $P(A_i)$ and $P(B/A_i), i = 1, 2.., n.$ We ask the following question:

Given that B has occured what is the probability that a particular event A_k *has occured*? In other words *what is* $P(A_k/B)$?

We have $P(B) = \sum_{i=1}^{n} P(A_i)\,P(B|A_i)$ (Using the theorem of total probability)

$$\therefore \; P(A_k|B) = \frac{P(A_k)\,P(B/A_k)}{P(B)}$$

$$= \frac{P(A_k)\,P(B/A_k)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}$$

This result is known as the Baye's theorem. The probability $P(A_k)$ is called the *a priori* probability and $P(A_k/B)$ is called the *a posteriori* probability. Thus the Bays' theorem enables us to determine the *a posteriori* probability $P(A_k/B)$ from the observation that *B* has occurred. This result is of practical importance and is the heart of Baysean classification, Baysean estimation etc.

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 $P(R_0 / T_1) = 0.08.$

(i) $P(R_1)$ = Probabilty of receiving 'one'

$$\begin{aligned} &= P(T_1)P(R_1/T_1) + P(T_0)P(R_1/T_0) \\ &= 0.4\times 0.92 + 0.6\times 0.1 \\ &= 0.448 \end{aligned}$$

(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 7: 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 $P(A_3/D)$.

$$P(A_1) = \frac{2}{9},\ P(A_2) = \frac{1}{3} \text{ and } P(A_3) = \frac{4}{9}$$

$$P(D/A_1) = 0.01,\ P(D/A_2) = 0.015 \text{ and } P(D/A_3) = 0.02$$

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

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 $A_1, A_2, \ldots, A_n$ are called independent if and only if

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

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:

Example1.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.

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

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).

Therefore P (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

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

:

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(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 = { 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(club) = $\frac{1}{4}$.

Solution 2: Let S = {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(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

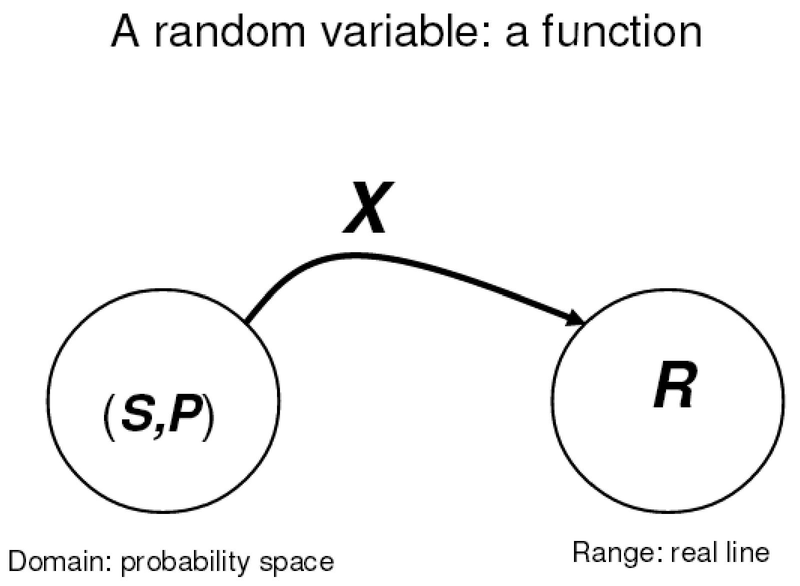
In many situations, we are interested in numbers associated with the outcomes of a random experiment. In application of probabilities, we are often concerned with numerical values which are random in nature. For example, we may consider the number of customers arriving at a service station at a particular 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.)



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 randomvariablesX1, X2 as
1 X1- the number of broken eggs in the 10 boxes
2 X2- the number of boxes rejected
Then the range of X1 is {0, 1,2,3,4-------------- 60} and X2 is {0,1,2------- 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 $X = \{1,2,3,4,5,6\}$.

Types of random variables:

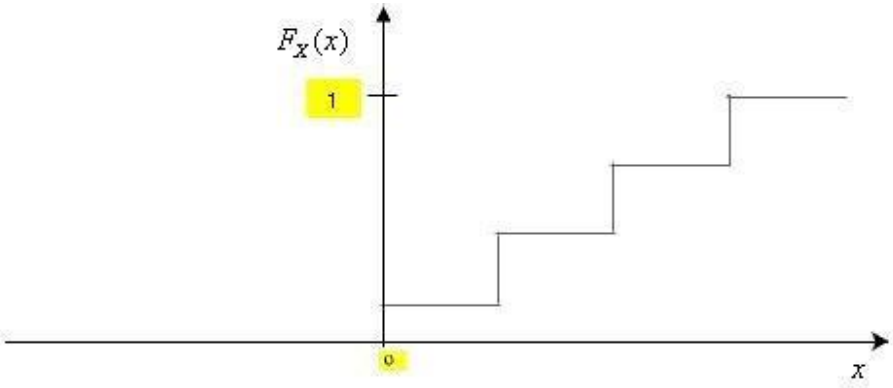
There are two types of random variables, ***discrete*** and ***continuous***.

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.



Plot of Distribution function of *discrete random variable*

•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 – 9:30AM Mon--Fri.
–E.g. Binomial, Poisson...

2. ***Continuous random variable:***

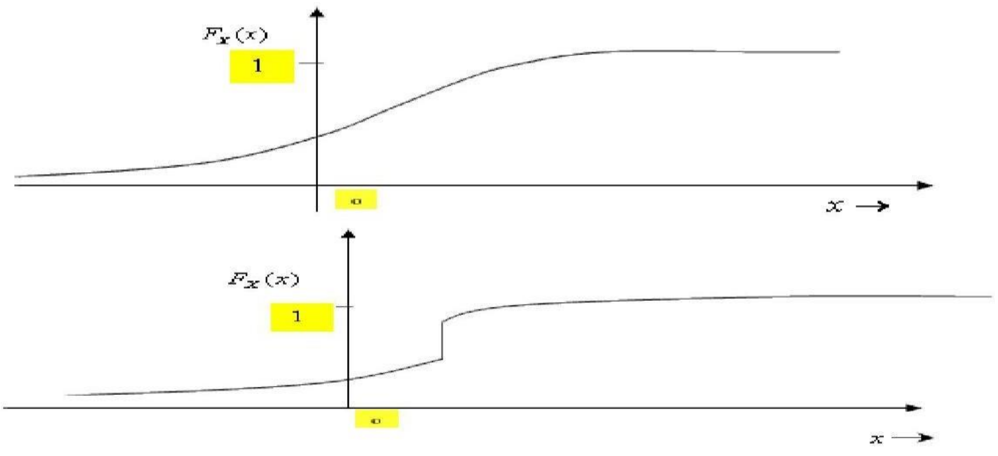
A ***continuous random variable*** is one which takes an infinite number of possible values. Continuous rando variables are usually measurements. E
A continuous random variable takes all values in an interval of real numbers.

(or)

X is called a ***continuous random variable*** if $F_X(x)$ ı 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.



Plot of Distribution function of *continuous and mixed random variables*

UNIT-II
DISTRIBUTION AND DENSITY FUNCTIONS AND OPERATIONS ON ONE RANDOM VARIABLE

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(xi) that the random variable equals xi, for each value xi:

p(xi) = P(X=xi)

It satisfies the following conditions:

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

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(xi):

xi	0	1	2	3	4	5
p(xi)	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:

xi	0	1	2	3	4	5

F(xi) 1/32 6/32 16/32 26/32 31/32 32/32

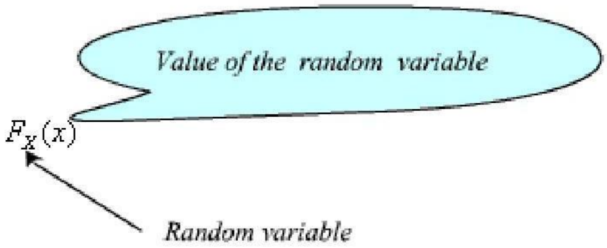
F(x) does not change at intermediate values. For example:

F(1.3) = F(1) = 6/32 and F(2.86) = F(2) = 16/32

Probability Distribution Function

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

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



Properties of the Distribution Function

1. $0 \le F_X(x) \le 1$

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

$$\begin{aligned} & x_1 < x_2 \\ & \Rightarrow \{X(s) \le x_1\} \subseteq \{X(s) \le x_2\} \\ & \Rightarrow P\{X(s) \le x_1\} \le P\{X(s) \le x_2\} \\ & \therefore\ F_X(x_1) < F_X(x_2) \end{aligned}$$

$F_X(x)$ Is right continuous.

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

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

3. $F_X(-\infty) = 0$

Because, $F_X(-\infty) = P\{s \mid X(s) \leq -\infty\} = P(\phi) = 0$.

4. $F_X(\infty) = 1$

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

We have,

$$F_X(x^-) = \lim_{\substack{h \to 0 \\ h > 0}} F_X(x-h)$$

$$= \lim_{h \to 0} P\{X(s) \leq x - h\}$$

6. $P(\{X > x\}) = P(\{x < X < \infty\}) = 1 - F_X(x)$

$$= P\{X(s) \leq x\} - P(X(s) = x)$$

$$= F_X(x) - P(X = x)$$

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 $x < 0$,

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

For $0 \le x < 1$,

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

For $1 \le x < 2$,

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

For $2 \le x < 3$,

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

For $x \ge 3$,

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

For $2 \le x < 3$,

$$F_X(x) = P(\{X \le x\})$$
$$= P(\{X = 0\} \cup \{X = 1\} \cup \{X = 2\})$$
$$= P(\{X = 0\}) + P(\{X = 1\}) + P(\{X = 2\})$$
$$= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

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 $F_X(x)\ \forall x \in X$ is a complete description of the random variable X.

Example 5 Consider the random variable X defined by

$$F_X(x) = 0, \quad x < -2$$
$$= \frac{1}{8}x + \frac{1}{4}, \quad -2 \le x < 0$$
$$= 1, \quad x \ge 0$$

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

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

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

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

Solutio

a) $P(X = 0) = F_X(0^+) - F_X(0^-)$

$$= 1 - \frac{1}{4} = \frac{3}{4}$$

b) $P\{X \le 0\} = F_X(0)$

$$= 1$$

c) $P\{X > 2\} = 1 - F_X(2)$

$$= 1 - 1 = 0$$

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:

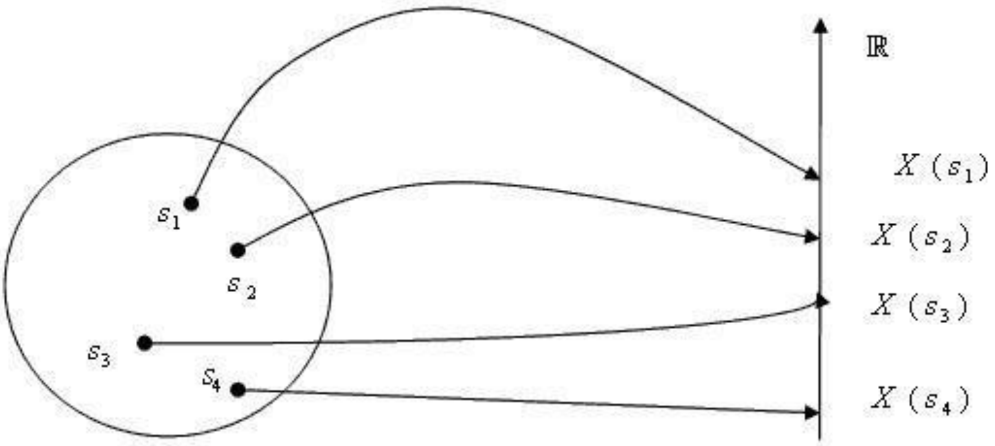
$$\int f(x)\,dx = F(b) - F(a) = P(a < X < b)$$

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

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

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

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

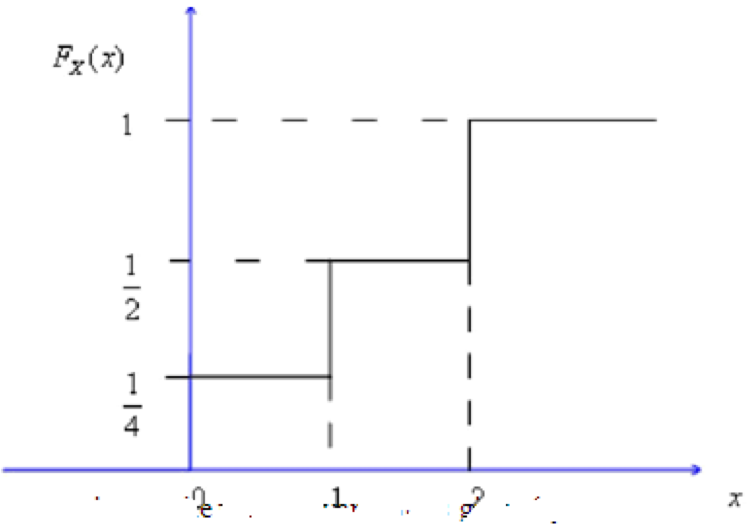


Example 1

Consider the random variable X with the distribution function

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

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. **Bernoulli random variable:**

Suppose X is a random variable that takes two values 0 and 1, with probability mass functions

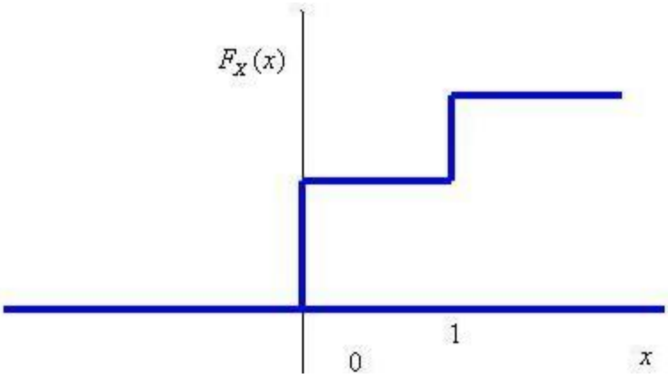
$$p_X(1) = P\{X = 1\} = p$$

And

$$p_X(0) = 1 - p, \qquad 0 \leq p \leq 1$$

Such a random variable *X* is called a **Bernoulli random variable,** because it describes the outcomes of a **Bernoulli trial.**

The typical CDF of the Bernoulli RV X is as shown in Figure 2



Mean and variance of the Bernoulli random:

$$\mu_X = EX = \sum_{k=0}^{1} k p_X(k) = 1 \times p + 0 \times (1-p) = p$$

$$EX^2 = \sum_{k=0}^{1} k^2 p_X(k) = 1 \times p + 0 \times (1-p) = p$$

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

Remark

- The Bernoulli RV is the simplest discrete RV. It can be used as the building block for many discrete RVs.
- For the Bernoulli RV,

  $$EX^m = p \quad m = 1, 2, 3 \ldots$$

  Thus all the moments of the Bernoulli RV have the same value of p.

2. **Binomial random variable**

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

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

where

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

The trials must meet the following requirements:

a. the total number of trials is fixed in advance;
b. there are just two outcomes of each trial; success and failure;
c. the outcomes of all the trials are statistically independent;
d. 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

p.

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

(a) Exactly 2 bit errors will occur

(b) At least 2 bit errors will occur

(c) More than 2 bit errors will occur

(d) 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,

(a) Probability that exactly 2 bit errors will occur

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

(b) Probability that at least 2 bit errors will occur

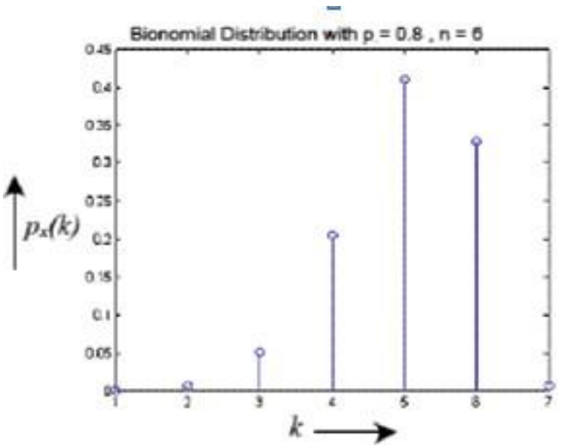
$$= p_X(0) + p_X(1) + p_X(2)$$
$$= 0.99^8 + {}^8C_1 \times 0.01^1 \times 0.99^7 + {}^8C_2 \times 0.01^2 \times 0.99^6$$
$$= 0.9999$$

(c) Probability that more than 2 bit errors will occur

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

(d) Probability that all 8 bits will be erroneous

$$= p_X(8)$$
$$= 0.01^8 = 10^{-16}$$



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

$$EX = \sum_{k=0}^{n} k p_X(k)$$

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

$$= 0 \times q^n + \sum_{k=1}^{n} k\, {}^nC_k p^k (1-p)^{n-k}$$

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

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

$$= np \sum_{k=1}^{n} \frac{n-1!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-1-k_1}$$

$$= np \sum_{k_1=0}^{n-1} \frac{n-1!}{k_1!(n-1-k_1)!} p^{k_1} (1-p)^{n-1-k_1} \quad \text{(Substituting } k_1 = k-1\text{)}$$

$$= np(p+1-p)^{n-1}$$

$$= np$$

Similarly

$$EX^2 = \sum_{k=0}^{n} k^2 p_X(k)$$

$$= \sum_{k=0}^{n} k^2 \, {}^nC_k p^k (1-p)^{n-k}$$

$$= 0^2 \times q^n + \sum_{k=1}^{n} k \, {}^nC_k p^k (1-p)^{n-k}$$

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

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

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

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

$$= np \times (n-1) p + np$$

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

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 $B(n-1, p)$

$$\therefore \sigma_X^2 = \text{variance of } X$$

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

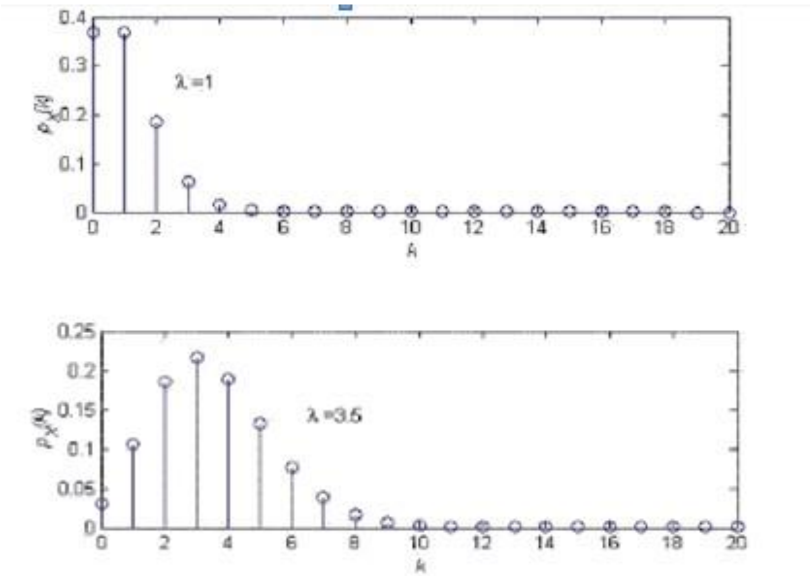
$$= np(1-p)$$

3. Poisson Random Variable

A discrete random variable X is *called a Poisson random variable* with the parameter λ if $\lambda > 0$ and

$P_X(k) = (e^{-\lambda}\lambda^k)/k!$

The plot of the pmf of the Poisson RV is shown



Mean and Variance of the Poisson RV

The mean of the Poisson RV X is given by

$$\mu_X = \sum_{k=0}^{\infty} k p_X(k)$$
$$= 0 + \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!}$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{k-1!}$$

$$= \lambda$$

$$EX^2 = \sum_{k=0}^{\infty} k^2 p_X(k)$$
$$= 0 + \sum_{k=1}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^k}{k-1!}$$
$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{(k-1+1)\lambda^k}{k-1!}$$

$$= e^{-\lambda}\left(0 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k-2!}\right) + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k-1!}$$
$$= e^{-\lambda}\lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{k-2!} + e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{k-1!}$$

$$= e^{-\lambda}\lambda^2 e^{\lambda} + e^{-\lambda}\lambda e^{\lambda}$$
$$= \lambda^2 + \lambda$$
$$\therefore \sigma_X^2 = EX^2 - \mu_X^2 = \lambda$$

Example: The number of calls received in a telephone exchange follows a Poisson distribution with an average of 10 calls per minute. What is the probability that in one-minute duration?

i. no call is received
ii. exactly 5 calls are received
iii. More than 3 calls are received.

Solution: Let X be the random variable representing the number of calls received. Given

Where $\lambda = 10$. Therefore,

i. probability that no call is received $= p_X(0) = e^{-10} =$ 0.000095

ii. probability that exactly 5 calls are received $= p_X(5) = \frac{e^{-10} \times 10^5}{5!} =$ 0.0378

iii. probability that more the 3 calls are received

$$= 1 - \sum_{k=0}^{3} p_X(k) = 1 - e^{-10}\left(1 + \frac{10}{1} + \frac{10^2}{2!} + \frac{10^3}{3!}\right) =$$ 0.9897

Poisson Approximation of the Binomial Random Variable

The Poisson distribution is also used to approximate the binomial distribution $B(n, p)$ when *n* is very large and *p* is small.

