LECTURE NOTES

ON

PROBABILITY THEORY AND STOCHASTIC PROCESS

B.Tech III-Sem ECE
IARE-R16

Dr. M V Krishna Rao
(Professor)

Mr. G.Anil kumar reddy
(Assistant professor)

Mrs G.Ajitha
(Assistant professor)

Mr. N Nagaraju
(Assistant professor)

ELECRTONICS AND COMMUNICATION ENGINEERING

INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
DUNDIGAL, HYDERABAD - 500043

UNIT – I
PROBABILITY AND RANDOM VARIABLE

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,

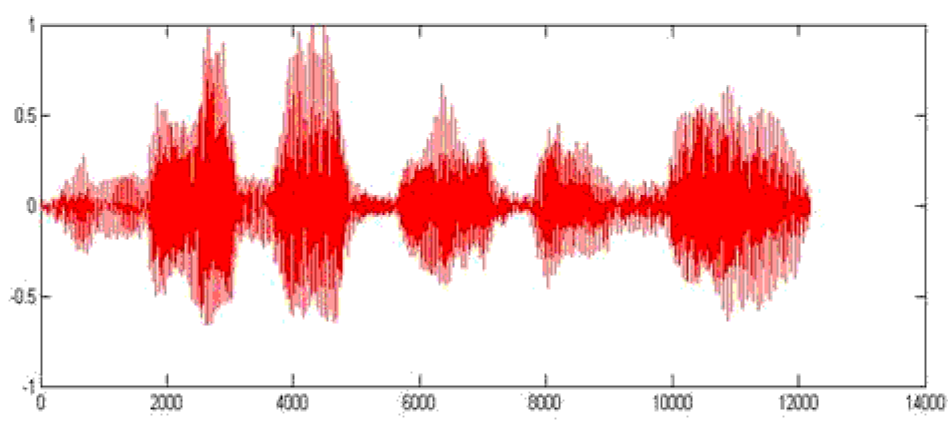
- We can determine the position of a particle moving under a constant force if we know the initial position of the particle and the magnitude and the direction of the force.
- We can determine the current in a circuit consisting of resistance, inductance and capacitance for a known voltage source applying 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,

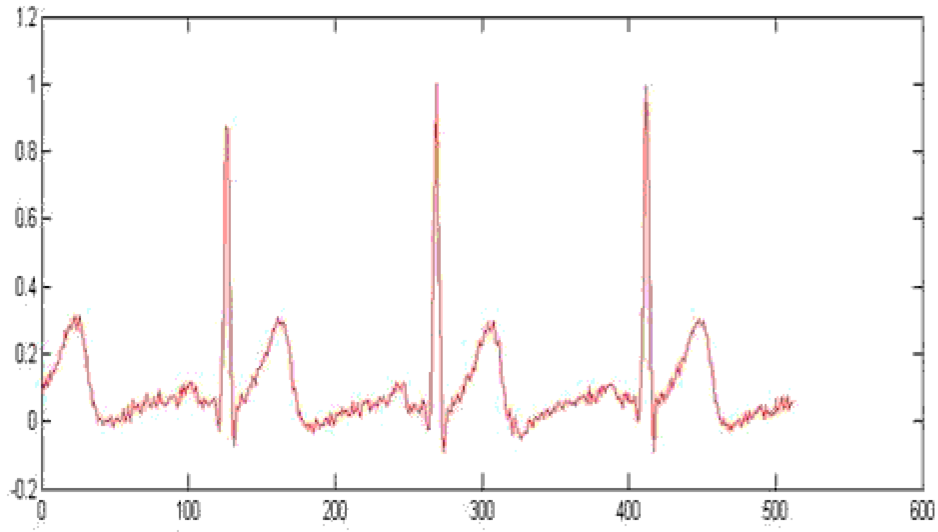
- The outcome of the tossing of a coin cannot be predicted with certainty. Thus the outcome of tossing a coin is random.
- The number of ones and zeros in a packet of binary data arriving through a communication channel cannot be precisely predicted is random.
- The ubiquitous *noise* corrupting the signal during acquisition, storage and transmission can be modelled only through statistical analysis.

Probability in Electrical Engineering

A *signal* is a physical quantity that varyies with time. The physical quantity is converted into the electrical form by means of some transducers . For example, the time-varying electrical voltage that is generated when one speaks through a telephone is a signal. More generally, a signal is a stream of information representing anything from stock prices to the weather data from a remote-sensing satellite**.**



A sample of a speech signal



An *analog signal* $\{x(t), t \in \Gamma\}$ is defined for a continuum of values of domain parameter $t \in \Gamma$ and it can take a continuous range of values.

A *digital signal* $\{x[n], n \in I\}$ is defined at discrete points and also takes a discrete set of values.

As an example, consider the case of an analog-to-digital (AD) converter. The input to the AD converter is an analog signal while the output is a digital signal obtained by taking the samples of the analog signal at periodic intervals of time and approximating the sampled values by a discrete set of values.

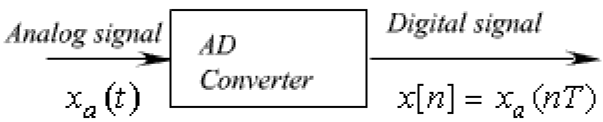


Figure 3 Analog-to-digital (AD) converters

Random Signal

Many of the signals encountered in practice behave randomly in part or as a whole in the sense that they cannot be explicitly described by deterministic mathematical functions such as a sinusoid or an exponential function. Randomness arises because of the random nature of the generation mechanism. Sometimes, limited understanding of the signal dynamics also necessitates the randomness assumption. In electrical engineering we encounter many *signals* that are random in nature. Some examples of random signals are:

i. **Radar signal:** Signals are sent out and get reflected by targets. The reflected signals are received and used to locate the target and target distance from the receiver. The received signals are highly noisy and demand statistical techniques for processing.

ii. **Sonar signal:** Sound signals are sent out and then the echoes generated by some targets are received back. The goal of processing the signal is to estimate the location of the target.

iii. **Speech signal:** A time-varying voltage waveform is produced by the speaker speaking over a microphone of a telephone. This signal can be modeled as a random signal. A sample of the speech signal is shown in Figure 1.

iv. **Biomedical signals:** Signals produced by biomedical measuring devices like ECG, EEG, etc., can display specific behavior of vital organs like heart and brain. Statistical signal processing can predict changes in the waveform patterns of these signals to detect abnormality. A sample of ECG signal is shown in Figure 2.

v. **Communication signals:** The signal received by a communication receiver is generally corrupted by noise. The signal transmitted may the digital data like video or speech and the channel may be electric conductors, optical fiber or the space itself. The signal is modified by the channel and corrupted by unwanted disturbances in different stages, collectively referred to as *noise*.

These signals can be described with the help of probability and other concepts in statistics. Particularly the signal under observation is considered as a realization of a *random process* or a *stochastic process.* The terms *random processes, stochastic processes* and *random signals* are used synonymously.

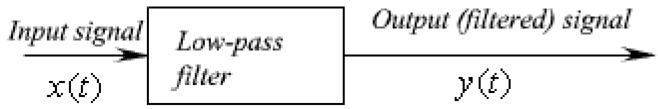
A deterministic signal is analyzed in the frequency-domain through Fourier series and Fourier transforms. We have to know how random signals can be analyzed in the frequency domain.