Consider binomial RV with $X \sim B(n, p)$ with

$n \to \infty$, $p \to 0$ so that $EX = np = \lambda$ remains constant.

Then

$$p_X(k) = {}^nC_k p^k (1-p)^{n-k}$$

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

$$= \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} p^k (1-p)^{n-k}$$

$$= \frac{n^k (1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{k-1}{n})}{k!} p^k (1-p)^{n-k}$$

$$= \frac{(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{k-1}{n})}{k!} (np)^k (1-p)^{n-k}$$

$$= \frac{n^k(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{k-1}{n})}{k!} p^k(1-p)^{n-k}$$

$$= \frac{(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{k-1}{n})}{k!} (np)^k(1-p)^{n-k}$$

$$= \frac{(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{k-1}{n})(\lambda)^k(1-\frac{\lambda}{n})^n}{k!(1-\frac{\lambda}{n})^k}$$

Note that $\lim_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$.

$$\therefore p_X(k) = \lim_{n\to\infty} \frac{(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{k-1}{n})(\lambda)^k(1-\frac{\lambda}{n})^n}{k!(1-\frac{\lambda}{n})^k} = \frac{e^{-\lambda}\lambda^k}{k!}$$

Thus the Poisson approximation can be used to compute binomial probabilities for large *n*. It also makes the analysis of such probabilities easier. Typical examples are:

- number of bit errors in a received binary data file
- Number of typographical errors in a printed page

Example 4 Suppose there is an error probability of 0.01 per word in typing. What is the probability that there will be more than 1 error in a page of 120 words?

Solution: Suppose X is the RV representing the number of errors per page of 120 words.

$X \sim B(120, p)$ Where $p = 0.01$. Therefore,

$$\therefore \lambda = 120 \times 0.01 = 0.12$$

P(more than one errors)

$$= 1 - p_X(0) - p_X(1)$$

$$\simeq 1 - e^{-\lambda} - \lambda e^{-\lambda}$$

$$= 0.0066$$

In the following we shall discuss some important continuous random variables.

4. 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

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

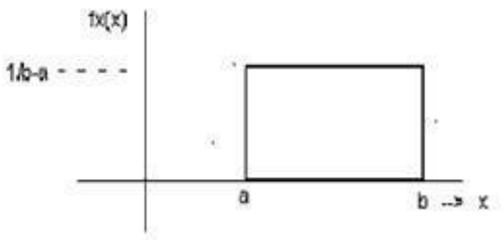


Figure 1

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 $a \le x \le b$

$$\int_{-\infty}^{x} f_X(u)du$$

$$= \int_{a}^{x} \frac{du}{b-a}$$

$$= \frac{x-a}{b-a}$$

For $x > b$,

$F_X(x) = 1$

Figure 2 illustrates the CDF of a uniform random variable.

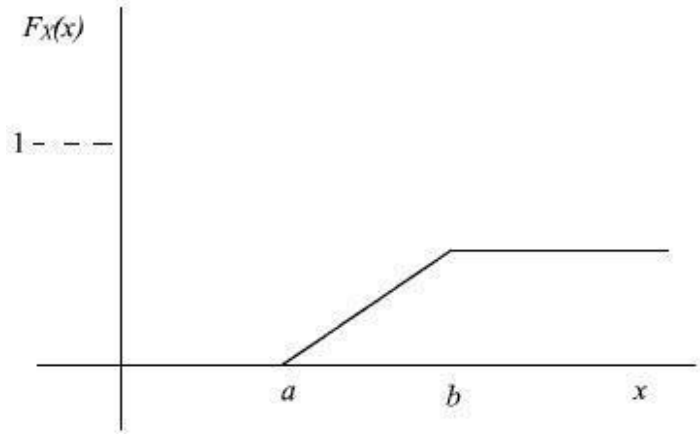


Figure 2: CDF of a uniform random variable

Mean and Variance of a Uniform Random Variable:

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

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

$$\phi_X(w) = Ee^{jwx} = \int_a^b \frac{e^{jwx}}{b-a}dx$$
$$= \frac{e^{jwb} - e^{jwa}}{jw(b-a)}$$

5. 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,

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

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.

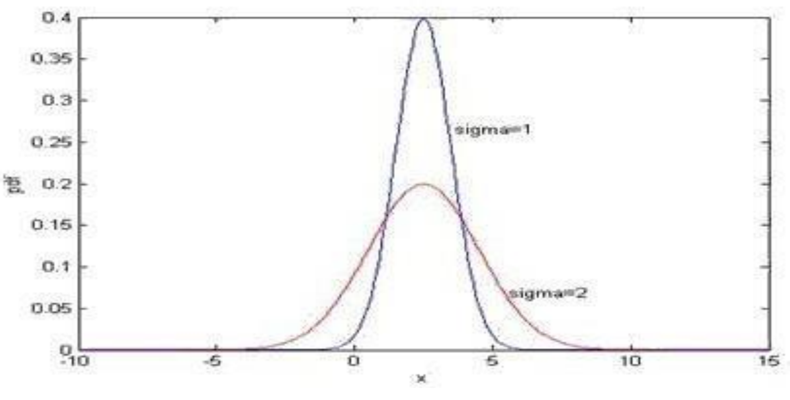


Figure 3

- $f_X(x)$ 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

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

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

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

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,

$$
\begin{aligned}
Q(x) &= 1 - \Phi(x) \\
&= \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{u^2}{2}} du
\end{aligned}
$$

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

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

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

$$EX = \mu_X$$
$$\operatorname{var}(X) = \sigma_X^2$$

Proof:

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

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

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

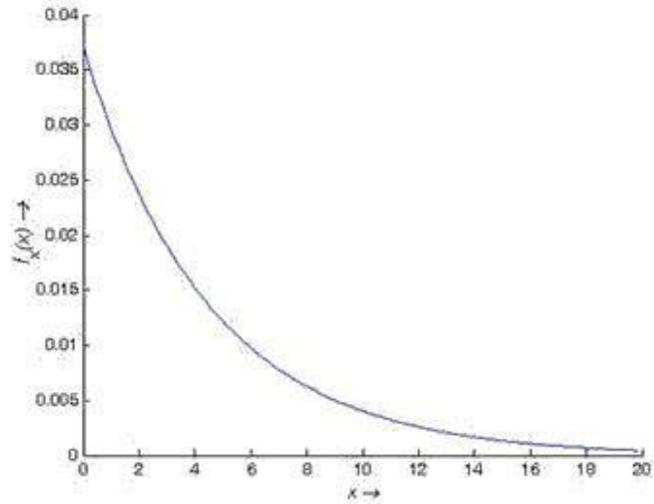
$$
\begin{aligned}
Var(X) &= E(X-\mu_X)^2 \\
&= \frac{1}{\sqrt{2\pi}\sigma_X}\int_{-\infty}^{\infty}(x-\mu_X)^2 e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}dx \\
&= \frac{1}{\sqrt{2\pi}\sigma_X}\int_{-\infty}^{\infty}\sigma_X^2 u^2 e^{-\frac{1}{2}u^2}\sigma_X du \qquad (\text{substituting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

6. Exponential Random Variable

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

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

$\lambda > 0$ 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

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

Then,

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

Rayleigh Random Variable

A Rayleigh random variable X is characterized by the PDF

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where σ is the parameter of the random variable.

probability density functions for the Rayleigh RVs are illustrated in Figure

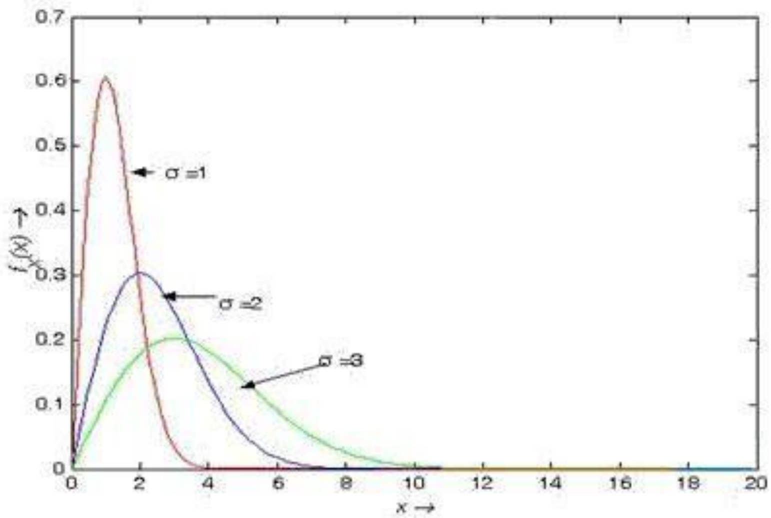


Figure 6

Mean and Variance of the Rayleigh Distribution

$$
\begin{aligned}
EX &= \int_{-\infty}^{\infty} x f_X(x)\,dx \\
&= \int_0^{\infty} x \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}\,dx \\
&= \frac{\sqrt{2\pi}}{\sigma} \int_0^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}\,dx \\
&= \frac{\sqrt{2\pi}}{\sigma} \frac{\sigma^2}{2} \\
&= \sqrt{\frac{\pi}{2}}\sigma
\end{aligned}
$$

similarly

$$
\begin{aligned}
EX^2 &= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx \\
&= \int_0^{\infty} x^2 \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}\,dx \\
&= 2\sigma^2 \int_0^{\infty} u e^{-u}\,du \quad \textbf{(Substituting } u = \frac{x^2}{2\sigma^2}\textbf{)} \\
&= 2\sigma^2 \qquad \textbf{(Noting that } \int_0^{\infty} u e^{-u}\,du \textbf{ is the mean of the exponential RV with } \lambda\textbf{=1)}
\end{aligned}
$$

$$
\begin{aligned}
\therefore \sigma_X^2 &= 2\sigma^2 - \left(\sqrt{\frac{\pi}{2}}\sigma\right)^2 \\
&= (2 - \frac{\pi}{2})\sigma^2
\end{aligned}
$$

Relation between the Rayleigh Distribution and the Gaussian distribution

A Rayleigh RV is related to Gaussian RVs as follow: If $X_1 \sim N(0, \sigma^2)$ and $X_2 \sim N(0, \sigma^2)$ are independent, then the envelope $X = \sqrt{X_1^2 + X_2^2}$ has the Rayleigh distribution with the parameter σ.

We shall prove this result in a later lecture. This important result also suggests the cases where the Rayleigh RV can be used.

Application of the Rayleigh RV

- ✓ Modeling the *root mean square error-*
- ✓ Modeling the envelope of a signal with two *orthogonal components* as in the case of a signal of the following form:

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

$$F_X(x/B) = P[\{X \leq x\}/B]$$
$$= \frac{P[\{X \leq x\} \cap B]}{P(B)} \qquad P(B) \neq 0$$

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

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

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

Example 1 Suppose X is a random variable with the distribution function $F_X(x)$. Define $B = \{X \leq b\}$

$$\begin{aligned} F_X(x/B) &= \frac{P(\{X \leq x\} \cap B)}{P(B)} \\ &= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{P\{X \leq b\}} \\ &= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)} \end{aligned}$$

Case 1: $x < b$

Then

$$\begin{aligned} F_X(x/B) &= \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)} \\ &= \frac{P(\{X \leq x\})}{F_X(b)} = \frac{F_X(x)}{F_X(b)} \end{aligned}$$

And

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

Case 2: $x \geq b$

$$F_X(x/B) = \frac{P(\{X \leq x\} \cap \{X \leq b\})}{F_X(b)}$$

$$= \frac{P(\{X \leq b\})}{F_X(b)} = \frac{F_X(b)}{F_X(b)} = 1$$

$F_X(x/B)$ and $f_X(x/B)$ are plotted in the following figures.

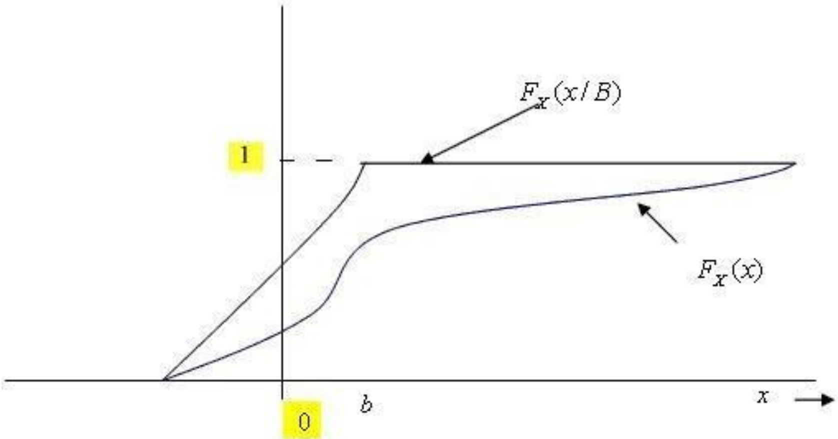
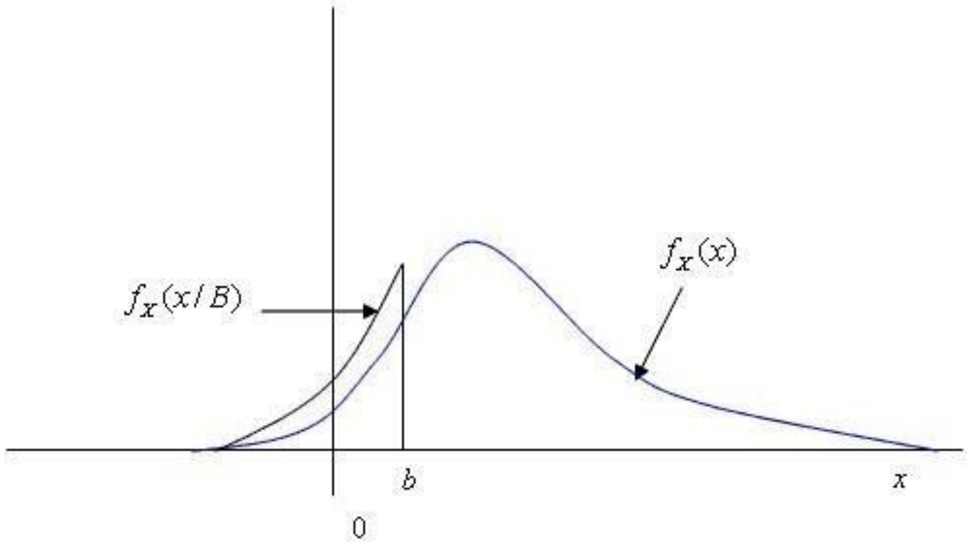


Figure 1



$f_X(x/B)$
$f_X(x)$
b
x
0

Example2: Suppose X is a random variable with the distribution function $F_X(x)$ and $B = \{X > b\}$.

Then

$$F_X(x/B) = \frac{P(\{X \le x\} \cap B)}{P(B)}$$
$$= \frac{P(\{X \le x\} \cap \{X > b\})}{P\{X > b\}}$$
$$= \frac{P(\{X \le x\} \cap \{X > b\})}{1 - F_X(b)}$$

For $x \le b$, $\{X \le x\} \cap \{X > b\} = \phi$. Therefore,

$$F_X(x/B) = 0 \qquad x \le b$$

For $x > b$, $\{X \le x\} \cap \{X > b\} = \{b < X \le x\}$. Therefore,

$$F_X(x/B) = \frac{P(\{b < X \le x\})}{1 - F_X(b)}$$
$$= \frac{F_X(x) - F_X(b)}{1 - F_X(b)}$$

Thus,

$$F_X(x/B) = \begin{cases} 0 \quad , & x \le b \\ \dfrac{F_X(x) - F_X(b)}{1 - F_X(b)} \quad , & \text{otherwise} \end{cases}$$

the corresponding pdf is given by

$$f_X(x/B) = \begin{cases} 0 \quad , & x \le b \\ \dfrac{f_X(x)}{1 - F_X(b)} & \text{otherwise} \end{cases}$$

OPERATION ON RANDOM VARIABLE-EXPECTATIONS

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

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

Provided $\int_{-\infty}^{\infty} x f_X(x)dx$ exists.

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

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

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

$$\begin{aligned}
\therefore \mu_X = E[X] &= \int_{-\infty}^{\infty} x \sum_{i=1}^{N} p_X(x_i)\delta(x - x_i)dx \\
&= \sum_{i=1}^{N} p_X(x_i) \int_{-\infty}^{\infty} x\delta(x - x_i)dx \\
&= \sum_{i=1}^{N} x_i p_X(x_i) \qquad \because \int_{-\infty}^{\infty} x\delta(x - x_i)dx
\end{aligned}$$

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

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

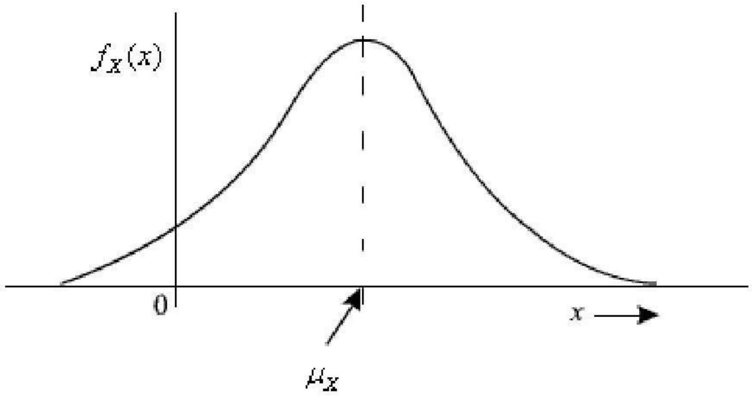


Figure1 Mean of a random variable

Example 1

Suppose X is a random variable defined by the pdf

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

Then

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

Example 2

Consider the random variable X with the pmf as tabulated below

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

Then

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

Example 3 Let *X* be a continuous random variable with

$$f_X(x) = \frac{\alpha}{\pi(x^2 + \alpha^2)} \qquad -\infty < x < \infty, \alpha > 0$$

Then

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

$$= \frac{\alpha}{\pi} \int_0^{\infty} \frac{2x}{x^2 + \alpha^2} dx$$

$$= \frac{\alpha}{\pi} \ln\left(1+x^2\right)\Big]_0^{\infty}$$

Hence EX does not exist. This density function is known as the *Cauchy density function*.

Expected value of a function of a random variable

Suppose $Y = g(X)$ is a real-valued 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 above result in the special case $g(X)$ when $y = g(x)$ is one-to-one and monotonically increasing function of **x** In this case,

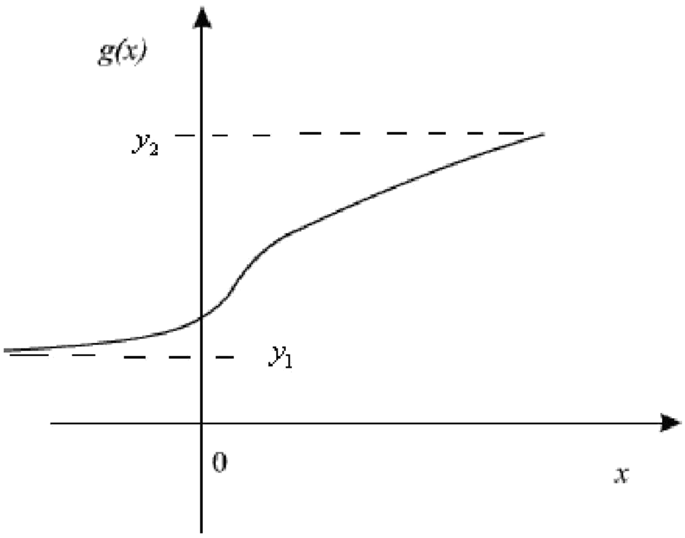


Figure 2

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

$$EY = \int_{-\infty}^{\infty} y f_Y(y)dy$$

$$= \int_{y_1}^{y_2} y \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} dy$$

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,

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

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

$$= \int_{-\infty}^{\infty} c_1 g_1(x) f_X(x)dx + \int_{-\infty}^{\infty} c_2 g_2(x) f_X(x)dx$$

$$= c_1 \int_{-\infty}^{\infty} g_1(x) f_X(x)dx + c_2 \int_{-\infty}^{\infty} g_2(x) f_X(x)dx$$

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

The above property means that E is a linear operator.

MOMENTS ABOUT THE ORIGIN:

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

MOMENTS ABOUT THE MEAN

Variance

Second central moment is called as 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 $p_X(x_i), i = 1, 2, \ldots, N,$

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

The standard deviation of X is defined as $\sigma_X = \sqrt{E(X-\mu_X)^2}$

Example 4

Find the variance of the random variable in the above example

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

Example 5

Find the variance of the random variable discussed in above example. As already computed

$$\mu_X = \frac{17}{8}$$

$$\begin{aligned}\sigma_X^2 &= E(X-\mu_X)^2\\ &= (0-\frac{17}{8})^2\times\frac{1}{8}+(1-\frac{17}{8})^2\times\frac{1}{8}+(2-\frac{17}{8})^2\times\frac{1}{4}+(3-\frac{17}{8})^2\times\frac{1}{2}\\ &= \frac{71}{64}\end{aligned}$$

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 mean.The variances are given by $\sigma_{X_1}^2 = \frac{1}{2}$ and $\sigma_{X_2}^2 = \frac{5}{3}$ implying that X_2 has more spread about the mean.

Properties of variance

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

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

$$\therefore \sigma_X{}^2 = EX^2 - \mu_X{}^2$$

(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,

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

***nth* moment of a random variable**

We can define the *nth* moment and the *nth central*- moment of a random variable *X* by the following relations nth-order moment $EX^n = \int_{-\infty}^{\infty} x^n f_X(x)dx \quad n = 1, 2, ..$

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

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

SKEWNESS

- 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 or 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 kurtosis.

Characteristic function

Consider a random variable X with probability density function $f_X(x)$ The characteristic function of X denoted by $\phi_X(\omega)$, is defined as

$$\phi_X(\omega) = Ee^{j\omega X}$$
$$= \int_{-\infty}^{\infty} e^{j\omega x} f_X(x)dx$$
$$\text{where } j = \sqrt{-1}$$

Note the following:

- $\phi_X(\omega)$, is a complex quantity, representing the Fourier transform of $f_X(x)$ and traditionally using $e^{j\omega X}$ instead of $e^{-j\omega X}$ This implies that the properties of the Fourier transform applies to the characteristic function.

We can get

$f_X(x)$ from $\phi_X(\omega)$, by the inverse transform

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

Example 1

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

$f_X(x) = \frac{1}{b-a} \qquad a \le x \le b$ = 0 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:

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

$$= \frac{1}{b-a} \left. \frac{e^{j\omega x}}{j\omega} \right]_a^b$$

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

Example 2

The characteristic function of the random variable X with

$$f_X(x) = \lambda e^{-\lambda x} \quad \lambda > 0, x > 0 \text{ is}$$

$$\phi_X(\omega) = \int_0^\infty e^{j\omega x} \lambda e^{-\lambda x}$$

$$= \lambda \int_0^\infty e^{-(\lambda - jw)x} dx$$

$$= \frac{\lambda}{\lambda - j\omega}$$

Characteristic function of a discrete random variable

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

Then,

$$\phi_X(\omega) = Ee^{j\omega X}$$
$$= \sum_{X_i \in R_X} p_X(x_i) e^{j\omega x_i}$$

If R_X is the set of integers, we can write

$$\phi_X(\omega) = Ee^{j\omega X}$$
$$= \sum_{X_i \in R_X} p_X(x_i) e^{j\omega x_i}$$

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

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

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

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

Moments and the characteristic function

Given the characteristics function $\phi_X(\omega)$, the *n*th moment is given by

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

To prove this consider the power series expansion of $e^{j\omega X}$

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

Taking expectation of both sides and assuming $EX, EX^2, \ldots, EX^n$ to exist, we get

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

Taking the first derivative of $\phi_X(\omega)$, with respect to ω at $\omega = 0$ we get

$$\left.\frac{d\phi_X(\omega)}{d\omega}\right|_{\omega=0} = jEX$$

Similarly, taking the nth derivative of $\phi_X(\omega)$, with respect to ω at $\omega = 0$ we get

$$\left.\frac{d^n\phi_X(\omega)}{d\omega^n}\right|_{\omega=0} = j^n EX^n$$

Thus ,

$$EX = \frac{1}{j}\left.\frac{d\phi_X(\omega)}{d\omega}\right|_{\omega=0}$$

and generally

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

TRANSFORMATION OF A RANDOM VARIABLE

Description:

Suppose we are given a random variable X with density $fX(x)$. We apply a function g to produce a random variable $Y = g(X)$. We can think of X as the input to a black box,and Y the output.

UNIT-3
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 colour 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 n random variables defined on the same sample space.

Jointly Distributed Random Variables

We may define two or more random variables on the same sample space. Let X and Y be two real random variables defined on the same probability space $(S, \mathbb{F}, P)$. The mapping $S \to \mathbb{R}^2$ such that for $s \in S,\ (X(s), Y(s)) \in \mathbb{R}^2$ is called a joint random variable.

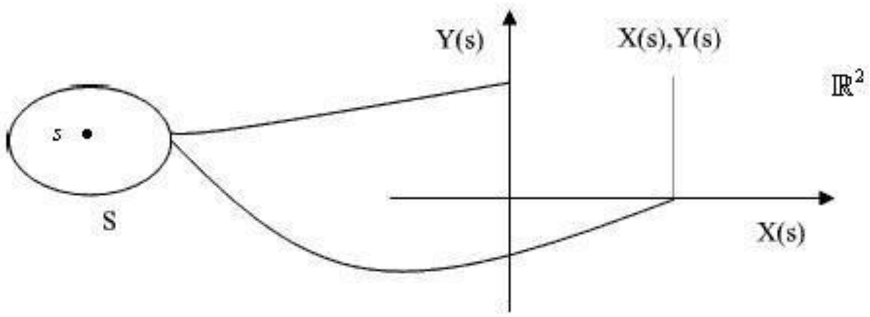


Figure 1

Joint Probability Distribution Function

Recall the definition of the distribution of a single random variable. The event $\{X \le x\}$ 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* or *the joint cumulative distribution function (CDF) of the random variables* X and Y and denoted by $F_{X,Y}(x,y)$.

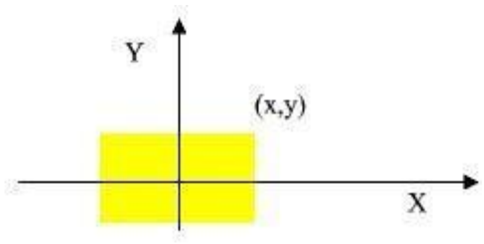


Figure 2

Properties of JPDF

$F_{X,Y}(x,y)$ satisfies the following properties:

1) $F_{X,Y}(x_1,y_1) \leq F_{X,Y}(x_2,y_2)$ if $x_1 \leq x_2$ and $y_1 \leq y_2$

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

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

3) $F_{X,Y}(-\infty,y) = F_{X,Y}(x,-\infty) = 0$

Note that $\{X \leq -\infty, Y \leq y\} \subseteq \{X \leq -\infty\}$

4) $F_{X,Y}(\infty,\infty) = 1$

5) $F_{X,Y}(x,y)$ is right continuous in both the variables.

6) If $x_1 < x_2$ and $y_1 < y_2$

$$P(\{x_1 < X \leq x_2,\ y_1 < Y \leq y_2\}) = F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1)$$

Given $F_{X,Y}(x,y),\ -\infty < x < \infty, -\infty < y < \infty$,we have a complete description of the random variables X and Y.

7) $F_X(x) = F_{XY}(x,+\infty)$

To prove this

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

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

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

Given $F_{X,Y}(x,y),\ -\infty < x < \infty, -\infty < y < \infty$, each of $F_X(x)$ and $F_Y(y)$ is called a marginal Distribution function *or* marginal cumulative distribution function (CDF).

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

- $p_{X,Y}(x,y) = 0 \text{ for } (x,y) \notin R_X \times R_Y$
- $\sum_{(x,y) \in R_X \times R_Y} p_{X,Y}(x,y) = 1$

$$\sum_{(x,y)\in R_X\times R_Y}\sum p_{X,Y}(x,y) = P(\bigcup_{(x,y)\in R_X\times R_Y}\{x,y\})$$
$$=P(R_X \times R_Y)$$
$$=P\{s \mid (X(s),Y(s)) \in (R_X \times R_Y)\}$$
$$=P(S) = 1$$

This is because

• *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

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

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 4 Consider the random variables X and Y with the joint probability mass function as tabulated in Table 1. The marginal probabilities $p_X(x)$ and $p_Y(y)$ are as shown in the last column and the last row respectively.

Y \ X	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	

Table 1

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

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

provided it exists.

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

Properties of Joint Probability Density Function

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

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

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

- The probability of any Borel set can be obtained by

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

Marginal density functions

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

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

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

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

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

Example 5 The joint density function $f_{X,Y}(x,y)$ of the random variables in *Example 3* is

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

$$= \frac{\partial^2}{\partial x \partial y}[(1-e^{-2x})(1-e^{-y})] \qquad x \geq 0, y \geq 0$$

$$= 2e^{-2x}e^{-y} \qquad x \geq 0, y \geq 0$$

Example 6 The joint pdf of two random variables X and Y are given by

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

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

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

$$= c\int_0^2 xdx \int_0^2 ydy$$

$$= 4c$$

$$\therefore 4c = 1$$

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

$$F_{X,Y}(x,y) = \frac{1}{4}\int_0^y\int_0^x uvdudv$$

$$= \frac{x^2y^2}{16} \qquad 0 \leq x \leq 2,\ 0 \leq y \leq 2$$

$$f_X(x) = \int_0^2 \frac{xy}{4} dy \qquad 0 \leq y \leq 2$$

$$= \frac{x}{2}$$

$$\therefore f_X(x) = \frac{x}{2} \qquad 0 \leq y \leq 2$$

Similarly

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

$P(0 < X \leq 1, 0 < Y \leq 1)$

$$= F_{X,Y}(1,1) + F_{X,Y}(0,0) - F_{X,Y}(0,1) - F_{X,Y}(1,0)$$
$$= \frac{1}{16} + 0 - 0 - 0$$
$$= \frac{1}{16}$$

Conditional Distributions

We discussed the conditional CDF and conditional PDF of a random variable conditioned on some events defined in terms of the same random variable. We observed that

$$F_X(x/B) = \frac{P(\{X \leq x\} \cap B)}{P(B)} \qquad P(B) \neq 0$$

and

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

We can define these quantities for two random variables. We start with the *conditional probability mass functions for* two random variables.

Conditional Probability Density Functions

Suppose X and Y are two discrete jointly random variable with the joint PMF $p_{X,Y}(x,y)$. The conditional PMF of Y given $X = x$ is denoted by $p_{Y/X}(y/x)$ and defined as

$$p_{Y/X}(y/x) = P(\{Y = y\}/\{X = x\})$$
$$= \frac{P(\{X = x\} \cap \{Y = y\})}{P\{X = x\}}$$
$$= \frac{p_{X,Y}(x,y)}{p_X(x)} \qquad \text{provided} \quad p_X(x) \neq 0$$

Thus,

$$p_{Y/X}(y/x) = \frac{p_{X,Y}(x,y)}{p_X(x)} \qquad \text{provided} \quad p_X(x) \neq 0$$

Similarly we can define the conditional probability mass function

$$p_{X/Y}(x/y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \quad \text{provided} \quad p_Y(y) \neq 0$$

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

$$F_{Y/X}(y/x) = P(Y \le y / X = x)$$
$$= \frac{P(Y \le y, X = x)}{P(X = x)}$$

as $P(X = x) = 0$ in the above expression. The conditional distribution function is defined in the *limiting sense* as follows:

$$F_{Y/X}(y/x) = lim_{\Delta x \to 0} P(Y \le y / x < X \le x + \Delta x)$$
$$= lim_{\Delta x \to 0} \frac{P(Y \le y, x < X \le x + \Delta x)}{P(x < X \le x + \Delta x)}$$
$$= lim_{\Delta x \to 0} \frac{\int_{\infty}^{y} f_{X,Y}(x,u) \Delta x du}{f_X(x) \Delta x}$$
$$= \frac{\int_{\infty}^{y} f_{X,Y}(x,u) du}{f_X(x)}$$

$$\therefore F_{Y/X}(y/x) = \frac{\int_{\infty}^{y} f_{X,Y}(x,u) du}{f_X(x)}$$

Conditional Probability Density Function

$f_{Y/X}(y / X = x) = f_{Y/X}(y/x)$ is called the *conditional probability density function* of Y given X

Let us define the conditional distribution function .

The conditional density is defined in the limiting sense as follows

$$f_{Y/X}(y/X=x) = \lim_{\Delta y\to 0}(F_{Y/X}(y+\Delta y/X=x) - F_{Y/X}(y/X=x))/\Delta y$$

$$\therefore f_{Y/X}(y/X=x) = \lim_{\Delta y\to 0,\Delta x\to 0}(F_{Y/X}(y+\Delta y/x<X\le x+\Delta x) - F_{Y/X}(y/x<X\le x+\Delta x))/\Delta y$$

Because, $(X=x) = \lim_{\Delta x\to 0}(x<X\le x+\Delta x)$

The right hand side of the highlighted equation is

$$\begin{aligned}
&\lim_{\Delta y\to 0,\Delta x\to 0}(F_{Y/X}(y+\Delta y/x<X<x+\Delta x) - F_{Y/X}(y/x<X<x+\Delta x))/\Delta y \\
&= \lim_{\Delta y\to 0,\Delta x\to 0}(P(y<Y\le y+\Delta y/x<X\le x+\Delta x))/\Delta y \\
&= \lim_{\Delta y\to 0,\Delta x\to 0}(P(y<Y\le y+\Delta y, x<X\le x+\Delta x))/P(x<X\le x+\Delta x)\Delta y \\
&= \lim_{\Delta y\to 0,\Delta x\to 0} f_{X,Y}(x,y)\Delta x\Delta y / f_X(x)\Delta x\Delta y \\
&= f_{X,Y}(x,y)/f_X(x)
\end{aligned}$$

$$\therefore f_{Y/X}(y/x) = f_{X,Y}(x,y)/f_X(x)$$

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

$$\begin{aligned}
f_{X,Y}(x,y) &= k \text{ for } 0\le x\le 1 \\
&= 0 \text{ otherwise}
\end{aligned}$$

find,

(a) k

(b) $f_X(x)$ and $f_Y(y)$

(c) $f_{X/Y}(x/y)$

Solution:

Since $\int\limits_{-\infty}^{\infty}\int f_{X,Y}(x,y)dydx = 1$

We get

$k\,x\frac{1}{2}\,x1x1 = 1$
$\Rightarrow k = 2$

$\therefore f_{X,Y}(x,y) = 2 \;\; for\; 0 \le x \le 1 \text{ as } y \le x$
$\qquad\qquad = 0 \text{ otherwise}$

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

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = 2\int_y^1 dx = 2(1-y)$$

Independent Random Variables (or) Statistical Independence

Let X and Y be two random variables characterized by the joint distribution function

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

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 $\forall (x,y) \in \mathbb{R}^2$, $\{X \le x\}$ and $\{Y \le y\}$ are independent events. Thus,

$$F_{X,Y}(x,y) = P\{X \le x, Y \le y\}$$
$$= P\{X \le x\}P\{Y \le y\}$$
$$= F_X(x)F_Y(y)$$
$$\therefore f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$
$$= \frac{dF_X(x)}{dx}\frac{dF_Y(y)}{dy}$$
$$= f_X(x)f_Y(y)$$
$$\therefore f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

and equivalently $f_{Y/X}(y) = f_Y(y)$

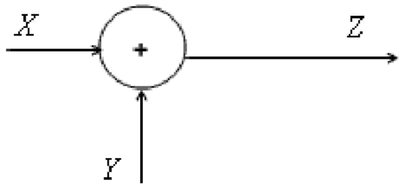
Sum of Two Random Variables

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.

Probability Density of the Function of Two Random Variables

We consider the transformation $g : \mathbb{R}^2 \to \mathbb{R}$.

Consider the event $\{Z \le z\}$ corresponding to each z. We can find a variable subset $D_Z \subseteq \mathbb{R}^2$ such that $D_Z = \{(x,y) \mid g(x,y) \le z\}$.

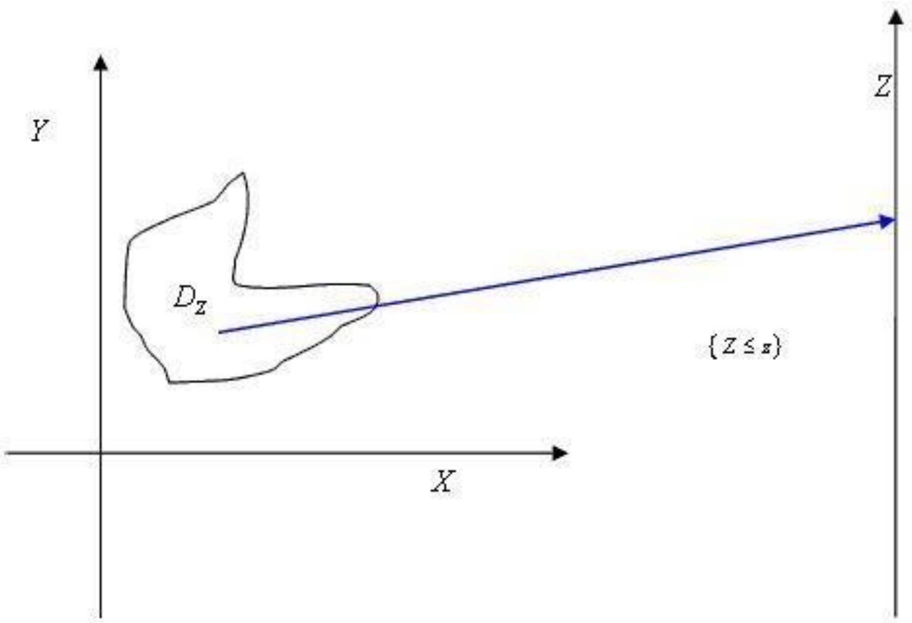


Figure 1

$$\begin{aligned}
\therefore F_Z(z) &= P(\{Z \le z\}) \\
&= P\{(x,y) \mid (x,y) \in D_z\} \\
&= \iint_{(x,y)\in D_z} f_{X,Y}(x,y)\,dydx
\end{aligned}$$

$$\text{and } f_Z(z) = \frac{dF_Z(z)}{dz}$$

Probability density function of $Z = X + Y$.

Consider Figure 2

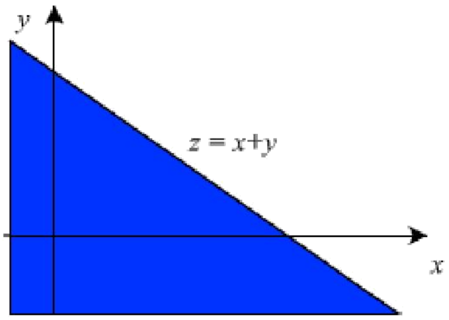


Figure 2

We have

$Z \le z$

$\Rightarrow X + Y \le z$

Therefore, D_Z is the colored region in the Figure 2.

$$\therefore F_Z(z) = \iint_{(x,y)\in D_z} f_{X,Y}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-x} f_{X,Y}(x,y)dy\right]dx$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z} f_{X,Y}(x,u-x)du\right]dx \quad \text{substituting } y = u - x$$

$$= \int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f_{X,Y}(x,u-x)dx\right]du \quad \text{interchanging the order of integration}$$

$$\therefore f_Z(z) = \frac{d}{dz}\int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f_{X,Y}(x,u-x)dx\right]du$$

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

$$\therefore f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x,u-x)\,dx$$

If X and Y are independent

$$f_{X,Y}(x, z-x) = f_X(x) f_Y(z-x)$$

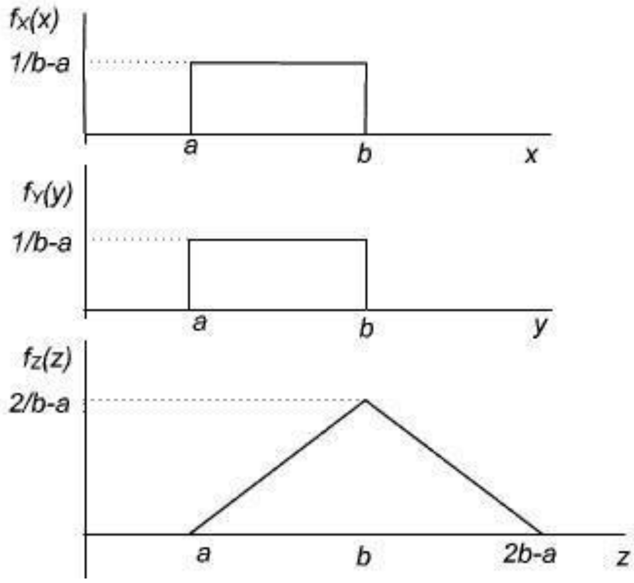
$$\therefore f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

$$= f_X(z) * f_Y(z)$$

Where * is the convolution operation.

Example 1

Suppose X and Y are independent random variables and each uniformly distributed over (a, b). $f_X(x)$ And $f_Y(y)$ are as shown in the figure below.



The PDF of $Z = X + Y$ is a triangular probability density function as shown in the figure.

Central Limit Theorem

Consider n **independent** random variables $X_1, X_2, \ldots, X_n$ .The mean and variance of each of the random variables are assumed to be known. Suppose $E(X_i) = \mu_{X_i}$ and $\operatorname{var}(X_i) = \sigma^2_{X_i}$. Form a random variable

$$Y_n = X_1 + X_2 + \ldots + X_n$$

The mean and variance of Y_n *are given by*

$$EY_n = \mu_{Y_n} = \mu_{X_1} + \mu_{X_2} + \ldots + \mu_{X_n}$$

$$\text{var}(Y_n) = \sigma^2_{Y_n} = E\{\sum_{i=1}^{n}(X_i - \mu_{X_i})\}^2$$

$$= \sum_{i=1}^{n} E(X_i - \mu_{X_i})^2 + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} E(X_i - \mu_i)(X_j - \mu_j)$$

$$= \sigma^2_{X_1} + \sigma^2_{X_2} + \ldots + \sigma^2_{X_n}$$

and $\because X_i$ **and** X_j **are independent for** $i \neq j$.

Thus we can determine the mean and the variance of Y_n.