Random Signal Processing

Processing refers to performing any operations on the signal. The signal can be amplified, integrated, differentiated and rectified. Any noise that corrupts the signal can also be reduced by performing some operations. Signal processing thus involves

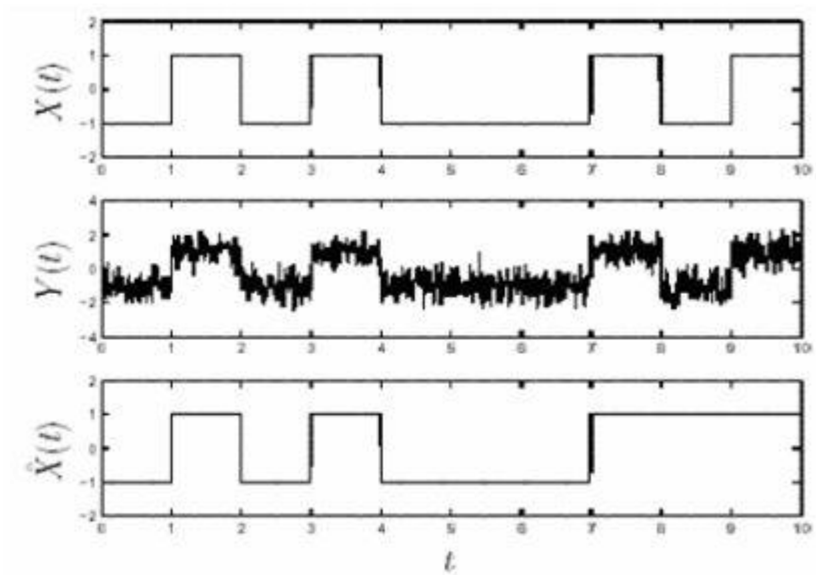
- *Amplification*
- *Filtering*
- *Integration and differentiation*
-
- Nonlinear operations like *rectification, squaring, modulation, demodulation* etc.

These operations are performed by passing the input signal to a *system* that performs the processing. For example, filtering involves selectively emphasising certain frequency components and attenuating others. In *low-pass filtering* illustrated in Fig.4, high-frequency components are attenuated



.

Figure 4 Low-pass filtering



Signal estimation and detection

A problem frequently come across in signal processing is the estimation of the true value of the signal from the received noisy data. Consider the received noisy signal $y(t)$ given by

$$y(t) = A\cos(\omega_0 t) + n(t)$$

where $A\cos(\omega_0 t)$ is the desired transmitted signal buried in the noise $n(t)$.

Simple frequency selective filters cannot be applied here, because random noise cannot be localized to any spectral band and does not have a specific spectral pattern. We have to do this by dissociating the noise from the signal in the probabilistic sense. *Optimal filters* like the *Wiener filter, adaptive filters* and *Kalman filter* deals with this problem.

In estimation, we try to find a value that is close enough to the transmitted signal. The process is explained in Figure 6. *Detection* is a related process that decides the best choice out of a finite number of possible values of the transmitted signal with minimum *error probability*. In binary communication, for example, the receiver has to decide about 'zero' and 'one' on the basis of the received waveform. Signal detection theory, also known as *decision theory*, is based on *hypothesis testing* and other related techniques and widely applied in pattern classification, target detection etc.

Figure 6 Signal estimation problem

Source and Channel Coding

One of the major areas of application of probability theory is *Information theory and coding*. In 1948 Claude Shannon published the paper "*A mathematical theory of communication"* which lays the foundation of modern digital communication. Following are two remarkable results stated in simple languages :

- Digital data is efficiently represented with number of bits for a symbol decided by its *probability of occurrence.*
- The data at a rate smaller than the *channel capacity* can be transmitted over a *noisy* channel with arbitrarily small probability of error. The channel capacity again is determined from the probabilistic descriptions of the signal and the noise.

Basic Concepts of Set Theory

The modern approach to probability based on axiomatically defining probability as function of a set. A background on the set theory is essential for understanding probability.

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.

Probability Concepts

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

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

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*

3. **Event***:* An event A is a subset of the sample space such that probability can be assigned to it. Thus

   - $A \subseteq S$
   - For a discrete sample space, all subsets are events.
   - S is the certain event (sure to occur) and ϕ is the impossible event.

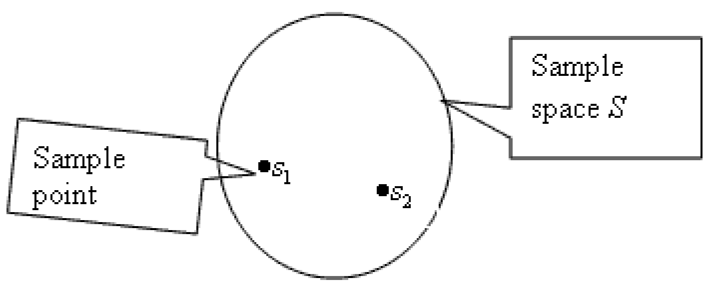


Figure 1

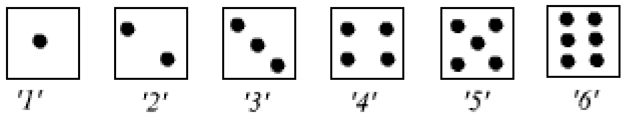
Consider the following examples.

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.

Definition of probability

Consider a random experiment with a finite number of outcomes N If all the outcomes of the experiment are *equally likely* , the probability of an event A is defined by

$$P(A) = \frac{N_A}{N}$$

where

$$N_A = \text{Number of outcomes favourable to } A.$$

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

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

$$\therefore N = 6 \quad \text{and} \quad N_A = 1$$
$$\therefore P(A) = \frac{1}{6}$$

Example 7 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}

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

Discussion

- The classical definition is limited to a random experiment which has only a finite number of outcomes. In many experiments like that in the above examples, the sample space is finite and each outcome may be assumed 'equally likely.' In such cases, the *counting method* can be used to compute probabilities of events.
- Consider the experiment of tossing a fair coin until a 'head' appears.As we have discussed earlier, there are countably infinite outcomes. Can you believe that all these outcomes are equally likely?
- The notion of *equally likely* is important here. *Equally likely* means equally probable. Thus this definition presupposes that all events occur with equal *probability* . Thus the definition includes a concept to be defined

Relative-frequency based definition of probability

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

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

Example 8 Suppose a die is rolled 500 times. The following table shows the frequency each face.

Face	*1*	*2*	*3*	*4*	*5*	*6*
Frequency	*82*	*81*	*88*	*81*	*90*	*78*
Relative frequency	*0.164*	*0.162*	*0.176*	*0.162*	*0.18*	*0.156*

We see that the relative frequencies are close to $\frac{1}{6}$. *How do we ascertain that these relative frequencies will approach to* $\frac{1}{6}$ *as we repeat the experiments infinite no of times?*

Discussion This definition is also inadequate from the theoretical point of view.

- We cannot repeat an experiment infinite number of times.
- How do we ascertain that the above ratio will converge for all possible sequences of outcomes of the experiment?

Axiomatic definition of probability

We have earlier defined an event as a subset of the sample space. *Does each subset of the sample space forms an event?*
The answer is *yes* for a finite sample space. However, we may not be able to assign probability meaningfully to all the subsets of a continuous sample space. We have to eliminate those subsets. The concept of the *sigma algebra* is meaningful now.