Can we guess about the probability distribution of Y_n?
The central limit theorem (CLT) provides an answer to this question.

The CLT states that under very general conditions $\left\{Y_n = \sum_{i=1}^{n} X_i\right\}$ converges *in distribution* to $Y \sim N(\mu_Y, \sigma_Y^2)$ as $n \to \infty$. The conditions are:

1. The random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed.
2. The random variables $X_1, X_2, \ldots, X_n$ are independent with same mean and variance, but not identically distributed.
3. The random variables $X_1, X_2, \ldots, X_n$ are independent with different means and same variance and not identically distributed.
4. The random variables $X_1, X_2, \ldots, X_n$ are independent with different means and each variance being neither too small nor too large.

We shall consider the first condition only. In this case, the central-limit theorem can be stated as follows:

Proof of the Central Limit Theorem:

We give a less rigorous proof of the theorem with the help of the characteristic function. Further we consider each of $X_1, X_2, \ldots, X_n$ to have zero mean. Thus, $Y_n = (X_1 + X_2 + \ldots + X_n)/\sqrt{n}$.

$$\mu_{Y_n} = 0,$$

$$\sigma^2_{Y_n} = \sigma^2_X,$$

Clearly, $E(Y_n^3) = E(X^3)/\sqrt{n}$ and so on.

The characteristic function of Y_n is given by

$$\phi_{Y_n}(\omega) = E\left(e^{j\omega Y_n}\right) = E\left(e^{\left(j\omega \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right)}\right)$$

We will show that as $n \to \infty$ the characteristic function ϕ_{Y_n} is of the form of the characteristic function of a Gaussian random variable.

Expanding $e^{j\omega Y_n}$ in power series

$$e^{j\omega Y_n} = 1 + j\omega Y_n + \frac{(j\omega)^2}{2!} Y_n^2 + \frac{(j\omega)^3}{3!} Y_n^3 + \ldots$$

Assume all the moments of Y_n to be finite. Then

$$\phi_{Y_n}(\omega) = E\left(e^{j\omega Y_n}\right) = 1 + j\omega \mu_{Y_n} + \frac{(j\omega)^2}{2!} E(Y_n^2) + \frac{(j\omega)^2}{3!} E(Y_n^3) + \ldots$$

Substituting $\mu_{Y_n} = 0$ and $E(Y_n^2) = \sigma_{Y_n}^2 = \sigma_X^2$, we get

$$\phi_{Y_n}(\omega) = 1 - (\omega^2 / 2!)\sigma_X^2 + \mathrm{R}(\omega, n)$$

where $\mathrm{R}(\omega, n)$ is the average of terms involving ω^3 and higher powers of ω .

Note also that each term in $\mathrm{R}(\omega, n)$ involves a ratio of a higher moment and a power of n and therefore,

$$\lim_{n \to \infty} \mathrm{R}(\omega, n) = 0$$

$$\therefore \lim_{n \to \infty} \phi_{Y_n}(\omega) = 1 - \frac{\omega^2}{2!}\sigma^2{}_X = e^{-\frac{\omega^2 \sigma_X^2}{2}}$$

which is the characteristic function of a Gaussian random variable with 0 mean and variance σ_X^2.

$$Y_n \xrightarrow{d} N(0, \sigma_X^2)$$

OPERATIONS ON MULTIPLE RANDOM VARIABLES

Expected Values of Functions of Random Variables

If $Y = g(X)$ is a function of a continuous random variable X, then

If $Y = g(X)$ is a function of a discrete random variable X, then $EY = Eg(X) = \sum_{x \in R_X} g(x) p_X(x)$

Suppose $Z = g(X,Y)$ is a function of continuous random variables X and Y then the expected value of Z is given by

$$EZ = Eg(X,y) = \int_{-\infty}^{\infty} z f_Z(z) dz$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dxdy$$

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 $(x_i, y_i),\ i = 1,2,..,n$ 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 $n = 3$.

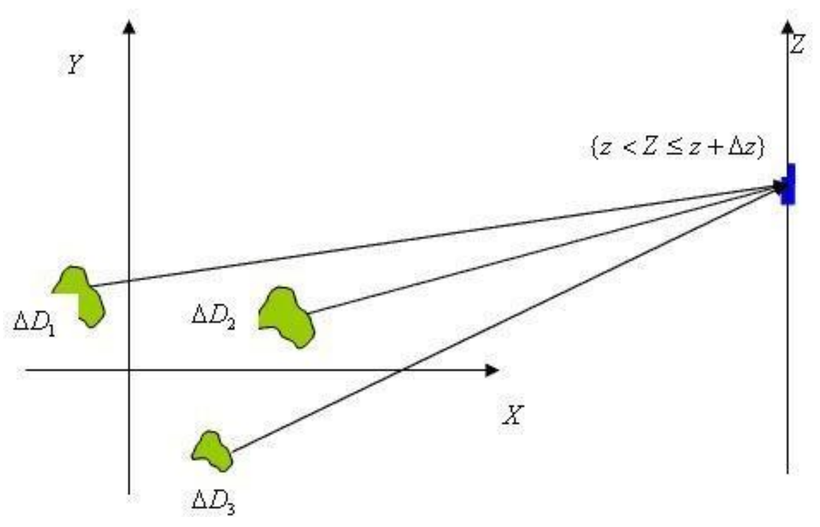


Figure 1

Note that

$$P(\{z < Z \leq z + \Delta z\}) = f_Z(z)\Delta z = \sum_{(x_i, y_i) \in D_i} f_{X,Y}(x_i, y_i)\Delta x_i \Delta y_i$$

$$\therefore z f_Z(z)\Delta z = \sum_{(x_i, y_i) \in \Delta D_i} z f_{X,Y}(x_i, y_i)\Delta x_i \Delta y_i$$

$$= \sum_{(x_i, y_i) \in \Delta D_i} g(x_i, y_i) f_{X,Y}(x_i, y_i)\Delta x_i \Delta y_i$$

As Z is varied over the entire Z axis, the corresponding (non-overlapping) differential regions in $X-Y$ plane cover the entire plane.

$$\therefore \int_{-\infty}^{\infty} z f_Z(z)dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy$$

Thus,

$$\boxed{Eg(X,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy}$$

If $Z = g(X,Y)$ is a function of discrete random variables X and Y, we can similarly show that

$$EZ = Eg(X,Y) = \sum_{x,y \in R_X \times R_Y} \sum g(x,y) p_{X,Y}(x,y)$$

Example 1 The joint pdf of two random variables X and Y is given by

$$f_{X,Y}(x,y) = \frac{1}{4}xy \quad 0 \leq x \leq 2,\ 0 \leq y \leq 2$$
$$= 0 \quad \text{otherwise}$$

Find the joint expectation of $g(X,Y) = X^2Y$

$$\begin{aligned}
Eg(X,Y) &= EX^2Y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dxdy \\
&= \int_0^2\int_0^2 x^2 y \frac{1}{4} xy\,dxdy \\
&= \frac{1}{4}\int_0^2 x^3 dx \int_0^2 y^2 dy \\
&= \frac{1}{4}\times\frac{2^4}{4}\times\frac{2^3}{3} \\
&= \frac{8}{3}
\end{aligned}$$

Example 2 If $Z = aX + bY$, where a and b are constants, then

$$EZ = aEX + bEY$$

Proof:

$$\begin{aligned}
EZ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax + by) f_{X,Y}(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ax f_{X,Y}(x,y)\,dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} by f_{X,Y}(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty} ax \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dydx + \int_{-\infty}^{\infty} by \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dxdy \\
&= a\int_{-\infty}^{\infty} x f_X(x)\,dx + b\int_{-\infty}^{\infty} y f_Y(y)\,dy \\
&= aEX + bEY
\end{aligned}$$

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 $g(X,Y) = XY$.

X \ 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	0.16	

$$\text{Clearly, } EXY = \sum_{x,y \in \mathbb{R}_X \times \mathbb{R}_Y} \sum g(x,y) p_{X,Y}(x,y)$$
$$= 1 \times 1 \times 0.35 + 1 \times 2 \times 0.01$$
$$= 0.37$$

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

$$Eg(X,Y) = Eg_1(X) g_2(Y)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(X) g_2(Y) f_{X,Y}(x,y) dx$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(X) g_2(Y) f_X(x) f_Y(y) dxdy$$
$$= \int_{-\infty}^{\infty} g_1(X) f_X(x) dx \int_{-\infty}^{\infty} g_2(Y) f_Y(y) dy$$
$$= Eg_1(X) Eg_2(Y)$$

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

$$E(X^m Y^n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^m y^n f_{X,Y}(x,y) dxdy$$

And the joint central moment of order $m+n$ is defined as

$$E(X-\mu_X)^m(Y-\mu_Y)^n = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)^m(y-\mu_Y)^n f_{X,Y}(x,y)\,dxdy$$

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

$$E(X^mY^n) = \sum_{(x,y)\in R_{X,Y}}\sum x^m y^n p_{X,Y}(x,y)$$

$$E(X-\mu_X)^m(Y-\mu_Y)^n = \sum_{(x,y)\in R_{X,Y}}\sum (x-\mu_X)^m(y-\mu_Y)^n p_{X,Y}(x,y)$$

(2) If $m=1$ and $n=1$, we have the second-order moment of the random variables X and Y given by

$$E(XY) = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dxdy & \text{if } X \text{ and } Y \text{ are continuous} \\ \sum_{(x,y)\in R_{X,Y}}\sum xy\, p_{X,Y}(x,y) & \text{if } X \text{ and } Y \text{ are discrete} \end{cases}$$

(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

$$Cov(X,Y) = E(X-\mu_X)(Y-\mu_Y)$$

Cov(X, Y) is also denoted as $\sigma_{X,Y}$.

Expanding the right-hand side, we get

$$\begin{aligned} Cov(X,Y) &= E(X-\mu_X)(Y-\mu_Y) \\ &= E(XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y) \\ &= EXY - \mu_Y EX - \mu_X EY + \mu_X\mu_Y \\ &= EXY - \mu_X\mu_Y \end{aligned}$$

The ratio $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X\sigma_Y}$ 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 $Z = aX + Y$

$$E(aX+Y)^2 \geq 0$$
$$\Rightarrow a^2 EX^2 + EY^2 + 2aEXY \geq 0.$$

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

$$\frac{E^2XY}{EX^2} + EY^2 - 2\frac{E^2XY}{EX^2}$$
$$= EY^2 - \frac{E^2XY}{EX^2}$$

Since the minimum is nonnegative,

$$EY^2 - \frac{E^2XY}{EX^2} \geq 0$$
$$\Rightarrow E^2XY \leq EX^2 EY^2$$
$$\Rightarrow |EXY| \leq \sqrt{EX^2}\sqrt{EY^2}$$

Now

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$
$$= \frac{E(X-\mu_X)(Y-\mu_Y)}{\sqrt{E(X-\mu_X)^2 E(Y-\mu_Y)^2}}$$
$$\therefore |\rho(X,Y)| = \frac{|E(X-\mu_X)(Y-\mu_Y)|}{\sqrt{E(X-\mu_X)^2 E(Y-\mu_Y)^2}}$$
$$\leq \frac{\sqrt{E(X-\mu_X)^2}\sqrt{E(Y-\mu_Y)^2}}{\sqrt{E(X-\mu_X)^2 E(Y-\mu_Y)^2}}$$
$$= 1$$

Thus $|\rho(X,Y)| \leq 1$

Uncorrelated random variables

Two random variables X and Y are called *uncorrelated* if

$$Cov(X,Y) = 0$$
which also means
$$E(XY) = \mu_X \mu_Y$$

Recall that if X and Y are independent random variables, then $f_{X,Y}(x,y) = f_X(x) f_Y(y)$.

$$EXY = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dx\,dy \qquad \text{assuming } X \text{ and } Y \text{ are continuous}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

then $= EXEY$

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

$$\phi_{X,Y}(\omega_1, \omega_2) = Ee^{j\omega_1 x + j\omega_2 y}$$

If X and Y are jointly continuous random variables, then

$$\phi_{X,Y}(\omega_1, \omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)e^{j\omega_1 x + j\omega_2 y}dydx$$

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:

$$\phi_{X,Y}(\omega_1, \omega_2) = \sum_{(x,y)\in \mathbb{R}_X \times \mathbb{R}_Y} \sum p_{X,Y}(x,y)e^{j\omega_1 x + j\omega_2 y}$$

Properties of the Joint Characteristic Function

The joint characteristic function has properties similar to the properties of the chacteristic 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

$$\begin{aligned}\phi_{X,Y}(\omega_1, \omega_2) &= Ee^{j\omega_1 X + j\omega_2 Y}\\ &= E(e^{j\omega_1 X}e^{j\omega_2 Y})\\ &= Ee^{j\omega_1 X}Ee^{j\omega_2 Y}\\ &= \phi_X(\omega_1)\phi_Y(\omega_2)\end{aligned}$$

4. We have,

$$\phi_{X,Y}(\omega_1,\omega_2) = Ee^{j\omega_1 X + j\omega_2 Y}$$
$$= E\left(1 + j\omega_1 X + j\omega_2 Y + \frac{j^2(\omega_1 X + j\omega_2 Y)^2}{2} + \ldots\ldots\ldots\ldots\right)$$
$$= 1 + j\omega_1 EX + j\omega_2 EY + \frac{j^2\omega_1^2 EX^2}{2} + \frac{j^2\omega_2^2 EY^2}{2} + \omega_1\omega_2 EXY + \ldots..$$

Hence,

$$\phi_{X,Y}(0,0) = 1$$
$$EX = \frac{1}{j}\frac{\partial}{\partial\omega_1}\phi_{X,Y}(\omega_1,\omega_2)\bigg|_{\omega_1=0}$$
$$EY = \frac{1}{j}\frac{\partial}{\partial\omega_2}\phi_{X,Y}(\omega_1,\omega_2)\bigg|_{\omega_2=0}$$
$$EXY = \frac{1}{j^2}\frac{\partial^2\phi_{X,Y}(\omega_1,\omega_2)}{\partial\omega_1\partial\omega_2}\bigg|_{\omega_1=0,\omega_2=0}$$

In general, the $(m+n)th$ order joint moment is given by

$$EX^mY^n = \frac{1}{j^{m+n}}\frac{\partial^m\partial^n\phi_{X,Y}(\omega_1,\omega_2)}{\partial\omega_1^m\partial\omega_2^n}\bigg|_{\omega_1=0,\omega_2=0}$$

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

$$X \sim N(\mu_X, \sigma_X^2)$$

$$\phi_X(\omega) = Ee^{j\omega X}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} . e^{j\omega x} dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{x^2-2(\mu_X-\sigma_X^2 j\omega)x+(\mu_X-\sigma_X^2 j\omega)^2-(\mu_X-\sigma_X^2 j\omega)^2+\mu_X^2}{\sigma_X^2}} dx$$

$$= e^{\frac{1}{2}\frac{(-\sigma_X^4\omega^2+2\mu_X\sigma_X^2 j\omega)}{\sigma_X^2}} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu_X-\sigma_X^2 j\omega}{\sigma_X}\right)^2} dx}_{\text{Area under a Gaussian}}$$

$$= e^{\mu_X j\omega-\sigma_X^2\omega^2/2} \times 1$$

$$= e^{\mu_X j\omega-\sigma_X^2\omega^2/2}$$

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

$$\phi_{X,Y}(\omega_1,\omega_2) = Ee^{j(X\omega_1+Y\omega_2)}$$
$$= e^{j\mu_X\omega_1+j\mu_Y\omega_2-\frac{1}{2}(\sigma_X^2\omega_1^2+2\rho_{X,Y}\sigma_X\sigma_Y\omega_1\omega_2+\sigma_Y^2\omega_2^2)}$$

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

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}-2\rho_{XY}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}, \quad -\infty < x < \infty, \ -\infty < y < \infty$$

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.

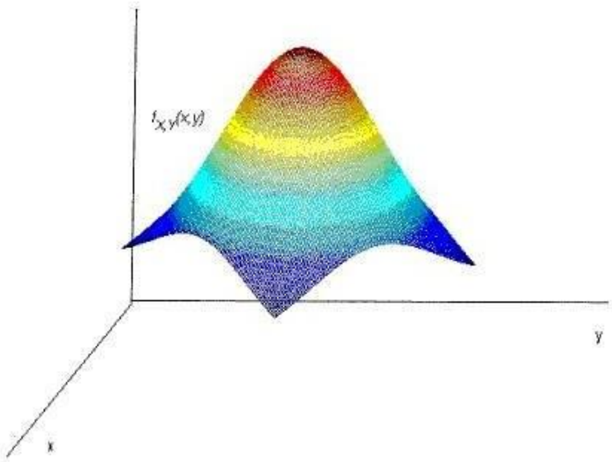


Figure 1 Jointly Gaussian PDF surface

Properties of jointly Gaussian random variables

(1) If X and Y are jointly Gaussian, then X and Y are both Gaussian.

We have

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

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{XY}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} dy$$

$$= \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left[\frac{\rho_{X,Y}^2(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{XY}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} dy$$

$$= \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2\sigma_Y^2(1-\rho_{X,Y}^2)}\left[(y-\mu_Y - \frac{\rho_{X,Y}\sigma_Y(x-\mu_X)}{\sigma_X}\right]^2} dy$$

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

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

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}(1+\sin x\sin y)dydx$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}dydx+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}\sin x\sin ydydx$$

$$=1+\frac{1}{2\pi\sigma_X\sigma_Y}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}}\sin xdx\underbrace{\int_{-\infty}^{\infty}e^{-\frac{1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}}\sin y\;dy}_{\text{integration of an odd function}}$$

$$=1+0$$

$$=1$$

The marginal density $f_X(x)$ is given by

$$f_X(x)=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}(1+\sin x\sin y)dy$$

$$=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}dy+\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}\underbrace{e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}\sin x\sin ydy}_{\text{integration of an odd function}}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}+0$$

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

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

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} e^{\frac{(x-\mu_X)^2}{\sigma_X^2}} \frac{1}{\sqrt{2\pi}\sigma_Y} e^{\frac{(y-\mu_Y)^2}{\sigma_Y^2}}$$

$$= f_X(x) f_Y(y)$$

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 $f_{X,Y}(x,y)$

$$\mu_X = \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 4 \text{ and } \operatorname{cov}(X,Y) = 1.$$

$$\therefore \rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{1}{1\times 2} = \frac{1}{2}$$

and

$$f_{X,Y}(x,y) = \frac{1}{2\pi\times 1\times 2\sqrt{1-\frac{1}{4}}} e^{-\frac{1}{2\times\frac{3}{4}}\left[\frac{x^2}{1} - 2\times\frac{1}{2}\times\frac{xy}{1\times 2} + \frac{y^2}{4}\right]}$$

We have

$$= \frac{1}{2\sqrt{3}\pi} e^{-\frac{2}{3}\left[x^2 - \frac{xy}{2} + \frac{y^2}{4}\right]}$$

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

$$\phi_Z(\omega) = \phi_{X,Y}(a\omega, b\omega)$$

$$= e^{j(\mu_X+\mu_Y)\omega - \frac{1}{2}(a^2\sigma_X^2 + 2\rho_{X,Y}ab\sigma_X\sigma_Y + b^2\sigma_Y^2)\omega^2}$$

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.

Example 3 If $Z = X + Y$ and X and Y are independent, then

$$\begin{aligned}\phi_Z(\omega) &= \phi_{X,Y}(\omega, \omega) \\ &= \phi_X(\omega)\phi_Y(\omega)\end{aligned}$$

Using the property of the Fourier transform, we get

$$f_Z(z) = f_X(z) * f_Y(z)$$

Hence proved.

Univariate transformations

When working on the probability density function (pdf) of a random variable X, one is often led to create a new variable Y defined as a function $f(X)$ of the original variable X. For example, if $X \sim \boldsymbol{N}(\mu, \sigma^2)$, then the new variable:

$$Y = f(X) = (X - \mu)/\sigma$$

Is $\boldsymbol{N}$ (0, 1).

It is also often the case that the quantity of interest is a function of another (random) quantity whose distribution is known. Here are a few examples:

*Scaling: from degrees to radians, miles to kilometers, light-years to parsecs, degrees Celsius to degrees Fahrenheit, linear to logarithmic scale, χ^2 to the distribution of the variance

* Laws of physics: what is the distribution of the kinetic energy of the molecules of a gas if the distribution of the speed of the molecules is known ?

So the general question is:

* If $Y = h(X)$,
* And if $f(x)$ is the pdf of X,

Then what is the pdf $g(y)$ of Y?

TRANSFORMATION OF A MULTIPLE RANDOM VARIABLES

Multivariate transformations

The problem extends naturally to the case when **several** variables Y_j are defined from **several** variables X_i through a transformation $\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x})$.

Here are some examples:

Rotation of the reference frame

Let $f(x, y)$ be the probability density function of the pair of r.v. $\{X, Y\}$. Let's rotate the reference frame $\{x, y\}$ by an angle θ. The new axes $\{x', y'\}$ define two new r. v. $\{X', Y'\}$. What is the joint probability density function of $\{X', Y'\}$?

Polar coordinates

Let $f(x, y)$ be the joint probability density function of the pair of r. v. $\{X, Y\}$, expressed in the Cartesian reference frame $\{x, y\}$. Any point (x, y) in the plane can also be identified by its polar coordinates (r, θ). So any realization of the pair $\{X, Y\}$ produces a pair of values of r and θ therefore defining two new r. v. R and θ.

What is the joint probability density function of R and? What are the (marginal) distributions of R and of θ

Sampling distributions

Let $f(x)$ is the pdf of the r. v. X. Let also $Z_1 = z_1(x_1, x_2... x_n)$ be a statistic, *e.g.* the sample mean. What is the pdf of Z_1?

Z_1 is a function of the n r. v. X_i (with n the sample size), that are lid with pdf $f(x)$. If it is possible to identify n - 1 other independent statistics Zi, $i = 2... n$, then a transformation $\boldsymbol{Z} = \boldsymbol{h}(\boldsymbol{X})$ is defined, and $g(\mathbf{z})$, the joint distribution of $\boldsymbol{Z} = \{Z_1, Z_2, ..., Z_n\}$ can be calculated. The pdf of Z_1 is then calculated as one of the marginal distributions of $\boldsymbol{Z}$ by integrating $g(\mathbf{z})$ over z_i , $i = 2, .., n$.

Integration limits

Calculations on joint distributions often involve multiple integrals whose integration limits are themselves variables. An appropriate change of variables sometimes allows changing all these variables but one into fixed integration limits, thus making the calculation of the integrals much simpler.

Linear Transformations of Random Variables

A **linear transformation** is a change to a variable characterized by one or more of the following operations: adding a constant to the variable, subtracting a constant from the variable, multiplying the variable by a constant, and/or dividing the variable by a constant.

When a linear transformation is applied to a random variable, a new random variable is created. To illustrate, let X be a random variable, and let m and b be constants. Each of the following examples show how a linear transformation of X defines a new random variable Y.

- Adding a constant: Y = X + b
- Subtracting a constant: Y = X - b

- Multiplying by a constant: Y = mX
- Dividing by a constant: Y = X/m
- Multiplying by a constant and adding a constant: Y = mX + b
- Dividing by a constant and subtracting a constant: Y = X/m - b

Suppose the vector of random variables $X = (X_1, ..., X_N)^T$ has the joint distribution $f(x) = f(x_1, ..., x_N)$. Set $Y = AX + B$ for some square matrix A and vector B. If $\det A \neq 0$ then Y has the joint distribution $\frac{1}{\det A} f(A^{-1}(y - B))$.

Indeed, suppose $Y \sim g(y)$ (this is the notation for "the $g(y)$ is the distribution density of Y ") and $X \sim f(x)$. For any domain D of the $Y-$ space we can write

$$\int_D g(y)dy = \text{Prob}(Y \in D) = \text{Prob}(AX + B \in D) =$$

$$= \text{Prob}\left(X \in A^{-1}(D - B)\right) = \int_{A^{-1}(D-B)} f(x)dx =$$

We make the change of variables $y = Ax + B$ in the last integral.

$$= \int_D f\left(A^{-1}(y - B)\right) \left| \frac{D(x)}{D(y)} \right| dy = \int_D f\left(A^{-1}(y - B)\right) \frac{1}{\det A} dy$$

(Linear transformation of random variables)

The linear transformation $\sigma\xi + \mu$ is distributed as $N(\mu, \sigma^2)$. The ξ was defined in the section (Definition of normal variable).

For two independent standard normal variables (s.n.v.) ξ_1 and ξ_2 the combination $\sigma_1\xi_1 + \sigma_2\xi_2$ is distributed as $N\left(0, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$.

A product of normal variables is not a normal variable. See the section on the chi - squared distribution.

UNIT 4

STOCHASTIC PROCESSES-TEMPORAL CHARACTERISTICS

Random Processes

In practical problems, we deal with time varying waveforms whose value at a time is random in nature. For example, the speech waveform recorded by a microphone, the signal received by communication receiver or the daily record of stock-market data represents random variables that change with time. ***How do we characterize such data?*** Such data are characterized as ***random*** or ***stochastic processes***. This lecture covers the fundamentals of random processes.

Recall that a random variable maps each sample point in the sample space to a point in the real line. A random process maps each sample point to a waveform.

Consider a probability space $\{S, \mathbb{F}, P\}$. A *random process* can be defined on $\{S, \mathbb{F}, P\}$ as an indexed family of random variables $\{X(s,t),\ s \in S, t \in \Gamma\}$ where Γ is an index set, which may be discrete or continuous, usually denoting time. Thus a random process is a function of the sample point s and index variable t and may be written as $X(t,s)$.

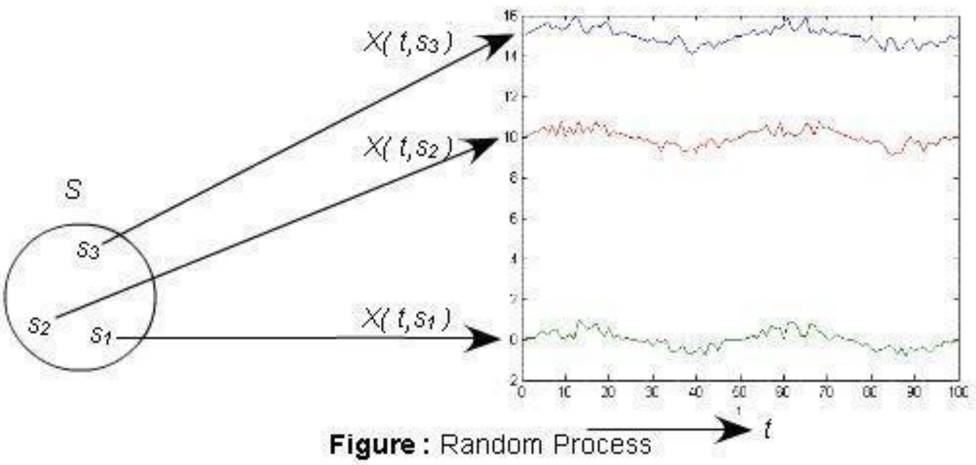


Figure : Random Process

Example 1 Consider a sinusoidal signal $X(t) = A\cos\omega t$ where A is a binary random variable with probability mass functions $p_A(1) = p$ and $p_A(-1) = 1 - p$.

Clearly, $\{X(t),\ t \in \Gamma\}$ is a random process with two possible realizations $X_1(t) = \cos \omega t$ and $X_2(t) = -\cos \omega t$. At a particular time t_0 $X(t_0)$ is a random variable with two values $\cos \omega t_0$ and $-\cos \omega t_0$.

Classification of a Random Process

a) Continuous-time vs. Discrete-time process

If the index set Γ is continuous, $\{X(t),\ t \in \Gamma\}$ is called a **continuous-time process**.

If the index set Γ is a countable set, $\{X(t),\ t \in \Gamma\}$ is called a **discrete-time process**. Such a random process can be represented as $\{X[n],\ n \in Z\}$ and called a *random sequence.* Sometimes the notation $\{X_n,\ n \geq 0\}$ is used to describe a random sequence indexed by the set of positive integers.

We can define a discrete-time random process on discrete points of time. Particularly, we can get a discrete-time random process $\{X[n],\ n \in Z\}$ by sampling a continuous-time process \$\{X(t),\ t \in \Gamma\}$ at a uniform interval T such that $X[n] = X(nT)$.

The discrete-time random process is more important in practical implementations. Advanced statistical signal processing techniques have been developed to process this type of signals.

b) Continuous-state vs. Discrete-state process

The value of a random process $X(t)$ is at any time t can be described from its probabilistic model.

The *state* is the value taken by $X(t)$ at a time t, and the set of all such states is called the ***state space***. A random process is discrete-state if the state-space is finite or countable. It also means that the corresponding sample space is also finite or countable. Otherwise , the random process is called ***continuous state***.

Firtst order and nth order Probability density function and Distribution functions

As we have observed above that $X(t)$ at a specific time t is a random variable and can be described by its *probability distribution function* $F_{X(t)}(x) = P(X(t) \leq x)$. This distribution function is called the *first-order probability distribution function.*

We can similarly define the *first-order probability density function* $f_{X(t)}(x) = \frac{dF_{X(t)}(x)}{dx}$.

To describe $\{X(t), t \in \Gamma\}$, we have to use joint distribution function of the random variables at all possible values of t. For any positive integer n, $X(t_1), X(t_2), \ldots, X(t_n)$ represents n jointly distributed random variables. Thus a random process $\{X(t), t \in \Gamma\}$ can thus be described by specifying the $n-\textbf{th order}$ joint distribution function .

$$F_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n) = P(X(t_1) \le x_1, X(t_2) \le x_2, \ldots, X(t_n) \le x_n), \forall n \ge 1 \text{ and } \forall t_n \in \Gamma$$

or th the $n-th\ order$ joint probability density function

$$f_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n)$$

If $\{X(t), t \in \Gamma\}$ is a discrete-state random process, then it can be also specified by the collection of $n-\textbf{th order}$ joint probability mass function

Moments of a random process

We defined the moments of a random variable and joint moments of random variables. We can define all the possible moments and joint moments of a random process $\{X(t), t \in \Gamma\}$. Particularly, following moments are important.

- $\mu_X(t) =$ Mean of the random process at $t = E(X(t))$
- $R_X(t_1,t_2) = autocorrelation\ function\ \textbf{of the process at times}\ t_1, t_2 = E(X(t_1)X(t_2))$.

Note that

$$R_X(t_1,t_2) = R_X(t_2,t_1) \textbf{ and}$$
$$R_X(t,t) = EX^2(t) = \textbf{sec}ond\ moment\ or\ mean\ square\ value\ at\ \textbf{time}\ t$$

• The *autocovariance function* $C_X(t_1,t_2)$ of the random process at time $t_1 \text{ and } t_2$ is defined by

$$C_X(t_1,t_2) = E(X(t_1) - \mu_X(t_1))(X(t_2) - \mu X(t_2))$$
$$= R_X(t_1,t_2) - \mu_X(t_1)\mu_X(t_2)$$

Clearly

$$C_X(t,t) = E(X(t) - \mu_X(t))^2 = \text{variance of the process at time } t$$

These moments give partial information about the process.

The ratio $\rho_X(t_1,t_2) = \dfrac{C_X(t_1,t_2)}{\sqrt{C_X(t_1,t_1)CX(t_2,t_2)}}$ is called the *correlation coefficient.*

The autocorrelation function and the autocovariance functions are widely used to characterize a class of random process called the wide-sense stationary process.

We can also define higher-order moments like

$R_X(t_1,t_2,t_3) = E(X(t_1), X(t_2), X(t_3))$ = Triple correlation function at t_1, t_2, t_3 etc.

The above definitions are easily extended to a random sequence $\{X[n], n \in \mathbb{Z}\}$.

Cross – covariance function of the processses at times t_1, t_2

$$\begin{aligned} C_{XY}(t_1,t_2) &= E(X(t_1)-\mu_X(t_1))(Y(t_2)-\mu_Y(t_2)) \\ &= R_{XY}(t_1,t_2)-\mu_X(t_1)\mu_Y(t_2) \end{aligned}$$

Cross-correlation coefficient

$$\rho_{XY}(t_1,t_2) = \frac{C_{XY}(t_1,t_2)}{\sqrt{C_X(t_1,t_1)\;C_Y(t_2,t_2)}}$$

On the basis of the above definitions, we can study the degree of dependence between two random processes

This also implies that for such two processes

$R_{XY}(t_1,t_2) = \mu_X(t_1)\mu_Y(t_2)$

Orthogonal processes: Two random processes $\{X(t), t \in \Gamma\}$ and $\{Y(t), t \in \Gamma\}$ are called orthogonal if

$R_{XY}(t_1,t_2) = 0 \; \forall \, t_1, t_2 \in \Gamma$

Stationary Random Process

The concept of stationarity plays an important role in solving practical problems involving random processes. Just like time-invariance is an important characteristic of many deterministic systems, stationarity describes certain time-invariant property of a class of random processes. Stationarity also leads to frequency-domain description of a random process.

Strict-sense Stationary Process

A random process $\{X(t)\}$ is called *strict-sense stationary* (SSS) if its probability structure is invariant with time. In terms of the joint distribution function, $\{X(t)\}$ is called SSS if

$$F_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n) = F_{X(t_1+t_0),X(t_2+t_0),\ldots,X(t_n+t_0)}(x_1, x_2, \ldots, x_n)$$

$$\forall n \in N, \ \forall t_0 \in \Gamma \ \text{ and for all choices of sample points } t_1, t_2, \ldots, t_n \in \Gamma.$$

Thus, the joint distribution functions of any set of random variables $X(t_1), X(t_2), \ldots, X(t_n)$ does not depend on the placement of the origin of the time axis. This requirement is a very strict. Less strict form of stationarity may be defined.

Particularly,

If $F_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1, x_2 \ldots x_n) = F_{X(t_1+t_0),X(t_2+t_0),\ldots,X(t_n+t_0)}(x_1, x_2 \ldots x_n)$ for $n = 1, 2, \ldots, k$, then $\{X(t)\}$ is called kth *order stationary.*

$\{X(t)\}$ Is called kth *order stationary* does not depend on the placement of the origin of the time axis. This requirement is a very strict. Less strict form of stationary may be defined.

- If $\{X(t)\}$ is stationary up to order 1

  $$F_{X(t_1)}(x_1) = F_{X(t_1+t_0)}(x_1), \ \forall t_0 \in T$$

  Let us assume $t_0 = -t_1$. Then

  $$F_{X(t_1)}(x_1) = F_{X(0)}(x_1) \ \text{ which is independent of time.}$$

  As a consequence

  $$EX(t_1) = EX(0) = \mu_X(0) = \text{constant}$$

- If $\{X(t)\}$ is stationary up to order 2

  Put $t_0 = -t_2$

$$F_{X(t_1),X(t_2)}(x_1, x_2) = F_{X(t_1-t_2),X(0)}(x_1, x_2)$$

This implies that the second-order distribution depends only on the time-lag $t_1 - t_2$.

As a consequence, for such a process

$$R_X(t_1,t_2) = E(X(t_1)X(t_2))$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X(0),X(t_1-t_2)}(x_1,x_2)dx_1dx_2$$
$$= R_X(t_1-t_2)$$

Similarly,

$$C_X(t_1,t_2) = C_X(t_1-t_2)$$

Therefore, the autocorrelation function of a SSS process depends only on the time lag

$t_1 - t_2$.

We can also define the joint stationary of two random processes. Two processes

$\{X(t)\}$ And $\{Y(t)\}$ are called jointly *strict-sense stationary* if their joint probability distributions of any order is invariant under the translation of time. A complex random process $\{Z(t) = X(t) + jY(t)\}$ is called SSS if $\{X(t)\}$ and $\{Y(t)\}$ are jointly SSS.

Example 1 A random process is SSS.

This is because $\forall n,$

$$F_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1,x_2,\ldots,x_n) = F_{X(t_1)}(x_1)F_{X(t_2)}(x_2)\ldots F_{X(t_n)}(x_n)$$
$$= F_X(x_1)F_X(x_2)\ldots F_X(x_n)$$
$$= F_{X(t_1+t_0)}(x_1)F_{X(t_2+t_0)}(x_2)\ldots F_{X(t_n+t_0)}(x_n)$$
$$= F_{X(t_1+t_0),X(t_2+t_0),\ldots,X(t_n+t_0)}(x_1,x_2,\ldots,x_n)$$

Wide-sense stationary process

It is very difficult to test whether a process is SSS or not. A subclass of the SSS process called the *wide sense stationary process* is extremely important from practical point of view.

A random process $\{X(t)\}$ is called *wide sense stationary process (WSS)* if

$EX(t) = \mu_X = \text{constant}$

and

$EX(t_1)X(t_2) = R_X(t_1 - t_2)$ is a function of time lag $t_1 - t_2$.

Remark

(1) For a WSS process $\{X(t)\}$

$$EX^2(t) = R_X(0) = \text{constant}$$
$$\text{var}(X(t)=EX^2(t) - (EX(t))^2 = \text{constant}$$
$$C_X(t_1, t_2) = EX(t_1)X(t_2) - EX(t_1)EX(t_2)$$
$$= R_X(t_2 - t_1) - \mu_X^2$$
$$\therefore C_X(t_1, t_2) \text{ is a function of the lag } (t_2 - t_1).$$

(2) An SSS process is always WSS, but the converse is not always true.

Example 3 Sinusoid with random phase

Consider the random process $\{X(t)\}$ given by

$X(t) = A\cos(\omega_0 t + \phi)$ where A and w_0 are constants and ϕ are unifirmly distributed between 0 and 2π.

This is the model of the carrier wave (sinusoid of fixed frequency) used to analyse the noise performance of many receivers.

Note that

$$f_\Phi(\phi) = \begin{cases} \dfrac{1}{2\pi} & 0 \le \phi \le 2\pi \\ 0 & \text{otherwise} \end{cases}$$

By applying the rule for the transformation of a random variable, we get

$$f_{X(t)}(x) = \begin{cases} \dfrac{1}{\pi\sqrt{A^2 - x^2}} & -A \le x \le A \\ 0 & \text{otherwise} \end{cases}$$

Which is independent of t. Hence $\{X(t)\}$ is *first-order stationary.*

Note that

$$\begin{aligned} EX(t) &= EA\cos(\omega_0 t + \phi) \\ &= \int_0^{2\pi} A\cos(\omega_0 t + \phi)\frac{1}{2\pi}d\phi \\ &= 0 \text{ which is a constant} \end{aligned}$$

and

$$\begin{aligned} R_X(t_1, t_2) &= EX(t_1)X(t_2) \\ &= EA\cos(\omega_0 t_1 + \phi)A\cos(\omega_0 t_2 + \phi) \\ &= \frac{A^2}{2}E[\cos(\omega_0 t_1 + \phi + \omega_0 t_2 + \phi) + \cos(\omega_0 t_1 + \phi - \omega_0 t_2 - \phi)] \\ &= \frac{A^2}{2}E[\cos(\omega_0 (t_1 + t_2) + 2\phi) + \cos(\omega_0 (t_1 - t_2))] \\ &= \frac{A^2}{2}\cos(\omega_0 (t_1 - t_2)) \text{ which is a function of the lag } t_1 - t_2. \end{aligned}$$

Hence $\{X(t)\}$ is *wide-sense stationary*

Properties of Autocorrelation Function of a Real WSS Random Process

Autocorrelation of a deterministic signal

Consider a deterministic signal $x(t)$ such that

$$0 < \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x^2(t)dt < \infty$$

Such signals are called *power signals.* For a power signal $x(t)$ the autocorrelation function is defined as

$$R_x(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x(t+\tau)x(t)dt$$

$R_x(\tau)$ Measures the similarity between a signal and its time-shifted version.

Particularly, $R_x(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)dt$ is the mean-square value. If $x(t)$ is a voltage waveform across a 1 ohm resistance, then $R_x(0)$ is the average power delivered to the resistance. In this sense, $R_x(0)$ represents the average power of the signal.

Example 1 Suppose $x(t) = A\cos \omega t$. The autocorrelation function of $x(t)$ at lag τ is given by

$$
\begin{aligned}
R_x(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A\cos \omega(t+\tau) A\cos \omega t dt \\
&= \lim_{T\to\infty} \frac{A^2}{4T} \int_{-T}^{T} [\cos(2\omega t+\tau) + \cos \omega\tau] dt \\
&= \frac{A^2 \cos \omega\tau}{2}
\end{aligned}
$$

We see that $R_x(\tau)$ of the above periodic signal is also periodic and its maximum occurs when $\tau = 0, \pm\frac{2\pi}{\omega}, \pm\frac{4\pi}{\omega}$, etc. The power of the signal is $R_x(0) = \frac{A^2}{2}$.

The autocorrelation of the deterministic signal gives us insight into the properties of the autocorrelation function of a WSS process. We shall discuss these properties next.

Properties of the autocorrelation function of a real WSS process

Consider a real WSS process $\{X(t)\}$. Since the autocorrelation function $R_X(t_1, t_2)$ of such a process is a function of the lag $\tau = t_1 - t_2$, we can redefine a one-parameter autocorrelation function as $R_X(\tau) = EX(t+\tau)X(t)$

If $\{X(t)\}$ is a complex WSS process, then

$$R_X(\tau) = EX(t+\tau)X^*(t)$$

Where $X^*(t)$ is the complex conjugate of $X(t)$. For a discrete random sequence, we can define the autocorrelation sequence similarly.

The autocorrelation function is an important function charactering a WSS random process. It possesses some general properties. We briefly describe them below.

1. $R_X(0) = EX^2(t)$ Is the mean-square value of the process? Thus,

$$R_X(0) = EX^2(t) \geq 0.$$

Remark If $X(t)$ is a voltage signal applied across a 1 ohm resistance, and then $R_X(0)$ is the ensemble average power delivered to the resistance.