Definition Let S be a sample space and $\mathbb{F}$ a sigma field defined over it. Let $P:\mathbb{F}\rightarrow\mathbb{R}$ be a mapping from the sigma-algebra $\mathbb{F}$ into the real line such that for each $A\in\mathbb{F}$, there exists a unique $P(A)\in\mathbb{R}$. Clearly P is a set function and is called probability, if it satisfies the following three axioms.

1. $P(A) \geq 0$ for all $A\in\mathbb{F}$
2. $P(S) = 1$
3. Countable additivity If $A_1, A_2, \ldots$ are pair-wise disjoint events, i.e. $A_i \cap A_j = \phi$ for $i \neq j$, then

   $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

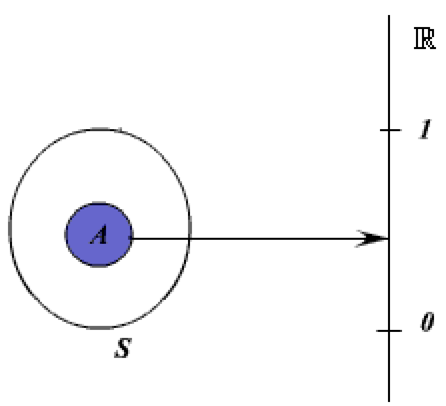


Figure 2

Discussion

- The triplet $(S,\mathbb{F},P)$ is called the probability space.
- Any assignment of probability assignment must satisfy the above three axioms

- If $if\ A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$

  This is a special case of axiom 3 and for a *discrete sample space* , this simpler version may be considered as the axiom 3. We shall give a proof of this result below.

- The events *A* and *B* are called *mutually exclusive* $if\ A \cap B = \emptyset$.

Basic results of probability

From the above axioms we established the following basic results:

1. $P(\phi) = 0$

   Suppose, $A_1 = \phi, A_2 = \phi, \ldots, A_n = \phi, \ldots$

   Then $A_i \cap A_j = \phi \text{ for } i \neq j$

   $$P(\phi) = P(\bigcup_{i=1}^{\infty} A_i)$$
   $$= \sum_{i=1}^{\infty} P(A_i)$$
   $$= \sum_{i=1}^{\infty} P(\phi)$$

   Therefore

   Thus $P(\phi) = \sum_{i=1}^{\infty} P(\phi)$ which is possible only if $P(\phi) = 0$

2. If $A, B \in \mathbb{F} \text{ and } A \cap B = \emptyset,\ P(A \cup B) = P(A) + P(B)$

   We have ,

   $$A \cup B = A \cup B \cup \phi \ldots \cup \phi ..$$
   $$\therefore P(A \cup B) = P(A) + P(B) + P(\phi) \ldots + P(\phi) + .. \text{ (using axiom 3)}$$
   $$\therefore P(A \cup B) = P(A) + P(B)$$

3. $P(A^c) = 1 - P(A)$ where where $A \in \mathbb{F}$

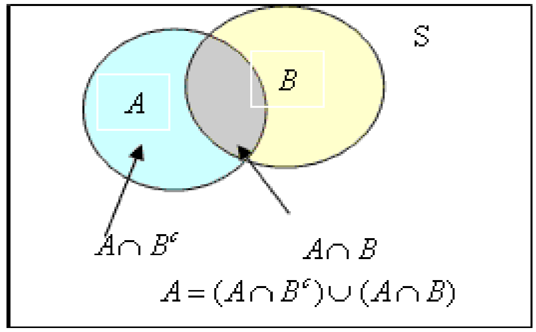
   $$A \cup A^c = S$$
   $$\Rightarrow P(A \cup A^c) = P(S)$$
   $$\Rightarrow P(A) + P(A^c) = 1 \quad \because A \cap A^c = \phi$$

   We have, $\therefore P(A) = 1 - P(A^c)$

4. If $A, B \in \mathbb{F},\ P(A \cap B^c) = P(A) - P(A \cap B)$

We have,

$$(A \cap B^c) \cup (A \cap B) = A$$
$$\therefore P[(A \cap B^c) \cup (A \cap B)] = P(A)$$
$$\Rightarrow P(A \cap B^c) + P(A \cap B) = P(A)$$
$$\Rightarrow P(A \cap B^c) = P(A) - P(A \cap B)$$



We can similarly show that ,

$$P(A^c \cap B) = P(B) - P(A \cap B)$$

5. If $A, B \in \mathbb{F},\ P(AUB) = P(A) + P(B) - P(A \cap B)$
   We have ,

$$A \cup B = (A^c \cap B) \cup (A \cap B) \cup (A \cap B^c)$$
$$\therefore P(A \cup B) = P[(A^c \cap B) \cup (A \cap B) \cup (A \cap B^c)]$$
$$= P(A^c \cap B) + P(A \cap B) + P(A \cap B^c)$$
$$= P(B) - P(A \cap B) + P(A \cap B) + P(A) - P(A \cap B)$$
$$= P(B) + P(A) - P(A \cap B)$$

6. We can apply the properties of sets to establish the following result for $A, B, C \in \mathbb{F}$,

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

The following generalization is known as the principle inclusion-exclusion.

Probability assignment in a discrete sample space

Consider a finite sample space . Then the sigma algebra is defined by the power set of S. For any elementary event , we can assign a probability P(si) such that,

$$\sum_{i=1}^{N} P\left(\{s_i\}\right) = 1$$

For any event $A \in \mathbb{F}$, we can define the probability

$$P(A) = \sum_{A_i \in A} P\left(A_i\right)$$

In a special case, when the outcomes are equi-probable, we can assign equal probability *p* to each elementary event.

$$\therefore \quad \sum_{i=1}^{n} p = 1$$

$$\Rightarrow \quad p = 1/n$$

$$\therefore \quad P(A) = P\left(\bigcup_{s_i \in A} \{A_i\}\right)$$

$$= n(A)\,\frac{1}{n}$$

$$= \frac{n(A)}{n}$$

Example 9 Consider the experiment of rolling a fair die considered in example 2.

Suppose $A_i, i = 1,..,6$ represent the elementary events. Thus A_1 is the event of getting '1', A_2 is the event of getting '2' and so on.

Since all six disjoint events are equiprobable and $S = A_1 \cup A_2 \cup .... \cup A_6$ we get ,

$$P(A_1) = P(A_2) = ... = P(A_6) = \frac{1}{6}$$

Suppose A is the event of getting an odd face. Then

$$A = A_1 \cup A_3 \cup A_5$$

$$\therefore P(A) = P(A_1) + P(A_3) + P(A_5) = 3 \times \frac{1}{6} = \frac{1}{2}$$

Example 10 Consider the experiment of tossing a fair coin until a head is obtained discussed in Example 3. Here $S = \{H,\ TH, TTH, ...\}$. Let us call

$$s_1 = H$$
$$s_2 = TH$$
$$s_3 = TTH$$

and so on. If we assign, $P(\{s_n\}) = \frac{1}{2^n}$ then $\sum_{s_n \in S} P(\{s_n\}) = 1.$ Let $A = \{s_1, s_2, s_3\}$ is the event of obtaining the head before the 4 th toss. Then

$$P(A) = P(\{s_1\}) + P(\{s_2\}) + P(\{s_3\})$$
$$= \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} = \frac{7}{8}$$

Probability assignment in a continuous space

Suppose the sample space *S* is continuous and un-countable. Such a sample space arises when the outcomes of an experiment are numbers. For example, such sample space occurs when the experiment consists in measuring the voltage, the current or the resistance. In such a case, the sigma algebra consists of the Borel sets on the real line.

$S = \mathbb{R}$ $\quad f : \mathbb{R} \to \mathbb{R}$

Suppose and is a non-negative integrable function such that,

$$\int_{\mathbb{R}} f(x)\, dx = 1$$

A

$$P(A) = \int_A f(x)\, dx \,,$$

defines the probability on the Borel sigma-a $\mathbb{R}^2, \mathbb{R}^3$ B .

We can similarly define probability on the continuous space of etc.

Example 11 Suppose

$$f_X(x) = \begin{cases} \dfrac{1}{b - a} & \text{for } x \in [a,\ b] \\ 0 & \text{otherwise} \end{cases}$$

Then for $[a_1,\ b_1] \subseteq [a,\ b]$

$$P([a_1,\ b_1]) = \frac{b_1 - a_1}{b - a}$$

Probability Using Counting Method

In many applications we have to deal with a finite sample space S and the elementary events formed by single elements of the set may be assumed equiprobable. In this case, we can define the probability of the event *A* according to the classical definition discussed earlier:

$$P(A) = \frac{n(A)}{n}$$

where n_A = number of elements favorable to *A* and *n* is the total number of elements in the sample space S.

Thus calculation of probability involves finding the number of elements in the sample space S and the event *A*. Combinatorial rules give us quick algebraic formulae to find the elements in S.We briefly outline some of these rules:

1. **Product rule** Suppose we have a set *A* with *m* distinct elements and the set *B* with *n* distinct elements and $A \times B = \{(a_i, b_j) \mid a_i \in A, b_j \in B\}$. Then $A \times B$ contains *mn* ordered pair of elements. This is illustrated in Fig for *m=5* and *n=4* n other words if we can choose element *a* in *m* possible ways and the element *b* in *n* possible ways then the ordered pair (a, *b)* can be chosen in *mn* possible ways.

B

a_1, b_4	a_2, b_4	a_3, b_4	a_4, b_4	a_5, b_4
a_1, b_3	a_2, b_3	a_3, b_3	a_4, b_3	a_5, b_3
a_1, b_2	a_2, b_2	a_3, b_2	a_4, b_2	a_5, b_2
a_1, b_1	a_2, b_1	a_3, b_1	a_4, b_1	a_5, b_1

A

Figure 1 Illustration of the product rule

The above result can be generalized as follows:

The number of distinct *k* -tupples in $A_1 \times A_2 \times \ldots \times A_k = \{(a_1, a_2, \ldots, a_k) \mid a_1 \in A_1, a_2 \in A_2, \ldots\ldots\ldots, a_k \in A_k\}$ is $n_1 n_2 \ldots\ldots\ldots n_k$ where n_i represents the number of distinct elements in A_i.

Example 1 A fair die is thrown twice. What is the probability that a 3 will appear at least once.

Solution: The sample space corresponding to two throws of the die is illustrated in the following table. Clearly, the sample space has $6 \times 6 = 36$ elements by the product rule. The event corresponding to getting at least one 3 is highlighted and contains 11 elements. Therefore, the required probability is $\frac{11}{36}$.

	Throw 1					
Throw 2	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)
	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)

Throw 1

Example 2 ***Birthday problem*** - Given a class of students, what is the probability of two students in the class having the same birthday? Plot this probability vs. number of students and be surprised!.

Let $k \leq 365$ be the number of students in the class.

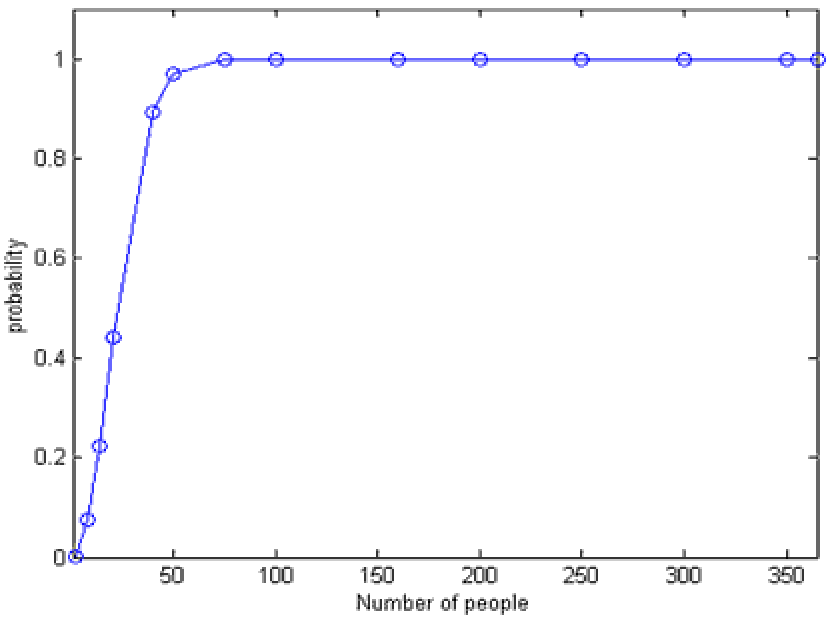
Then the number of possible birth days=365.365....365 (k-times) = 365^k

The number of cases with each of the k students having a different birth day is $= {}^{365}P_k = 365\,.364.....(365-k+1)$

Therefore, the probability of common birthday $= 1 - \frac{{}^{365}P_k}{365^k}$

Number of persons	*Probability*
2	*0.0027*
10	*0.1169*
15	*0.4114*
25	*0.5687*
40	*0.8912*
50	*0.9704*
60	*0.9941*
80	*0.9999*
100	

The plot of **probability vs number of students** is shown in above table. Observe the steep rise in the probability in the beginning. In fact this probability for a group of 25 students is greater than 0.5 and that for 60 students onward is closed to 1. This probability for 366 or more number of students is exactly one.



Example 3 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?

Solution :

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

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

Example 4 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

Consider the probability space $(S, \mathbb{F}, P)$. Let A and B two events in $\mathbb{F}$. We ask the following question –
Given that A has occurred, what is the probability of B?

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

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

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

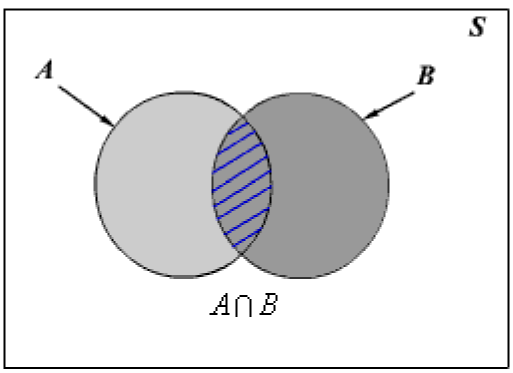


Figure 1

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

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

Clearly ,

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

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

We can similarly define the *conditional probability of A given B* , denoted by $P(A/B)$.

From the definition of conditional probability, we have the joint probability $P(A \cap B)$ of two events *A* and *B* as follows

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

Example 1 Consider the example tossing the fair die. Suppose

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

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

Clearly,

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

Conditional probability and the axioms of probability

In the following we show that the conditional probability satisfies the axioms of probability.

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

Axiom 1:

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

Axiom 2 :

We have , $S \cap A = A$

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

Axiom 3 :

Consider a sequence of disjoint events $B_1, B_2, \ldots, B_n \cdots$.