2. For a real WSS process $X(t)$, $R_X(\tau)$ is an even function of the time τ. Thus,

$$R_X(-\tau) = R_X(\tau).$$

Because,

$$\begin{aligned} R_X(-\tau) &= EX(t-\tau)X(t) \\ &= EX(t)X(t-\tau) \\ &= EX(t_1+\tau)X(t_1) \quad (\text{Substituting } t_1 = t-\tau) \\ &= R_X(\tau) \end{aligned}$$

Remark For a complex process $R_X(-\tau) = R_X^*(\tau)$

3. $|R_X(\tau)| \leq R_X(0)$. This follows from the Schwartz inequality

$$|< X(t), X(t+\tau) >|^2 \leq \|X(t)\|^2 \|X(t+\tau)\|^2$$

We have

$$\begin{aligned} R_X^2(\tau) &= \{EX(t)X(t+\tau)\}^2 \\ &\leq EX^2(t)EX^2(t+\tau) \\ &= R_X(0)R_X(0) \end{aligned}$$
$$\therefore |R_X(\tau)| \leq R_X(0)$$

4. $R_X(\tau)$ is a positive semi-definite function in the sense that for any positive integer n and real a_j, a_j, $\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j R_X(t_i - t_j) \geq 0$

Proof

Define the random variable

$$Y = \sum_{j=1}^{n} a_i X(t_i)$$

Then we have

$$0 \le EY^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j EX(t_i)X(t_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j R_X(t_i - t_j)$$

It can be shown that *the sufficient condition for a function* $R_X(\tau)$ *to be the autocorrelation function of a real WSS process* $\{X(t)\}$ *is that* $R_X(\tau)$ *be real, even and positive semidefinite.*

If $X(t)$ is MS periodic, then $R_X(\tau)$. Is also periodic with the same period.

Proof: Note that a real WSS random process $\{X(t)\}$ is called mean-square periodic (*MS periodic*) with a period T_p if for every $t \in \Gamma$

$$E(X(t+T_p) - X(t))^2 = 0$$
$$\Rightarrow EX^2(t+T_p) + EX^2(t) - 2EX(t+T_p)X(t) = 0$$
$$\Rightarrow R_X(0) + R_X(0) - 2R_X(T_p) = 0$$
$$\Rightarrow R_X(T_p) = R_X(0)$$

Again

$$(E(X(t+\tau+T_p) - X(t+\tau))X(t))^2 \le E(X(t+\tau+T_p) - X(t+\tau))^2 EX^2(t)$$

(By applying Cauchy Schwartz inequality)

$$\Rightarrow (R_X(\tau+T_p) - R_X(\tau))^2 \le 2(R_X(0) - R_X(T_p))R_X(0)$$
$$\Rightarrow (R_X(\tau+T_p) - R_X(\tau))^2 \le 0 \qquad \because R_X(0) = R_X(T_p)$$
$$\therefore R_X(\tau+T_p) = R_X(\tau)$$

Cross correlation function of jointly WSS processes

If $\{X(t)\}$ and $\{Y(t)\}$ are two real jointly WSS random processes, their cross-correlation functions are independent of t and depends on the time-lag. We can write the cross-correlation function

$$R_{XY}(\tau) = EX(t+\tau)Y(t)$$

The cross correlation function satisfies the following properties:

$$
\begin{aligned}
R_{XY}(\tau) &= EX(t+\tau)Y(t) \\
&= EY(t)X(t+\tau) \\
&= R_{YX}(-\tau)
\end{aligned}
$$

$(ii)\ |R_{XY}(\tau)| \leq \sqrt{R_X(0)R_Y(0)}$

We Have

$$
\begin{aligned}
|R_{XY}(\tau)|^2 &= |EX(t+\tau)Y(t)|^2 \\
&\leq EX^2(t+\tau)EY^2(t) \quad \text{using Cauchy-Schwartz Inequality} \\
&= R_X(0)R_Y(0) \\
\therefore |R_{XY}(\tau)| &\leq \sqrt{R_X(0)R_Y(0)}
\end{aligned}
$$

Further,

$$\sqrt{R_X(0)R_Y(0)} \leq \frac{1}{2}\left(R_X(0)+R_Y(0)\right) \quad \because \text{ Geometric mean} \leq \text{Arithmatic mean}$$

iii. If $X(t)$ and Y (t) are uncorrelated, $R_{XY}(\tau) = EX(t+\tau)EY(t) = \mu_X\mu_Y$

iv. If $X(t)$ and Y (t) are orthogonal processes, $R_{XY}(\tau) = EX(t+\tau)Y(t) = 0$

Example 2

Consider a random process $Z(t)$ which is sum of two real jointly WSS random processes.

$X(t)$ and $Y(t)$. We have

$$
\begin{aligned}
Z(t) &= X(t)+Y(t) \\
R_Z(\tau) &= EZ(t+\tau)Z(t) \\
&= E[X(t+\tau)+Y(t+\tau)][X(t)+Y(t)] \\
&= R_X(\tau)+R_Y(\tau)+R_{XY}(\tau)+R_{YX}(\tau)
\end{aligned}
$$

If $X(t)$ and $Y(t)$ are orthogonal processes,then $R_{XY}(\tau) = R_{YX}(\tau) = 0$

$$\therefore R_Z(\tau) = R_X(\tau)+R_Y(\tau)$$

Example 3

Suppose

$$Z_1(t) = X(t)\cos(\omega_0 t + \Phi) \; and$$
$$Z_2(t) = X(t)\sin(\omega_0 t + \Phi)$$

Where X (*t*) is a WSS process and $\Phi \sim U[0\,2\pi]$

$$\begin{aligned} R_{Z_1Z_2}(\tau) &= E\left[X_1(t)X_2(t-\tau)\right] = \frac{1}{2\pi}\int_0^{2\pi} x_1(t)x_2(t-\tau)d\phi \\ &= E\left[X(t)X(t-\tau)\right]E\left[\cos(\omega_0 t + \Phi)\sin(\omega_0 t - \omega_0\tau + \Phi)\right] \\ &= \frac{1}{2}R_X(\tau)\left\{E\left[\underline{\sin(2\omega_0 t - \omega_0\tau + 2\Phi)}\right] - E\left[\sin(\omega_{0c}\tau)\right]\right\} \\ &= -\frac{1}{2}R_X(\tau)\sin(\omega_0\tau) \end{aligned}$$

Time averages and Ergodicity

Often we are interested in finding the various ensemble averages of a random process $\{X(t)\}$ by means of the corresponding time averages determined from single realization of the random process. For example we can compute the time-mean of a single realization of the random process by the formula

$$\langle \mu_x \rangle_T = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t)dt$$

which is constant for the selected realization. Note that $\langle \mu_x \rangle_T$ represents the dc value of $x(t)$.

Another important average used in electrical engineering is the rms value given by

$$\langle x_{rms} \rangle_T = \lim_{T\to\infty} \sqrt{\frac{1}{2T}\int_{-T}^{T} x^2(t)dt}$$

Time averages of a random process

The time-average of a function $g(X(t))$ of a continuous random process $\{X(t)\}$ is defined by

$$\langle g(X(t)) \rangle_T = \frac{1}{2T}\int_{-T}^{T} g(X(t))dt$$

where the integral is defined in the mean-square sense.

Similarly, the time-average of a function $g(X_n)$ of a continuous random process $\{X_n\}$ is defined by

$$\langle g(X_n)\rangle_N = \frac{1}{2N+1}\sum_{i=-N}^{N} g(X_i)$$

The above definitions are in contrast to the corresponding ensemble average defined by

$$Eg(X(t)) = \int_{-\infty}^{\infty} g(x) f_{X(t)}(x)dx \qquad \text{for continuous case}$$
$$= \sum_{i \in R_{X(t)}} g(x_i) p_{X(t)}(x_i) \qquad \text{for discrete case}$$

The following time averages are of particular interest

(a) Time-averaged mean

$$\langle \mu_X \rangle_T = \frac{1}{2T}\int_{-T}^{T} X(t)dt \qquad \text{(continuous case)}$$
$$\langle \mu_X \rangle_N = \frac{1}{2N+1}\sum_{i=-N}^{N} X_i \qquad \text{(discrete case)}$$

(b) Time-averaged autocorrelation function

$$\langle R_X(\tau) \rangle_T = \frac{1}{2T}\int_{-T}^{T} X(t)X(t+\tau)dt \qquad \text{(continuous case)}$$
$$\langle R_X[m] \rangle_N = \frac{1}{2N+1}\sum_{i=-N}^{N} X_i X_{i+m} \qquad \text{(discrete case)}$$

Note that, $\langle g(X(t))\rangle_T$ and $\langle g(X_n)\rangle_N$ are functions of random variables and are governed by respective probability distributions. However, determination of these distribution functions is difficult and we shall discuss the behaviour of these averages in terms of their mean and variances. We shall further assume that the random processes $\{X(t)\}$ and $\{X_n\}$ are WSS.

Mean and Variance of the Time Averages

Let us consider the simplest case of the time averaged mean of a discrete-time WSS random process $\{X_n\}$ given by

$$\langle \mu_X \rangle_N = \frac{1}{2N+1}\sum_{i=-N}^{N} X_i$$

The mean of $\langle \mu_X \rangle_N$

$$E\langle \mu_X \rangle_N = E\frac{1}{2N+1}\sum_{i=-N}^{N} X_i$$
$$= \frac{1}{2N+1}\sum_{i=-N}^{N} EX_i$$
$$= \mu_X$$

and the variance

$$E\left(\langle \mu_X \rangle_N - \mu_X\right)^2 = E\left(\frac{1}{2N+1}\sum_{i=-N}^{N} X_i - \mu_X\right)^2$$
$$= E\left(\frac{1}{2N+1}\sum_{i=-N}^{N} (X_i - \mu_X)\right)^2$$
$$= \frac{1}{(2N+1)^2}\left[\sum_{i=-N}^{N} E(X_i - \mu_X)^2 + 2\sum_{i=-N}^{N}\sum_{j=-N,\, i\neq j}^{N} E(X_i - \mu_X)(X_j - \mu_X)\right]$$

If the samples $X_{-N}, X_{-N+1}, \ldots, X_1, X_2, \ldots, X_N$ are uncorrelated,

$$E\left(\langle \mu_X \rangle_N - \mu_X\right)^2 = E\left(\frac{1}{2N+1}\sum_{i=-N}^{N} X_i - \mu_X\right)^2$$
$$= \frac{1}{(2N+1)^2}\left[\sum_{i=-N}^{N} E(X_i - \mu_X)^2\right]$$
$$= \frac{\sigma_X^2}{2N+1}$$

We also observe that $\lim_{N\to\infty} E\left(\langle \mu_X \rangle_N - \mu_X\right)^2 = 0$

From the above result, we conclude that $\langle \mu_X \rangle_N \xrightarrow{m.s.} \mu_X$

Let us consider the time-averaged mean for the continuous case. We have

$$\langle \mu_X \rangle_T = \frac{1}{2T}\int_{-T}^{T} X(t)\,dt$$
$$\therefore E\langle \mu_X \rangle_T = \frac{1}{2T}\int_{-T}^{T} EX(t)\,dt$$
$$= \frac{1}{2T}\int_{-T}^{T} \mu_X\,dt = \mu_X$$

and the variance

$$
\begin{aligned}
E\left(\langle \mu_X \rangle_T - \mu_X\right)^2 &= E\left(\frac{1}{2T}\int_{-T}^{T} X(t)dt - \mu_X\right)^2 \\
&= E\left(\frac{1}{2T}\int_{-T}^{T} (X(t) - \mu_X)dt\right)^2 \\
&= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} E(X(t_1) - \mu_X)(X(t_2) - \mu_X)dt_1dt_2 \\
&= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C_X(t_1 - t_2)dt_1dt_2
\end{aligned}
$$

The above double integral is evaluated on the square area bounded by $t_1 = \pm T$ and $t_2 = \pm T$. We divide this square region into sum of trapezoidal strips parallel to $t_1 - t_2 = 0$. (See Figure 1)Putting $t_1 - t_2 = \tau$ and noting that the differential area between $t_1 - t_2 = \tau$ and $t_1 - t_2 = \tau + d\tau$ is $(2T - |\tau|)d\tau$, the above double integral is converted to a single integral as follows:

$$
\begin{aligned}
E\left(\langle \mu_X \rangle_T - \mu_X\right)^2 &= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C_X(t_1 - t_2)dt_1dt_2 \\
&= \frac{1}{4T^2}\int_{-2T}^{2T} (2T - |\tau|)C_X(\tau)d\tau \\
&= \frac{1}{2T}\int_{-2T}^{2T}\left(1 - \frac{|\tau|}{2T}\right)C_X(\tau)d\tau
\end{aligned}
$$

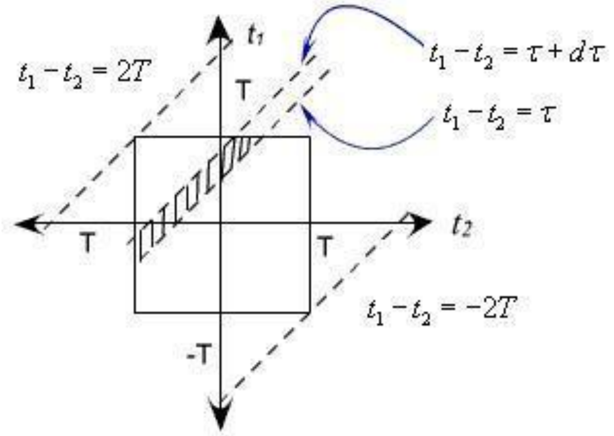


Figure 1

Ergodicity Principle

If the time averages converge to the corresponding ensemble averages in the probabilistic sense, then a time-average computed from a large realization can be used as the value for the corresponding ensemble average. Such a principle is the *ergodicity principle* to be discussed below:

Mean ergodic process

A WSS process $\{X(t)\}$ is said to be ergodic in mean, if $\langle \mu_X \rangle_T \xrightarrow{m.s.} \mu_X$ as $T \to \infty$. Thus for a mean ergodic process $\{X(t)\}$,

$$\lim_{T\to\infty} E\langle \mu_X \rangle_T = \mu_X$$
and
$$\lim_{T\to\infty} \operatorname{var}\langle \mu_X \rangle_T = 0$$

We have earlier shown that

$$E\langle \mu_X \rangle_T = \mu_X$$

and

$$\operatorname{var}\langle \mu_X \rangle_T = \frac{1}{2T}\int_{-2T}^{2T} C_X(\tau)\left[1-\frac{|\tau|}{2T}\right]d\tau$$

therefore, the condition for ergodicity in mean is

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T} C_X(\tau)\left[1-\frac{|\tau|}{2T}\right]d\tau = 0$$

Further,

$$\frac{1}{2T}\int_{-2T}^{2T} C_X(\tau)\left[1-\frac{|\tau|}{2T}\right]d\tau \le \frac{1}{2T}\int_{-2T}^{2T} |C_X(\tau)|d\tau$$

Therefore, a sufficient condition for mean ergodicity is

$$\int_{-2T}^{2T} |C_X(\tau)| d\tau < \infty$$

Example 1 Consider the random binary waveform $\{X(t)\}$ discussed in Example 5 of lecture 32.The process has the auto-covariance function given by

$$C_X(\tau) = \begin{cases} 1 - \dfrac{|\tau|}{T_p} & |\tau| \le T_p \\ 0 & \text{otherwise} \end{cases}$$

Here

$$\begin{aligned} \int_{-2T}^{2T} |C_X(\tau)| d\tau &= 2\int_{0}^{2T} |C_X(\tau)| d\tau \\ &= 2\int_{0}^{T_o} \left(1 - \frac{\tau}{T_p}\right) d\tau \\ &= 2\left(T_p + \frac{T_p^3}{3T_p^2} - \frac{T_p^2}{T_p}\right) \\ &= \frac{2T_p}{3} \end{aligned}$$

$$\int_{-2T}^{2T} |C_X(\tau)| d\tau < \infty$$

hence $\{X(t)\}$ is mean ergodic.

Autocorrelation ergodicity

$$\langle R_X(\tau) \rangle_T = \frac{1}{2T} \int_{-T}^{T} X(t)X(t+\tau)\,dt$$

We consider $Z(t) = X(t)X(t+\tau)$ so that, $\mu_Z = R_X(\tau)$

Then $\{X(t)\}$ will be autocorrelation ergodic if $\{Z(t)\}$ is mean ergodic.

Thus $\{X(t)\}$ will be autocorrelation ergodic if

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\tau_1|}{2T}\right) C_Z(\tau_1)\,d\tau_1 = 0$$

where

$$C_Z(\tau_1) = EZ(t)Z(t-\tau_1) - EZ(t)EZ(t-\tau_1)$$
$$= EX(t)X(t-\tau)X(t-\tau)X(t-\tau-\tau_1) - R_X^2(\tau)$$

$C_Z(\tau_1)$ involves fourth order moment.

Simpler condition for autocorrelation ergodicity of a jointly Gaussian process can be found.

Example 2

Consider the random–phased sinusoid given by
$X(t) = A\cos(\omega_0 t + \phi)$ where A and ω_0 are constants and $\phi \sim U[0,\ 2\pi]$ is a random variable. We have earlier proved that this process is WSS with $\mu_X = 0$ and $R_X(\tau) = \frac{A^2}{2}\cos\omega_0\tau$

For any particular realization $x(t) = A\cos(w_0 t + \phi_1)$,

$$\langle \mu_x \rangle_T = \frac{1}{2T}\int_{-T}^{T} A\cos(w_0 t + \phi_1)dt$$
$$= \frac{1}{Tw_0} A\sin(w_0 T)$$

and

$$\langle R_x(\tau) \rangle_T = \frac{1}{2T}\int_{-T}^{T} A\cos(w_0 t + \phi_1) A\cos(w_0(t+\tau) + \phi_1)\mathrm{d}t$$
$$= \frac{A^2}{4T}\int_{-T}^{T} [\cos w_0\tau + A\cos(w_0(2t+\tau) + 2\phi_1)]\mathrm{d}t$$
$$= \frac{A^2\cos w_0\tau}{2} + \frac{A^2\sin(w_0(2T+\tau))}{4w_0 T}$$

We see that as $T \to \infty$ $\langle \mu_x \rangle_T \to 0$ and $\langle R_x(\tau) \rangle_T \to \frac{A^2\cos w_0\tau}{2}$

For each realization, both the time-averaged mean and the time-averaged autocorrelation function converge to the corresponding ensemble averages. Thus the random-phased sinusoid is ergodic in both mean and autocorrelation.

UNIT 5

STOCHASTIC PROCESSES—SPECTRAL CHARACTERISTICS

Definition of Power Spectral Density of a WSS Process

Let us define the truncated random process $\{X_T(t)\}$ as follows

$$X_T(t) = X(t) \quad -T < t < T$$
$$= 0 \quad \text{otherwise}$$
$$= X(t)rect(\frac{t}{2T})$$

where $rect(\frac{t}{2T})$ is the unity-amplitude rectangular pulse of width $2T$ centering the origin. As $t \to \infty,\ \{X_T(t)\}$ 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 Pareseval'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

$$\frac{1}{2T}\int_{-T}^{T} X_T^2(t)dt = \frac{1}{2T}\int_{-\infty}^{\infty} |FTX_T(\omega)|^2 d\omega$$ 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 $\frac{E|FTX_T(\omega)|^2}{2T}$ the contribution to the average is power at frequency w and represents the power spectral

density of $\{X_T(t)\}$. As $T \to \infty$, 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

$$S_X(\omega) = \lim_{T\to\infty} \frac{E|FTX_T(\omega)|^2}{2T}$$

Relation between the autocorrelation function and PSD: Wiener-Khinchin-Einstein theorem

We have

$$E\frac{|FTX_T(\omega)|^2}{2T} = E\frac{FTX_T(\omega)FTX_T^*(\omega)}{2T}$$

$$= \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} EX_T(t_1)X_T(t_2)e^{-j\omega t_1}e^{+j\omega t_2}dt_1dt_2$$

$$= \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} R_X(t_1-t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2$$

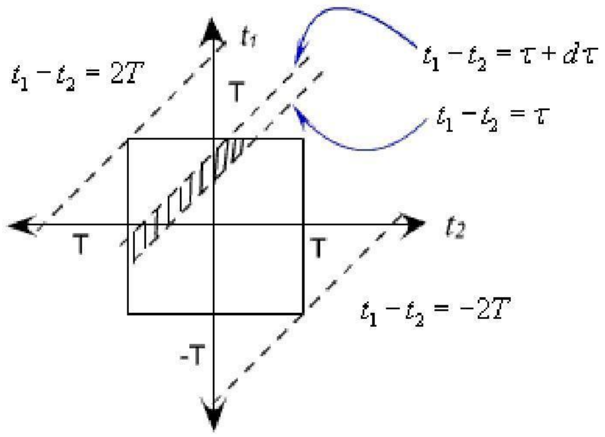


Figure 1

Note that the above integral is to be performed on a square region bounded by $t_1 = \pm T$ and $t_2 = \pm T$ as illustrated in Figure 1.Substitute $t_1 - t_2 = \tau$ so that $t_1 = t_2 + \tau$ 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,

$$E\frac{FTX_T(\omega)X_T^*(\omega)}{2T} = \frac{1}{2T}\int_{-2T}^{2T} R_x(\tau)e^{-j\omega\tau}(2T-|\tau|)d\tau$$

$$= \int_{-2T}^{2T} R_X(\tau)e^{-j\omega\tau}(1-\frac{|\tau|}{2T})d\tau$$

If $R_X(\tau)$ is integral then the right hand integral converges to $\int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau$ as $T \to \infty$

$$\therefore \lim_{T\to\infty}\frac{E|FTX_T(\omega)|^2}{2T} = \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau$$

As we have noted earlier, the power spectral density $S_X(\omega) = \lim_{T\to\infty}\frac{E|FTX_T(\omega)|^2}{2T}$ is the contribution to the average power at frequency ω and is called the power spectral density of $\{X(t)\}$. 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.