We have , $(\bigcup_{i=1}^{\infty} B_i) \cap A = \bigcup_{i=1}^{\infty} (B_i \cap A)$

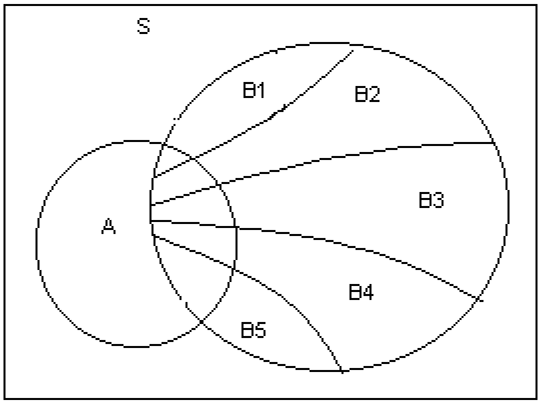


Figure 2

Note that the sequence $B_i \cap A, \quad i = 1, 2, \ldots$ is also sequence of disjoint events.

$$\therefore P(\bigcup_{i=1}^{\infty}(B_i \cap A)) = \sum_{i=1}^{\infty} P(B_i \cap A)$$

$$\therefore P(\bigcup_{i=1}^{\infty} B_i / A) = \frac{P(\bigcup_{i=1}^{\infty} B_i \cap A)}{P(A)} = \frac{\sum_{i=1}^{\infty} P(B_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} P(B_i / A)$$

Properties of Conditional Probabilities

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

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

Chain Rule of Probability

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

We have ,

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

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

We can generalize the above to get the *chain rule of probability*

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

Joint Probability

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

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 | 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∩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.

Total Probability

Let $A_1, A_2, \ldots, A_n$ be *n* events such that $S = A_1 \cup A_2 \ldots\ldots \cup A_n$ 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 $\bigcup_{i=1}^{n} B \cap A_i = B$ and the sequence $B \cap A_i$ is disjoint.

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

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

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

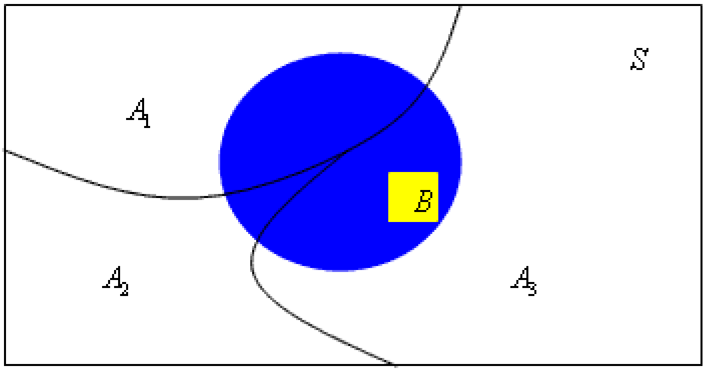


Figure 3

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 $i \neq j$.

(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**.**

Example 3 Suppose a box contains 2 white and 3 black balls. Two balls are picked at random without replacement.

Let A_1 = event that the first ball is white and

Let A_1^c = event that the first ball is black.

Clearly A_1 and A_1^c form a partition of the sample space corresponding to picking two balls from the box.

Let B = the event that the second ball is white. Then .

$$\begin{aligned} P(B) &= P(A_1)P(B/A_1) + P(A_1^c)P(B/A_1^c) \\ &= \frac{2}{5}\times\frac{1}{4} + \frac{3}{5}\times\frac{2}{4} = \frac{2}{5} \end{aligned}$$

Bayes' Theorem

Suppose $A_1, A_2, \ldots A_n$ *are partitions on S such that* $S = A_1 \cup A_2 \ldots \cup A_n$ 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 $P(B) = \sum_{i=1}^{n} P(A_i) P(B \mid A_i)$ (Using the theorem of total probability)

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

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

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.

Example 6

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.

Let *S* be 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'

$$= P(T_1)P(R_1/T_1) + P(T_0)P(R_1/T_0)$$
$$= 0.4\times 0.92 + 0.6\times 0.1$$
$$= 0.448$$

(ii) Using the Baye's rule

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

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 $P(A_3/D)$.

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

The conditional probabilities are $P(D/A_1) = 0.01,\ P(D/A_2) = 0.015$ and $P(D/A_3) = 0.02$

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

and

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

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 4 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.

Example 5 Consider the experiment of picking two balls at random discussed in above example

In this case, $P(B) = \frac{2}{5}$ and $P(B / A_1) = \frac{1}{4}$.

Therefore, $P(B) \neq P(B / A_1)$ and A_1 and *B* are dependent.

RANDOM VARIABLE

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.

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

Consider the probability space $(S, \mathbb{F}, P)$ and function $X : S \to \mathbb{R}$ mapping the sample space S

into the real line. Let us define the probability of a subset $B \subseteq \mathbb{R}$ by

$$P_X(\{B\}) = P(X^{-1}(B)) = P(\{s \mid X(s) \in B\})$$

Such a definition will be valid if $(X^{-1}(B))$ is a valid event. If S is a discrete sample space, $(X^{-1}(B))$ is always a valid event, but the same may not be true if S is infinite. The concept of sigma algebra is again necessary to overcome this difficulty. We also need the Borel sigma algebra $\mathbb{B}$ -the sigma algebra defined on the real line.

The function $X : S \to \mathbb{R}$ is called a *random variable* if the inverse image of all Borel sets under X is an event. Thus, if X is a random variable, then

$$X^{-1}(\mathbb{B}) = \{s \mid X(s) \in \mathbb{B}\} \in \mathbb{F}$$

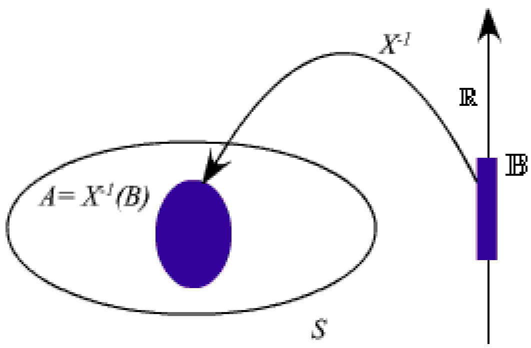


Figure: Random Variable

Observations:

- S is the domain of X.
- The *range* of X denoted by R_X ,is given by

$$R_X = \{X(s) \mid s \in S\}$$

Clearly $R_X \subseteq \mathbb{R}$.

- The above definition of the random variable requires that the mapping X is such that $(X^{-1}(B))$ is a valid event in S. If S is a discrete sample space, this requirement is met by any mapping $X : S \to \mathbb{R}$. Thus *any mapping defined on the discrete sample space is a random variable.*

Example 2 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 3 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 $X = \{1,2,3,4,5,6\}$.

.

Discrete, Continuous and Mixed-type Random Variables

- 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.

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

- *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*,

$$F_X(x) = pF_D(x) + (1-p)F_C(x)$$

where $F_D(x)$ is the distribution function of a discrete RV, $F_C(x)$ is the distribution function of a continuous RV and *o< p <1.*

Typical plots of $F_X(x)$ for discrete, continuous and mixed-random variables are shown in Figure 1, Figure 2 and Figure 3 respectively.

The interpretation of $F_D(x)$ and $F_C(x)$ will be given later.

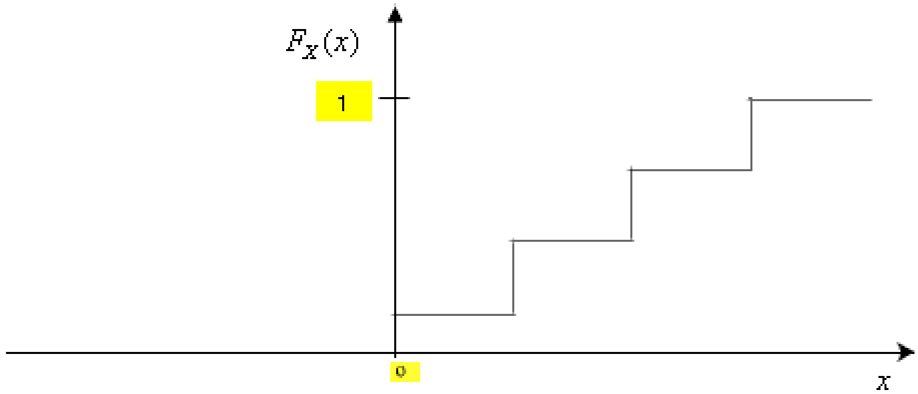
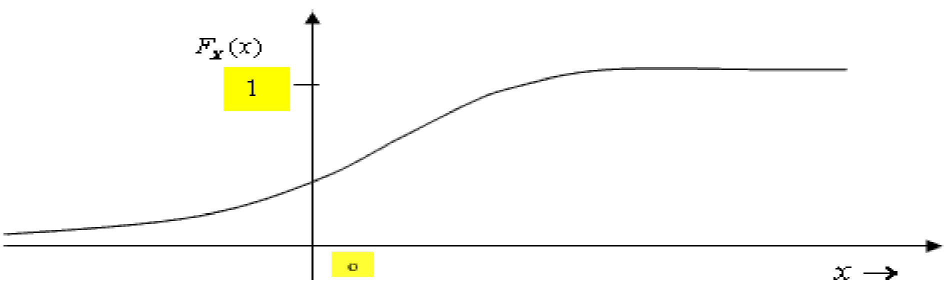
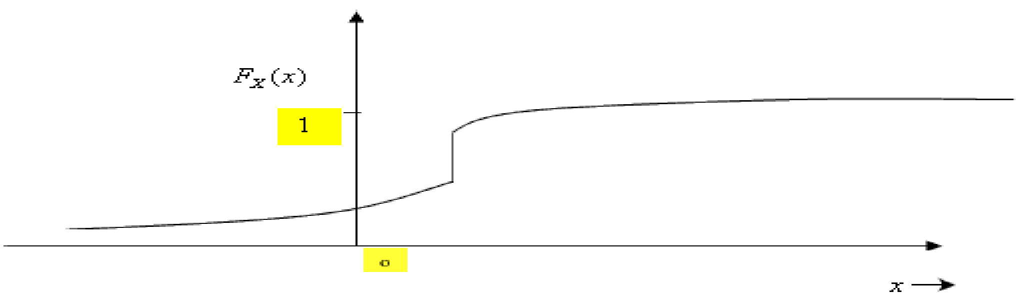


Figure 1 Plot of $F_X(x)$ vs. x for a discrete random variable





UNIT – II
DISTRIBUTION AND DENSITY FUNCTIONS

We have seen that the event B and $\{s \mid X(s) \in B\}$ are equivalent and $P_X(\{B\}) = P(\{s \mid X(s) \in B\})$.The underlying sample space is omitted in notation and we simply write $\{X \in B\}$ and $P(\{X \in B\})$ instead of $\{s \mid X(s) \in B\}$ and $P(\{s \mid X(s) \in B\})$ respectively.

Consider the Borel set $(-\infty,\ x]$, where x represents any real number. The equivalent event $X^{-1}((-\infty,\ x]) = \{s \mid X(s) \le x,\ s \in S\}$ is denoted as $\{X \le x\}$.The event $\{X \le x\}$ can be taken as a representative event in studying the probability description of a random variable X. Any other event can be represented in terms of this event. For example,

$$\{X > x\} = \{X \le x\}^c, \{x_1 < X \le x_2\} = \{X \le x_2\} \setminus \{X \le x_1\},$$

$$\{X = x\} = \bigcap_{n=1}^{\infty}\left(\{X \le x\} \setminus \{X \le x - \frac{1}{n}\}\right)$$

and so on.

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

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

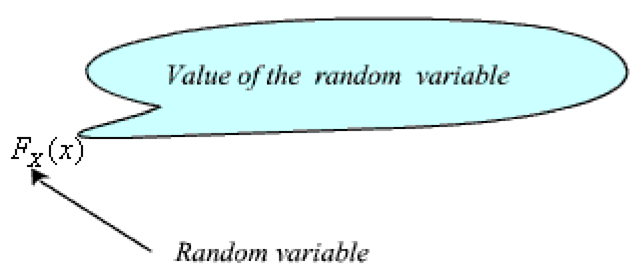


Figure 4

Example 4: 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 $x \ge 3$,

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

Figure 5 shows the plot of $F_X(x)$

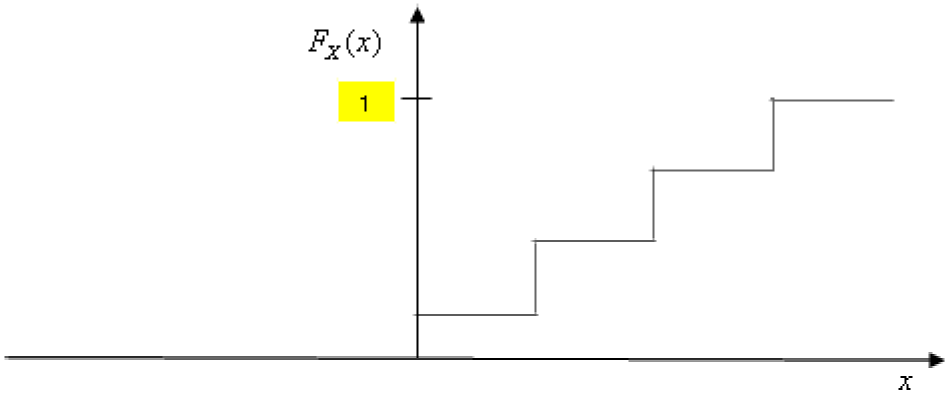


Figure 5

For $1 \le x < 2$,

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

For $2 \le x < 3$,

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

Properties of the Distribution Function

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

  This follows from the fact that $F_X(x)$ is a probability and its value should lie between 0 and 1.

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

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

- $F_X(x)$ Is right continuous.

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

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

- $F_X(-\infty) = 0$

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

- $F_X(\infty) = 1$

$$\text{Because, } F_X(\infty) = P\{s \mid X(s) \le \infty\} = P(S) = 1.$$

- $P(\{x_1 < X \le x_2\}) = F_X(x_2) - F_X(x_1)$

We have ,

$$\{X \le x_2\} = \{X \le x_1\} \cup \{x_1 < X \le x_2\}$$
$$\therefore P(\{X \le x_2\}) = P(\{X \le x_1\}) + P(\{x_1 < X \le x_2\})$$
$$\Rightarrow P(\{x_1 < X \le x_2\}) = P(\{X \le x_2\}) - P(\{X \le x_1\}) = F_X(x_2) - F_X(x_1)$$

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

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

We can further establish the following results on probability of events on the real line:

$$P\{x_1 \le X \le x_2\} = F_X(x_2) - F_X(x_1) + P(X = x_1)$$

$$P(\{x_1 \le X < x_2\}) = F_X(x_2) - F_X(x_1) + P(X = x_1) - P(X = x_2)$$

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

$$F_X(x), \quad -\infty < x < \infty$$

Thus we have seen that given X, $F_X(x)\ \forall x \in X$ line the probability of any event involving values of the random variable X .Thus is a complete description of the random variable .