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau$$

$$= \int_{-\infty}^{\infty} a^2 e^{-b|\tau|}e^{-j\omega\tau}d\tau$$

$$= \int_{-\infty}^{0} a^2 e^{b\tau}e^{-j\omega\tau}d\tau + \int_{0}^{\infty} a^2 e^{-b\tau}e^{-j\omega\tau}d\tau$$

$$= \frac{a^2}{b-j\omega} + \frac{a^2}{b+j\omega}$$

$$= \frac{2a^2 b}{b^2+\omega^2}$$

The autocorrelation function and the PSD are shown in Figure 2

Figure 2

Example 3 Find the PSD of the amplitude-modulated random-phase sinusoid

$$X(t) = M(t)\ \cos(\omega_c t + \Phi), \qquad \Phi \sim U[0,\ 2\pi]$$

Where *M(t)* is a WSS process independent of Φ.

$$R_X(\tau) = E\ M(t+\tau)\ \cos(\omega_c(t+\tau)+\Phi)\ \ M(t)\cos(\omega_c\tau+\Phi)$$

$$= E\ M(t+\tau)\,M(t)\ E\cos(\omega_c(t+\tau)+\Phi)\ \cos(\omega_c\tau+\Phi)$$

(Using the independence of $M(t)$ and the sinusoid)

$$= R_M(\tau)\,\frac{A^2}{2}\cos\omega_c\tau$$

$$\therefore\ S_X(\omega) = \frac{A^2}{4}\left(S_M(\omega+\omega_c) + S_M(\omega-\omega_c)\right)$$

where $S_M(\omega)$ is the PSD of $M(t)$.

Figure 4 illustrates the above result.

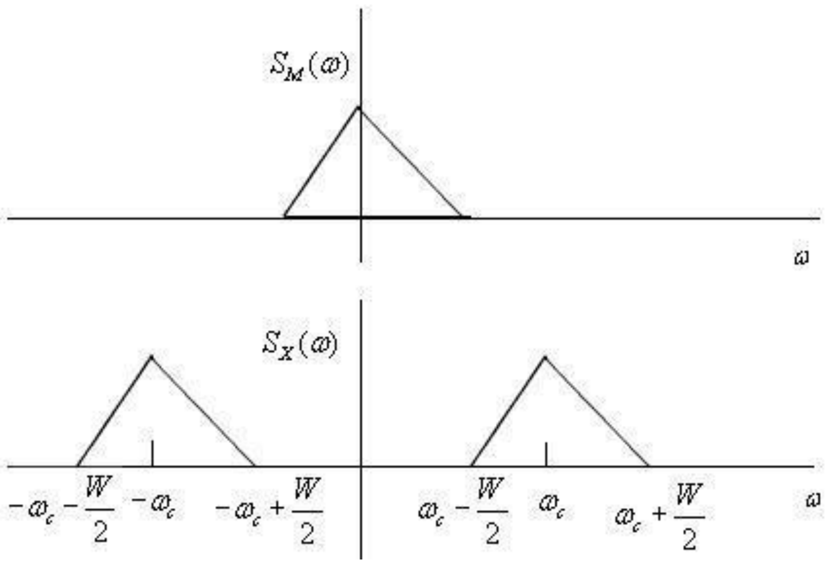


Figure 4

Properties of the PSD

$S_X(\omega)$ being the Fourier transform of $R_X(\tau)$ it shares the properties of the Fourier transform. Here we discuss

important properties of $S_X(\omega)$

1) the average power of a random process $X(t)$ is

$$EX^2(t) = R_X(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)dw$$

2) If $X(t)$ is real, $R_X(\tau)$ is a real and even function of τ. Therefore,

$$\begin{aligned} S_X(\omega) &= \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau \\ &= \int_{-\infty}^{\infty} R_X(\tau)(\cos\omega\tau + j\sin\omega\tau)d\tau \\ &= \int_{-\infty}^{\infty} R_X(\tau)\cos\omega\tau d\tau \\ &= 2\int_{0}^{\infty} R_X(\tau)\cos\omega\tau d\tau \end{aligned}$$

Thus for a real WSS process, the PSD is always real.

3) Thus $S_X(\omega)$ is a real and even function of ω.

4) From the definition $S_X(w) = \lim_{T\to\infty} \frac{E|X_T(\omega)|^2}{2T}$ is always non-negative. Thus $S_X(\omega) \geq 0$.

5) If $X(t)$ has a periodic component, $R_X(\tau)$ is periodic and so $S_X(\omega)$ will have impulses.

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

$$S_{XY}(\omega) = \lim_{T\to\infty} E\frac{FTX_T^*(\omega)FTY_T(\omega)}{2T}$$

Where $FTX_T(\omega)$ and $FTY_T(\omega)$ are the Fourier transform of the truncated processes $X_T(t) = X(t)rect(\frac{t}{2T})$ and $Y_T(t) = Y(t)rect(\frac{t}{2T})$ respectively and $*$ denotes the complex conjugate operation.

We can similarly define $S_{YX}(\omega)$ by

$$S_{YX}(\omega) = \lim_{T\to\infty} E\frac{FTY_T^*(\omega)FTX_T(\omega)}{2T}$$

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 'w'. Some properties of the CPSD of two jointly WSS processes $\{X(t)\}$ and $\{Y(t)\}$ are listed below:

(1) $S_{XY}(\omega) = S_{YX}^{*}(\omega)$

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) $\mathrm{Re}(S_{XY}(\omega))$ is an even function of ω and $\mathrm{Im}(S_{XY}(\omega))$ is an odd function of ω .

We have

$$\begin{aligned}S_{XY}(\omega) &= \int_{-\infty}^{\infty} R_{XY}(\tau)(\cos\omega\tau + j\sin\omega\tau) d\tau \\ &= \int_{-\infty}^{\infty} R_{XY}(\tau)\cos\omega\tau\, d\tau + j\int_{-\infty}^{\infty} R_{XY}(\tau)\sin\omega\tau) d\tau \\ &= \mathrm{Re}(S_{XY}(\omega)) + j\,\mathrm{Im}(S_{XY}(\omega))\end{aligned}$$

where

$\mathrm{Re}(S_{XY}(\omega)) = \int_{-\infty}^{\infty} R_{XY}(\tau)\cos\omega\tau\, d\tau$ is an even function of ω and

$\mathrm{Im}(S_{XY}(\omega)) = \int_{-\infty}^{\infty} R_{XY}(\tau)\sin\omega\tau\, d\tau$ is an odd function of ω and

(3) If $\{X(t)\}$ and $\{Y(t)\}$ are uncorrelated and have constant means, then

$$S_{XY}(\omega) = S_{YX}(\omega) = \mu_X \mu_Y \delta(\omega)$$

Where $\partial(\omega)$ is the Dirac delta function?

Observe that

$$\begin{aligned} R_{XY}(\tau) &= EX(t+\tau)Y(t) \\ &= EX(t+\tau)EY(t) \\ &= \mu_X \mu_Y \\ &= \mu_Y \mu_X \\ &= R_{YX}(\tau) \\ \therefore S_{XY}(\omega) &= S_{YX}(\omega) = \mu_X \mu_Y \delta(\omega) \end{aligned}$$

(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, we have

$$\begin{aligned} R_{XY}(\tau) &= EX(t+\tau)Y(t) \\ &= 0 \\ &= R_{XY}(\tau) \\ \therefore S_{XY}(\omega) &= S_{YX}(\omega) = 0 \end{aligned}$$

(5) the *cross power* P_{XY} between $\{X(t)\}$ and $\{Y(t)\}$ is defined by

$$P_{XY} = \lim_{T \to \infty} \frac{1}{2T} E \int_{-T}^{T} X(t)Y(t)dt$$

Applying Parseval's theorem, we get

$$
\begin{aligned}
P_{XY} &= \lim_{T\to\infty} \frac{1}{2T} E \int_{-T}^{T} X(t)Y(t)dt \\
&= \lim_{T\to\infty} \frac{1}{2T} E \int_{-\infty}^{\infty} X_T(t)Y_T(t)dt \\
&= \lim_{T\to\infty} \frac{1}{2T} E \frac{1}{2\pi} \int_{-\infty}^{\infty} FTX_T^*(\omega) FTY_T(\omega) d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T\to\infty} \frac{EFTX_T^*(\omega) FTY_T(\omega)}{2T} d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}(\omega) d\omega
\end{aligned}
$$

$$
\therefore \; P_{XY} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}(\omega) d\omega
$$

Similarly,

$$
\begin{aligned}
P_{YX} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{YX}(\omega) d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}^*(\omega) d\omega \\
&= P_{XY}^*
\end{aligned}
$$

Example 1 Consider the random process given by $Z(t) = X(t) + Y(t)$ 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,

$$
P_Z(\omega) = P_X(\omega) + P_Y(\omega) + P_{XY}(\omega) + P_{YX}(\omega)
$$

Wiener-Khinchin-Einstein theorem

The Wiener-Khinchin-Einstein theorem is also valid for discrete-time random processes. The power spectral density $S_X(\omega)$ of the WSS process $\{X[n]\}$ is the discrete-time Fourier transform of autocorrelation sequence.

$$S_X(\omega) = \sum_{m=-\infty}^{\infty} R_X[m] e^{-j\omega m} \qquad -\pi \le w \le \pi$$

$R_x[m]$ Is related to $S_X(\omega)$ by the inverse discrete-time Fourier transform and given by

$$R_X[m]) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(\omega) e^{j\omega m} d\omega$$

Thus $R_x[m]$ and $S_X(\omega)$ forms a discrete-time Fourier transform pair. A generalized PSD can be defined in terms of $z-\text{transform}$ as follows

$$S_X(z) = \sum_{m=-\infty}^{\infty} R_x[m] z^{-m}$$

clearly,

$$S_X(\omega) = S_X(z)\big|_{z=e^{j\omega}}$$

Linear time-invariant systems

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 differential 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:

- Analysis of the response of a system
- Finding an LTI system that can optimally 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) as shown in Figure 1. We can thus write,

$$y(t) = T[x(t)]$$

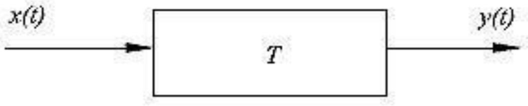


Figure 1

Linear system

The system is called linear if the principle of superposition applies: **the weighted sum of inputs results in the weighted sum of the corresponding outputs.** Thus for a linear system

$$T\left[a_1 x_1(t) + a_2 x_2(t)\right] = a_1 T\left[x_1(t)\right] + a_2 T\left[x_2(t)\right]$$

Example 1 Consider the output of a differentiator, given by

$$y(t) = \frac{d\,x(t)}{dx}$$

Then, $\frac{d}{dt}\left(a_1 x_1(t) + a_2 x_2(t) \right)$

$$= a_1 \frac{d}{dt} x_1(t) + a_2 \frac{d}{dt} x_2(t)$$

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

$$T\,x(t - t_0) = y(t - t_0) \qquad \forall \; t_0$$

It is easy to check that that the differentiator in the above example is a linear time - invariant system.

Response of a linear time-invariant system to deterministic input

As shown in Figure 2, a linear system can be characterised by its impulse response $h(t) = T\delta(t)$ where $\delta(t)$ is the *Dirac delta function.*

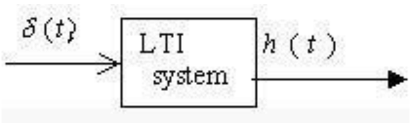


Figure 2

Recall that any function *x*(*t*) can be represented in terms of the Dirac delta function as follows

$$x(t) = \int_{-\infty}^{\infty} x(s)\,\delta(t-s)\,ds$$

If *x*(*t*) is input to the linear system *y* (*t*) = *T x* (*t*), then

$$y(t) = T\int_{-\infty}^{\infty} x(s)\,\delta(t-s)\,ds$$

$$= \int_{-\infty}^{\infty} x(s)\,T\delta(t-s)\,ds \quad [\text{ Using the linearity property }]$$

$$= \int_{-\infty}^{\infty} x(s)\,h(t,s)\,ds$$

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.

We also note that $x(t) * h(t) = h(t) * x(t).$

Thus for a LTI System,

$$y(t) = x(t) * h(t) = h(t) * x(t)$$

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

Figure 3 shows the input-output relationship of an LTI system in terms of the impulse response and the frequency response.

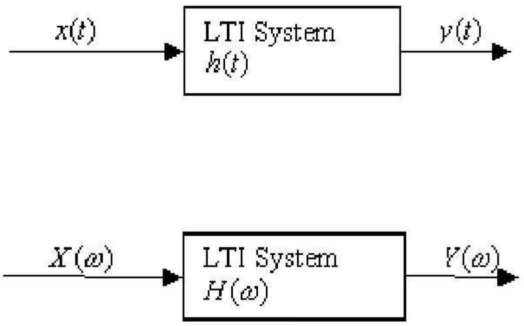


Figure 3

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 sense.

Mean and autocorrelation of the output process $\{Y(t)\}$

$$EY(t) = E\int_{-\infty}^{\infty} h(s)X(t-s)\,ds$$
$$= \int_{-\infty}^{\infty} h(s)EX(t-s)\,ds$$
$$= \int_{-\infty}^{\infty} h(s)\mu_X\,ds$$
$$= \mu_X \int_{-\infty}^{\infty} h(s)\,ds$$
$$= \mu_X H(0)$$

Where $H(0)$ is the frequency response at 0 frequency ($\omega = 0$) and given by

$$H(\omega)\Big|_{\omega=0} = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}dt\Big|_{\omega=0} = \int_{-\infty}^{\infty} h(t)\,dt$$

T The Cross correlation of the input {X*(t)*} and the out put *{Y (t)}* is given by

$$E\big(X(t+\tau)Y(t)\big) = E\,X(t+\tau)\int_{-\infty}^{\infty} h(s)\,X(t-s)\,ds$$
$$= \int_{-\infty}^{\infty} h(s)\;E\,X(t+\tau)\,X(t-s)\,ds$$
$$= \int_{-\infty}^{\infty} h(s)\;R_X(\tau+s)\,ds$$
$$= \int_{-\infty}^{\infty} h(-u)\;R_X(\tau-u)\,du \qquad [\text{ Put } s = -u\ ]$$
$$= h(-\tau) * R_X(\tau)$$

$$\therefore \quad R_{XY}(\tau) = h(-\tau) * R_X(\tau)$$
$$also \quad R_{YX}(\tau) = R_{XY}(-\tau) = h(\tau) * R_X(-\tau)$$
$$= h(\tau) * R_X(\tau)$$

Therefore, the mean of the output process $\{Y(t)\}$ is a constant

The Cross correlation of the input {X *(t)*} and the out put *{Y (t)}* is given by

$$E\left(X(t+\tau)Y(t)\right) = E\, X(t+\tau)\int_{-\infty}^{\infty} h(s)\, X(t-s)\, ds$$

$$= \int_{-\infty}^{\infty} h(s)\;\; E\, X(t+\tau)\, X(t-s)\, ds$$

$$= \int_{-\infty}^{\infty} h(s)\;\; R_X(\tau+s)\,ds$$

$$= \int_{-\infty}^{\infty} h(-u)\;\; R_X(\tau-u)\,du \qquad [\text{ Put } s = -u\,]$$

$$= h(-\tau) * R_X(\tau)$$

$$\therefore \quad R_{XY}(\tau) = h(-\tau) * R_X(\tau)$$

$$\text{also} \quad R_{YX}(\tau) = R_{XY}(-\tau) = h(\tau) * R_X(-\tau)$$

$$= h(\tau) * R_X(\tau)$$

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,

$$\therefore \quad E\left(Y(t+\tau)Y(t)\right) = E\int_{-\infty}^{\infty} h(s)\; X(t+\tau-s)\,ds\,Y(t)$$

$$= \int_{-\infty}^{\infty} h(s)\; E\, X(t+\tau-s)\; Y(t)\; ds$$

$$= \int_{-\infty}^{\infty} h(s)\; R_{XY}(\tau-s)\; ds$$

$$= h(\tau) * R_{XY}(\tau) = h(\tau) * h(-\tau) * R_X(\tau)$$

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

- the output is WSS and
- The input and output are jointly WSS.

The average power of the output process $\{Y(t)\}$ is given by

$$\begin{aligned} P_Y &= R_Y(0) \\ &= R_X(0) * h(0) * h(0) \end{aligned}$$

Power spectrum of the output process

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

$$\begin{aligned} R_{XY}(\tau) &= h(-\tau) * R_X(\tau) \\ and \ R_{YX}(\tau) &= h(\tau) * R_X(\tau) \end{aligned}$$

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

$$S_{YX}(\omega) = H(\omega) S_X(\omega)$$

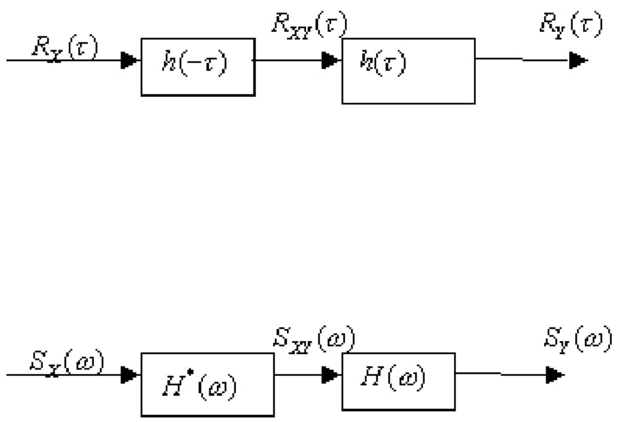


Figure 4

Example 3

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 Figure 7.

Find (a) output PSD $S_Y(\omega)$

(b) output auto correlation function $R_Y(\tau)$

(c) average output power $EY^2(t)$

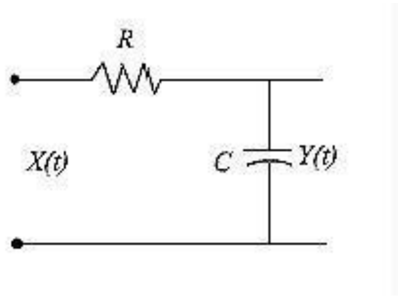


Figure 7

The frequency response of the system is given by

$$H(\omega) = \frac{\frac{1}{jC\omega}}{R + \frac{1}{jC\omega}}$$

$$= \frac{1}{jRC\omega + 1}$$

Therefore,

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega)$$

$$= \frac{1}{R^2C^2\omega^2 + 1} S_X(\omega)$$

(a) $$= \frac{1}{R^2C^2\omega^2 + 1} \frac{N_0}{2}$$

(b) Taking the inverse Fourier transform

$$R_Y(\tau) = \frac{N_0}{4RC} e^{-\frac{|\tau|}{RC}}$$

(c) Average output power

$$EY^2(t) = R_Y(0) = \frac{N_0}{4RC}$$

Rice's representation or quadrature representation of a WSS process

An arbitrary zero-mean WSS process $\{X(t)\}$ can be represented in terms of the slowly varying components $X_c(t)$ and $X_s(t)$ as follows:

$$X(t) = X_c(t)\cos\omega_0 t - X_s(t)\sin\omega_0 t \qquad (1)$$

where ω_0 is a center frequency arbitrary chosen in the band $\{(\omega_0 - \frac{B}{2} \leq |\omega| \leq \omega_0 + \frac{B}{2})\}$. $X_c(t)$ and $X_s(t)$ are respectively called the in-phase and the quadrature-phase components of $X(t)$.

Let us choose a *dual process* $\{Y(t)\}$ such that

$$X(t) + jY(t) = \left(X_c(t) + jX_s(t)\right)e^{j\omega_0 t}$$
$$= \underbrace{\left(X_c(t)\cos\omega_0 t - X_s(t)\sin\omega_0 t\right)}_{X(t)} + j\underbrace{\left(X_c(t)\sin\omega_0 t + X_s(t)\cos\omega_0 t\right)}_{Y(t)}$$

then ,

$$X_c(t) = X(t)\cos\omega_0 t + Y(t)\sin\omega_0 t \qquad (2)$$

and

$$X_s(t) = X(t)\cos\omega_0 t - Y(t)\sin\omega_0 t \qquad (3)$$

For such a representation, we require the processes $\{X_c(t)\}$ and $\{X_s(t)\}$ to be WSS.

Note that

$$EX(t) = \cos\omega_0 t EX_c(t) - \sin\omega_0 t EX_s(t)$$

As $\{X(t)\}$ is zero mean, we require that

$$EX_c(t) = 0$$

And

$$EX_s(t) = 0$$

Again

$$EX_c(t) = \cos\omega_0 t\ EX(t) + \sin\omega_0 t\ EY(t)$$
$$EX_s(t) = \cos\omega_0 t EX(t) - \sin\omega_0 t EY(t)$$

As each of $EX_c(t)$, $EX_s(t)$ and $EX(t)$ is zero-mean, we require that $EY(t) = 0$.