Example 5 Consider the random variable X defined by

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

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

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

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

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

Solution:

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

$$\text{b) } P\{X \le 0\} = F_X(0)$$
$$= 1$$

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

$$\text{d) } P\{-1 < X \leq 1\}$$
$$= F_X(1) - F_X(-1)$$
$$= 1 - \frac{1}{8} = \frac{7}{8}$$

Figure 6 shows the plot of $F_X(x)$.

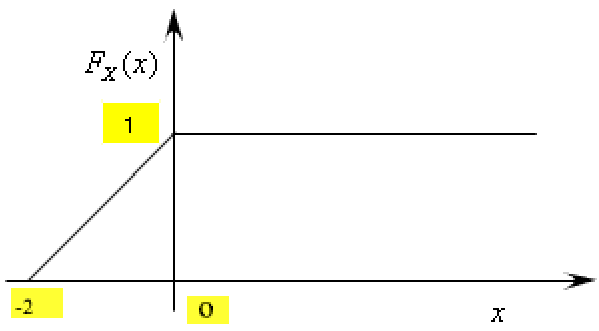


Figure 6

Discrete Random Variables and Probability DENSITY functions

A random variable is said to be *discrete* if the number of elements in the range R_X is finite or countably infinite.

First assume R_X to be countably finite. Let $x_1, x_2, x_3, \ldots, x_N$ be the elements of R_X. Here the mapping $X(s)$ partitions S into N subsets $\{s \mid X(s) = x_i\}, i = 1, 2, \ldots, N$.

The discrete random variable in this case is completely specified by the *probability mass function* (pmf) $p_X(x_i) = P(\{s \mid X(s) = x_i\}, i = 1, 2, \ldots, N)$.

Clearly,

- $p_X(x_i) \geq 0 \quad \forall x_i \in R_X$ and
- $\sum_{i \in R_X} p_X(x_i) = 1$
- $P_X(x_i) = F_X(x_i) - F_X(x_i^-) \quad \text{for all } x_i \in R_x$
- Suppose $D \subseteq R_X$ .Then $P(\{x \in D\}) = \sum_{x_i \in D} p_X(x_i)$

Figure 6 illustrates a discrete random variable.

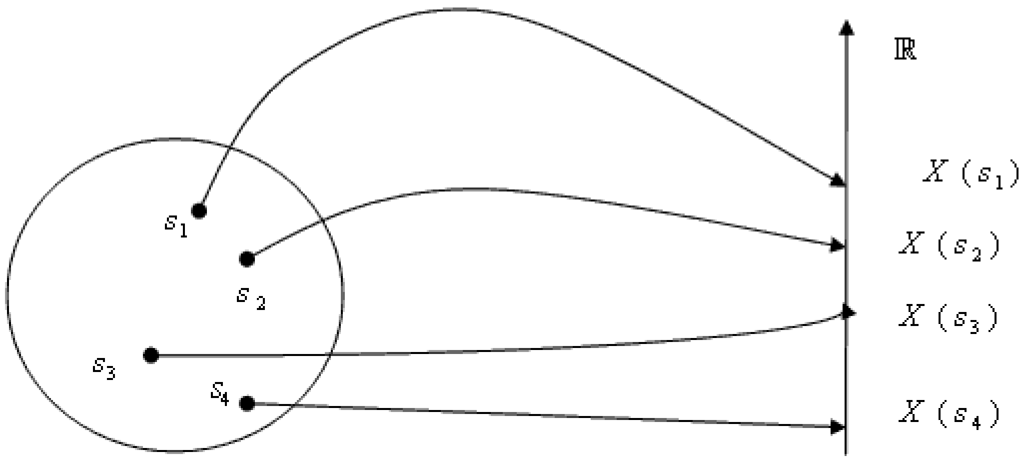


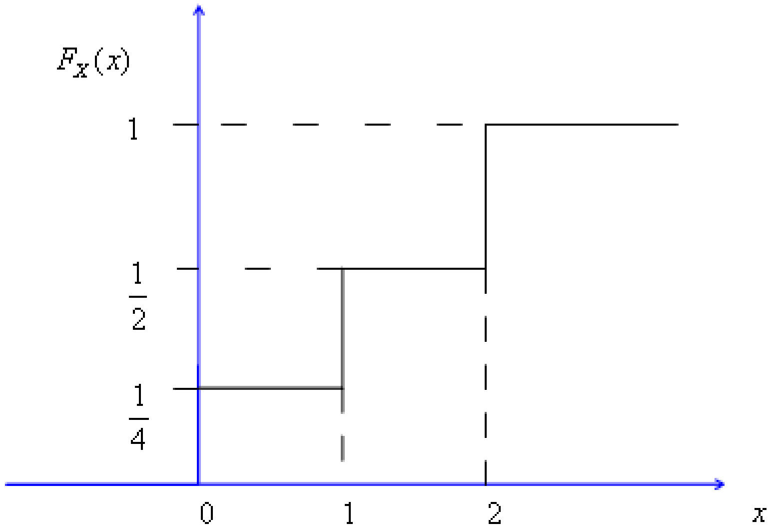
Figure 6 Discrete Random Variable

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	$\frac{1}{2}$

Continous Random Variables and Probability Density Functions

For a continuous random variable X, $F_X(x)$ is continuous everywhere. Therefore,

$$F_X(x) = F_X(x^-) \quad \forall x \in \mathbb{R}$$

This implies that for $x \in \mathbb{R}$,

$$\begin{aligned} p_X(x) &= P(\{X = x\}) \\ &= F_X(x) - F_X(x^-) \\ &= 0 \end{aligned}$$

Therefore, the probability mass function of a continuous RV X is zero for all x .A continuous random variable cannot be characterized by the probability mass function. A continuous random variable has a very important chacterisation in terms of a function called the *probability density function.*

If $F_X(x)$ is differentiable, the probability density function (pdf) of X denoted by $f_X(x)$ is defined as

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

Interpretation of $f_X(x)$

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{P(\{x < X \le x + \Delta x\})}{\Delta x} \end{aligned}$$

so that

$$P(\{x < X \le x + \Delta x\}) = f_X(x)\Delta x$$

Thus the probability of X lying in some interval $(x, x+\Delta x]$ is determined by $f_X(x)$. In that sense, $f_X(x)$ represents the concentration of probability just as the density represents the concentration of mass.

Properties of the Probability Density Function

- $f_X(x) \ge 0$.

  This follows from the fact that $F_X(x)$ is a non-decreasing function

- $F_X(x) = \int_{-\infty}^{x} f_X(u)du$
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- $P(x_1 < X \le x_2) = \int_{-x_1}^{x_2} f_X(x)dx$

Figure 8 below illustrates the probability of an elementary interval in terms of the pdf.