Also

$$R_{X_c}(t+\tau,t) = E[X(t+\tau)\cos\omega_0(t+\tau) + Y(t+\tau)\sin\omega_0(t+\tau)][X(t)\cos\omega_0 t + Y(t)\sin\omega_0 t]$$
$$= R_X(\tau)\cos\omega_0(t+\tau)\cos\omega_0 t + R_Y(\tau)\sin\omega_0(t+\tau)\sin\omega_0 t + R_{XY}(\tau)\cos\omega_0(t+\tau)\sin\omega_0 t$$
$$+ R_{YX}(\tau)\sin\omega_0(t+\tau)\cos\omega_0 t$$

and

$$R_{X_s}(t+\tau,t) = R_X(\tau)\cos\omega_0(t+\tau)\cos\omega_0 t + R_Y(\tau)\sin\omega_0(t+\tau)\sin\omega_0 t$$
$$- R_{XY}(\tau)\cos\omega_0(t+\tau)\sin\omega_0 t - R_{YX}(\tau)\sin\omega_0(t+\tau)\cos\omega_0 t$$

and

$$R_{X_cX_s}(t+\tau,t) = R_X(\tau)\cos\omega_0(t+\tau)\cos\omega_0 t - R_Y(\tau)\sin\omega_0(t+\tau)\sin\omega_0 t$$
$$- R_{XY}(\tau)\cos\omega_0(t+\tau)\sin\omega_0 t + R_{YX}(\tau)\sin\omega_0(t+\tau)\cos\omega_0 t$$

Thus, $R_{X_c}(t+\tau,t)$, $R_{X_s}(t+\tau,t)$ and $R_{X_cX_s}(t+\tau,t)$ will be independent of t if and only if

and

$$\begin{aligned}
R_{X_cX_s}(\tau) &= R_X(\tau)\cos\omega_0(t+\tau)\cos\omega_0 t - R_Y(\tau)\sin\omega_0(t+\tau)\sin\omega_0 t \\
&\quad - R_{XY}(\tau)\cos\omega_0(t+\tau)\sin\omega_0 t + R_{YX}(\tau)\sin\omega_0(t+\tau)\cos\omega_0 t \\
&= R_X(\tau)[\cos\omega_0(t+\tau)\cos\omega_0 t - \sin\omega_0(t+\tau)\sin\omega_0 t] \\
&\quad - R_{XY}(\tau)[\cos\omega_0(t+\tau)\sin\omega_0 t - \sin\omega_0(t+\tau)\cos\omega_0 t] \\
&= R_X(\tau)\cos\omega_0\tau \; - R_{XY}(\tau)\sin(-\omega_0\tau) \\
&= R_X(\tau)\cos\omega_0\tau \; - R_{YX}(\tau)\sin\omega_0\tau
\end{aligned}$$

How to find $\{Y(t)\}$ *satisfying the above two conditions?*

For this, consider $\{Y(t)\}$ to be the *Hilbert transform* of $\{X(t)\}$, i.e.

$$Y(t) = \int_{-\infty}^{\infty} X(s)h(t-s)ds$$

Where $h(t) = \frac{1}{\pi t}$ and the integral is defined in the mean-square sense. See the illustration in Figure 2.

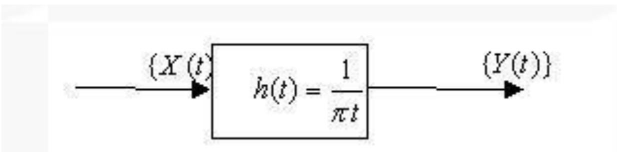


Figure 2

The frequency response $H(\omega)$ of the Hilbert transform is given by

$$H(\omega) = \begin{cases} -j, & \text{if } \omega > 0 \\ j, & \text{if } \omega < 0 \\ 0, & \text{if } \omega = 0 \end{cases}$$

$$\therefore H(\omega) = -j\,\text{sgn}(\omega)$$

$$\text{and } |H(\omega)|^2 = 1$$

$$\therefore S_Y(\omega) = |H(\omega)|^2 S_X(\omega) = S_X(\omega)$$

and

$$S_{XY}(\omega) = H(\omega) S_{XX}(\omega) = \begin{cases} jS_{XX}(\omega), & \text{for } \omega > 0 \\ -jS_{XX}(\omega), & \text{for } \omega < 0 \end{cases}$$

$$S_{YX}(\omega) = H^*(\omega) S_{XX}(\omega) = \begin{cases} -jS_{XX}(\omega), & \text{for } \omega > 0 \\ jS_{XX}(\omega), & \text{for } \omega < 0 \end{cases}$$

The Hilbert transform of *Y*(*t*) satisfies the following spectral relations

$$S_Y(\omega) = S_X(\omega)$$
$$\text{and}$$
$$S_{XY}(\omega) = -S_{YX}(\omega)$$

From the above two relations, we get

$$R_X(\tau) = R_Y(\tau)$$
$$\text{and}$$
$$R_{XY}(\tau) = -R_{YX}(\tau)$$

The Hilbert transform of $X(t)$ is generally denoted as $\hat{X}(t)$. Therefore, from (2) and (3) we establish

$$X_C(t) = X(t)\cos\omega_0 t + \hat{X}(t)\sin\omega_0 t,$$
$$X_s(t) = X(t)\cos\omega_0 t - \hat{X}(t)\sin\omega_0 t$$

and

$$X(t) = X_c(t)\cos\omega_0 t - X_S(t)\sin\omega_0 t$$

The realization for the in phase and the quadrature phase components is shown in Figure 3 below.

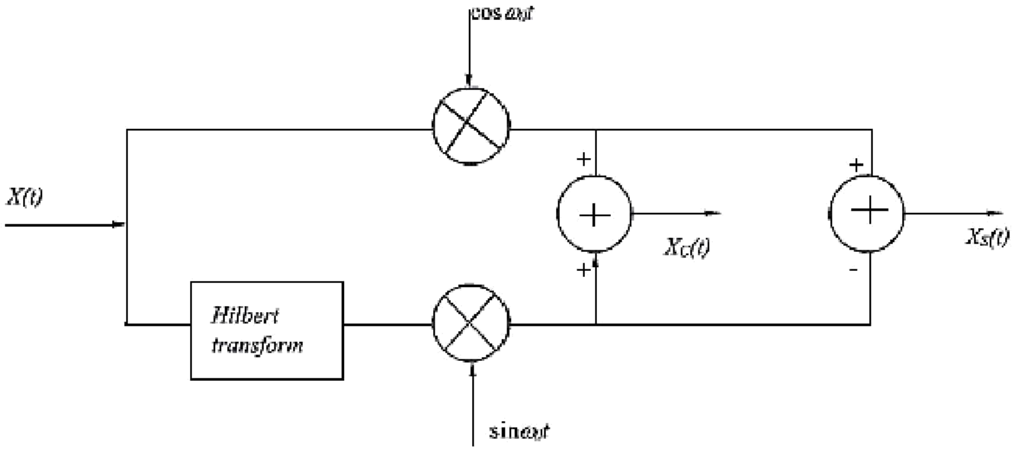


Figure 3

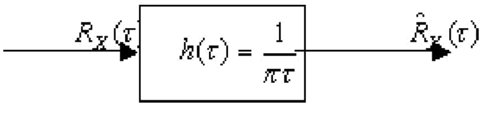
From the above analysis, we can summarize the following expressions for the autocorrelation functions

$$R_{X_c}(\tau) = R_{X_s}(\tau)$$
$$= R_X(\tau)\cos\omega_0\tau + R_{\hat{X}X}(\tau)\sin\omega_0\tau$$
$$= R_X(\tau)\cos\omega_0\tau + h(t) * R_X(\tau)\sin\omega_0\tau \qquad \because R_{\hat{X}X}(\tau) = h(\tau) * R_X(\tau)$$
$$= R_X(\tau)\cos\omega_0\tau + \hat{R}_X(\tau)\sin\omega_0\tau$$

Where

$$\hat{R}_X(\tau) = \text{Hilbert transform of } R_X(\tau)$$
$$= \int_{-\infty}^{\infty} \frac{1}{\pi s} R_X(\tau - s)ds$$

See the illustration in Figure 4



The variances $\sigma^2_{X_c}$ and $\sigma^2_{X_s}$ are given by

$$\sigma^2_{X_c} = \sigma^2_{X_s} = R_X(0).$$

Taking the Fourier transform of $R_{X_c}(\tau)$ and $R_{X_s}(\tau)$, we get

$$S_{X_c}(\omega) = S_{X_s}(\omega) = \begin{cases} S_X(\omega - \omega_0) + S_X(\omega + \omega_0) & |\omega| \le B \\ 0 & \text{otherwise} \end{cases}$$

Similarly,

$$\begin{aligned} R_{X_c X_s}(\tau) &= R_X(\tau)\sin\omega_0\tau - R_{\hat{X}X}(\tau)\cos\omega_0\tau \\ &= R_X(\tau)\sin\omega_0\tau - \hat{R}_X(\tau)\cos\omega_0\tau \end{aligned}$$

and

$$S_{X_c X_s}(\omega) = \begin{cases} j[S_X(\omega + \omega_0) - S_X(\omega - \omega_0)] & |\omega| \le B \\ 0 & \text{otherwise} \end{cases}$$

Notice that the cross power spectral density $S_{X_c X_s}(\omega)$ is purely imaginary. Particularly, if $S_X(\omega)$ is locally symmetric about ω_0

$$S_{X_c X_s}(\omega) = 0$$

Implying that

$$R_{X_c X_s}(\tau) = 0$$

Consequently, the zero-mean processes $\{X_c(t)\}$ and $\{X_s(t)\}$ are also uncorrelated

PROBABILITY THEORY AND STOCHASTIC PROCESSES

Important Questions

UNIT-I

Two marks Questions:

1. Define Demorgans‟ law.

2. Give the classical definition of Probability.

3. Define probability using the axiomatic approach.

4. Write the statement of multiplication theorem.

5. What are the conditions for a function to be a random variable?

Three Marks Questions:

1. Define sample space and classify the types of sample space.

2. Define Joint and Conditional Probability.

3. Define Equally likely events, Exhaustive events and Mutually exclusive events.

4. Show that P(AUB)=P(A)+P(B)-P(A∩B).

5. Define Random variable and Write the classifications of Random variable.

6. In the experiment of tossing a dice, what is the probability of face having 3 dots or 6 dots to appear?

Ten Marks Questions:

1.a) State and Prove Bayes‟ theorem.

b) Write the Mathematical model of experiment.

2. 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

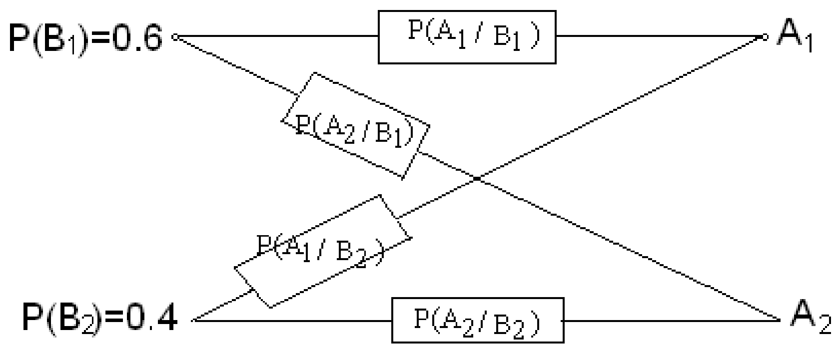
3. a) Two boxes are selected randomly. The first box contains 2 white balls and 3 black balls. The second box contains 3 white and 4 black balls. What is the probability of drawing a white ball.

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?

4. a) 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.

b) 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. In a binary communication system, the errors occur with a probability of "p", In a block of "n" bits transmitted, what is the probability of receiving

i) at the most 1 bit in error

ii at least 4 bits in error

7. Let A and B are events in a sample space S. Show that if A and B are independent, then so are

a) A and *B* b) *A* and B c) *A* and *B*

9.a) An experiment consist of rolling a single die. Two events are defined as A = { a 6 shows up}: and B= {a 2 or a 5 shows up}

i) Find P(A) and P(B)

ii) Define a third event C so that P(C) = 1-P(A)-P(B)

b) The six sides of a fair die are numbered from t to 6. The die is rolled 4 times. How many sequences of the four resulting numbers are possible?

10.a) State and prove the total probability theorem?

b) Explain about conditional probability.

11.In the experiment of tossing a die, all the even numbers are equally likely to appear and similarly the odd numbers. An odd number occurs thrice more frequently than an even number. Find the probability that

a) an even number appears

b) a prime number appears

c) an odd numbers appears

d) an odd prime number appears.

UNIT-II

Two marks Questions:

1. Define Probability density and distribution function.

2. Define the expected value of Discrete Random variable and Continuous Random Variable.

3. Define Moment generating function and Characteristic Function of a Random variable.

4. Define moments about origin and central moments.

5. Show that $Var(kX)=k^2\ var(X)$, here k is a constant.

6. Define skew and coefficient of skewness.

7. Find the Moment generating function of two independent Random variables X_1 & X_2.

8. Write the statement of Chebychev‟s inequality.

Three marks Questions:

1. Derive the expression for the density function of Discrete Random variable.

2. Find the variance of X for uniform density function.

3. Define various types of transformation of Random variables.

4. Write the properties of Gaussian density curve.

5. Find the maximum value of Gaussian density function.

6. In an experiment when two dice are thrown simultaneously, find expected value of the sum of number of points on them.
7. Derive the expression for distribution function of uniform Random variable.

Ten Marks Questions:

1.a) The exponential density function given by

$$f_X(x) = \begin{cases} (1/b)e^{-(x-a)/b} & x > a \\ = 0 & x < a \end{cases}$$

Find the mean and variance.

b) Define Moment Generating Function and write any two properties.

2. Derive the Binomial density function and find mean & variance.
3. Derive the Poisson density function and find mean & variance.

4. 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) $Y=\frac{X+a}{b}$

5. 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) $p\{6< X \leq 9\}$

6. Let X be a Continuous random variable with density function

$$f(x)= \begin{cases} \frac{x}{9} + K & 0 \leq x \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of K and also find $P\{2 \leq X \leq 5\}$

7. a) Verify the Characteristic function of a random variable is having its maximum magnitude at ω=0 and find its maximum value.

b) Find the Moment generating function of exponential distribution?

8. The probability density function of a random variable X is given by $f(x) = \frac{x^2}{81}$ for -3<x<6 and equal to zero otherwise. Find the density function of $Y=\frac{1}{3}(12-x)$

9. a)Write short notes on Gaussian distribution and also find its mean?
   b) 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.

10.a) State and prove the chebychev"s inequality theorem?
   b) b) Find the probability of getting a total of 5, at-least once in 4 tosses of a pair of fair dice.

UNIT-III

Two marks Questions:

1. Define the statistical Independence of the Random variables.

2. Define point conditioning & interval conditioning distribution function.

3. Give the statement of central limit theorem.

4. Define correlation and covariance of two random variables X& Y.

5. Define the joint Gaussian density function of two random variables.

Three Marks Questions:

1. If E[X]=2, E[Y]=3, E[XY]=10, $E[X^2]=9$, and $E[Y^2]=16$ then find variance & covariance of X&Y.

2. The joint probability density function of X&Y is

$$f_{X,Y}(x,y) = \begin{cases} c(2x+y); & 0 \le x \le 2,\ 0 \le y \le 3 \\ 0; & \text{else} \end{cases}$$

Then find the value of constant c.

3. Define correlation coefficient with two properties.

4. Show that var(X+Y) = var(x)+var(Y), if X&Y are statistical independent random variables.

5. Define Marginal distribution & Density functions.

Ten Marks Questions:

1. a) State and prove the density function of sum of two random variables.

b) The joint density function of two random variables X and Y is

$$f_{XY}\ x, y = \frac{x + y^{\ 2}}{40} \ ; -1 < x < 1 \ and -3 < y < 3$$
$$0; otherwise$$

Find the variances of X and Y.

2. a) 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.

b) State and prove any three properties of joint characteristic function.

3.a) State and explain the properties of joint density function

b) The joint density function of random variables X and Y is

$$f_{XY}(x, y) = \begin{cases} 8xy; & 0 \le x < 1, 0 < y < 1 \\ 0, otherwise \end{cases}$$

Find f(y/x) and f(x/y)

4. The input to a binary communication system is a random variable X, takes on one of two values 0 and 1, with probabilities ¾ and ¼ 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 $P\left(\frac{Y=1}{X=1}\right) = \frac{3}{4}$ 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) Probability that the transmitted message is a1. If the received is a 1.

5. Let X and Y be the random variables defined as X=Cosθ and Y=Sinθ where θ is a uniform random variable over (0, 2π)

a) Are X and Y Uncorrelated?

b) Are X and Y Independent?

6. a) Define and State the properties of joint cumulative distribution function of two random variables X and Y.

b) A joint probability density function is $f_{x,y}(x,y) = \begin{cases} \frac{1}{24} & 0 < x < 6, 0 < y < 4 \\ 0 & \text{else where} \end{cases}$

Find the expected value of the function g(X,Y)= $(XY)^2$

7. State and prove the central limit theorem.

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. A certain binary system transmits two binary states X = +1 and X = -1 with equal probability. There are three possible states with the receiver, such as Y = +1, 0 and -1. The performance of the communication system is given as

P(y = +1/X = +1) =0.2;

P(Y= +1/X= -1) = 0.1; P(Y = 0/X = +1) = P(Y = 0/X = -1) = 0.05. Find

a) P(Y = 0)

b) P(X = +1/Y = +1)

c) P(X = -1/Y = 0).

10. 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.

UNIT-IV

Two marks Questions:

1. Define wide sense stationary random processes.

2. Give the statement of ergodic theorem.

3. Define the auto covariance & cross covariance functions of Random processes X(t).

4. When two random processes X(t)& Y(t) are said to be independent.

5. Define the cross correlation function between two random processes X(t) & Y(t).

Three Marks Questions:

1. Differentiate between Random Processes and Random variables with example

2. Prove that the Auto correlation function has maximum value at the origin i.e $|R_{XX}(\tau)| = R_{XX}(0)$

3. A stationary ergodic random processes has the Auto correlation function with the periodic components as $R_{XX}(\tau) = 25 + \frac{4}{1+6\tau^2}$

4. Define mean ergodic random processes and correlation ergodic Random processes.

5. Find the mean value of Response of a linear system.

Ten Marks Questions:

1. a) Define Wide Sense Stationary Process and write it‟s conditions.

b) 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 6+5 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. Explain the following
i) Stationarity
ii) Ergodicity
iii) Statistical independence with respect to random processes

4. a) Given the RP $X(t) = A\cos(w_0 t) + B\sin(w_0 t)$ 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.

b) Define Covariance of the Random processes with any two properties.

5. a) A Gaussian RP has an auto correlation function $R_{XX}(\tau)=\frac{6\sin(\pi\tau)}{\pi\tau}$. Determine a covariance matrix for the Random variable X(t)

b) Derive the expression for cross correlation function between the input and output of a LTI system.

6. Explain about Poisson Random process and also find its mean and variance.

7. 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 var(z).

8. Briefly explain the distribution and density functions in the context of stationary and independent random processes.

9. Explain about the following random process

(i) Mean ergodic process

(ii) Correlation ergodic process

(iii) Gaussian random process

10. State and prove the auto correlation and cross correlation function properties.

UNIT-V

Two marks Questions:

1. Define Power Spectrum Density.

2. Give the statement of Wiener-Khinchin relation.

3. Define spectrum Band width and RMS bandwidth.

4. Write any two properties of Power Spectrum Density.

5. Define linear system.

Three Marks Questions:

1. Show that $S_{XX}(-\omega) = S_{XX}(\omega)$. i.e., Power spectrum density is even function of ω.

2. If the Power spectrum density of x(t) is $S_{XX}(\omega)$, find the PSD of $\frac{d}{dt}x(t)$.

3. If the Auto correlation function of wide sense stationary X(t) is $R_{XX}(\tau)=4+2e^{-2\tau}$. Find the area enclosed by the power spectrum density curve of X(t).

4. Define linear system and derive the expression for output response.

5. If X(t) & Y(t)are uncorrelated and have constant mean values $X \& Y$ then show that $S_{XX}(\omega)= 2\Pi XY\delta(\omega)$

Ten Marks Questions:

1. a)Check the following power spectral density functions are valid or not

$i)\ \frac{cos8(\omega)}{2+\omega^4} \quad ii)\ e^{-(\omega-1)^2}$

b) Derive the relation between input PSD and output PSD of an LTI system

2. Derive the relationship between cross-power spectral density and cross correlation function.

3. A stationery 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)

4. a) 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.

b) Define Power Spectral density with three properties.

5. a) 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

b) State the properties of power spectral density

6. 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.

b) Find the auto correlation function of the random process whose psd is $\frac{16}{\omega^2+4}$

7. a) Find the cross correlation function corresponding to the cross power spectrum

$S_{XY}(\omega)=\frac{6}{(9+\omega^2)(3+j\omega)^2}$

b) Write short notes on cross power density spectrum.

8. a) Consider a random process X(t)=cos$(\omega t+\theta)$where ω is a real constant and $\theta$is a uniform random variable in (0, π/2). Find the average power in the process.

b) Define and derive the expression for average power of Random process.

9. a) The power spectrum density function of a stationary random process is given by

$$S_{XX}(\omega)=\begin{cases} A, & -K<\omega<K \\ 0, & \text{other wise} \end{cases}$$

Find the auto correlation function.

b) Derive the expression for power spectrum density.

10. a) Define and derive the expression for average cross power between two random process X(t) and Y(t).

b) Find the cross power spectral density for $R_{XX}(\tau)=\frac{A^2}{2}\sin(\omega_0\tau)$