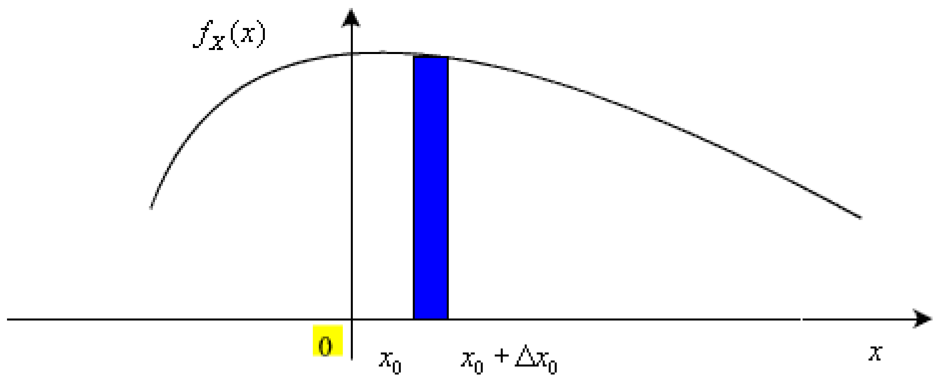


Figure 8 Illustration of $P(\{x_0 < X \le x_0 + \Delta x_0\}) = fX(x_0)\Delta x_0$

Example 2 Consider the random variable X with the distribution function

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-ax}, a > 0 & x \ge 0 \end{cases}$$

The pdf of the RV is given by

$$f_X(x) = \begin{cases} 0 & x < 0 \\ e^{-ax} \; a > 0 \;, & x \ge 0 \end{cases}$$

Remark: Using the Dirac delta function we can define the density function for a discrete random variables.

Consider the random variable X defined by the *probability mass function* (pmf)

$$p_X(x_i) = P(s \mid X(s) = x_i), i = 1, 2, \ldots, n.$$

The distribution function $F_X(x)$ can be written as

$$F_X(x) = \sum_{i=1}^{n} p_X(x_i) u(x - x_i)$$

where $u(x - x_i)$ is the shifted unit-step function given by

$$u(x - x_i) = \begin{cases} 1 & \text{for} \quad x \geq x_i \\ 0 & \text{otherwise} \end{cases}$$

Then the density function $f_X(x)$ can be written in terms of the Dirac delta function as

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

Example 3

Consider the random variable defined with the distribution function $F_X(x)$ given by,

$$F_X(x) = \frac{1}{4} u(x) + \frac{1}{4} u(x-1) + \frac{1}{2} u(x-2)$$

Then

$$f_X(x) = \frac{1}{4} \delta(x) + \frac{1}{4} \delta(x-1) + \frac{1}{2} \delta(x-2)$$

Probability Density Function of Mixed Random Variable

Suppose X is a mixed random variable with $F_X(x)$ having jump discontinuity at $X = x_i, i = 1, 2, \ldots, n$. As already stated, the CDF of a mixed random variable X is given by

$$\boxed{F_X(x) = pF_D(x) + (1-p) F_C(x)}$$

where $F_D(x)$ is a discrete distribution function of X and $F_C(x)$ is a continuous distribution function of X.

The corresponding pdf is given by

$$f_X(x) = pf_D(x) + (1-p)f_C(x)$$

where

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

and $f_C(x)$ is a continuous pdf. We can establish the above relations as follows.

Suppose $R_D = \{x_1, x_2, \ldots, x_n\}$ denotes the countable subset of points on R_X such that the random variable is characterized by the probability mass function $p_X(x), x \in R_D$. Similarly, let $R_C = R_X \setminus R_D$ be a continuous subset of points on R_X such that RV is characterized by the probability density function $f_C(x), x \in R_C$.

Clearly the subsets R_D and R_C partition the set R_X If $P(R_D) = p$, then $P(R_C) = 1 - p$. Thus the probability of the event $\{X \le x\}$ can be expressed as

$$\begin{aligned} P\{X \le x\} &= P(R_D)P(\{X \le x\} \mid R_D) + P(R_C)P(\{X \le x\} \mid R_C) \\ &= pF_D(x) + (1-p)F_C(x) \\ \therefore F_X(x) &= pF_D(x) + (1-p)F_C(x) \end{aligned}$$

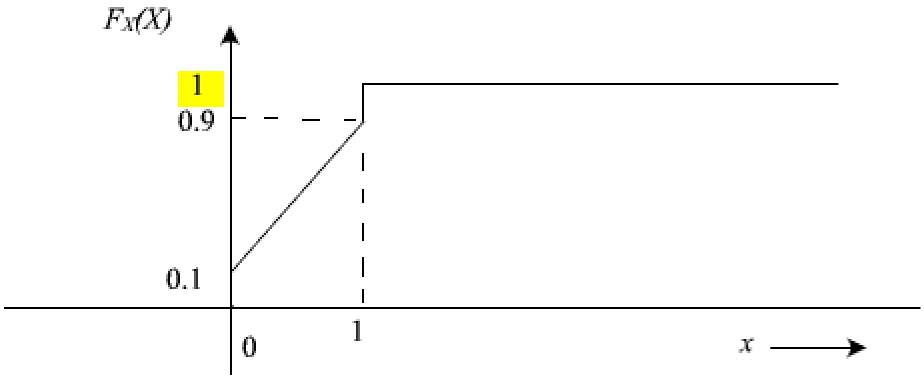
Taking the derivative with respect to *x* , we get

$$f_X(x) = pf_D(x) + (1-p)f_C(x)$$

Example 4 Consider the random variable X with the distribution function

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.1 & x = 0 \\ 0.1 + 0.8x & 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$

The plot of $F_X(x)$ is shown in Figure 9 on next page



where

$$F_D(x) = \begin{cases} 0 & x < 0 \\ 0.5 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

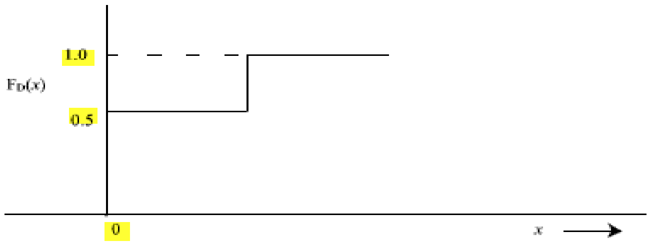


Figure 10

The pdf is given by

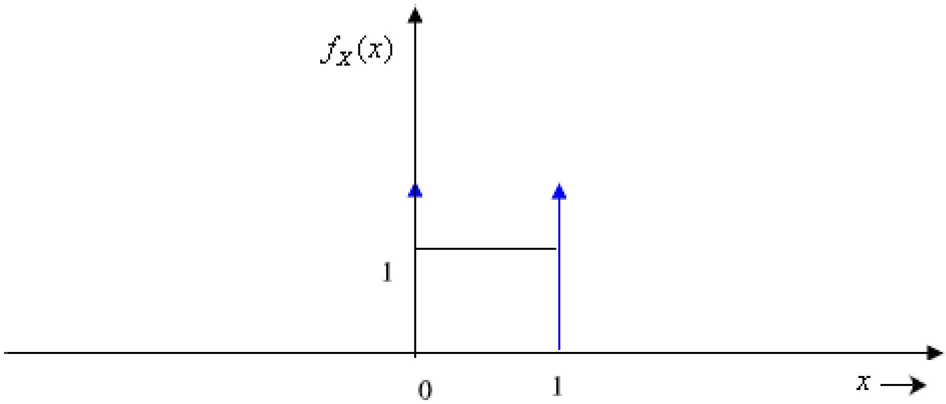
$$f_X(x) = 0.2 f_D(x) + 0.8 f_C(x)$$

where

$$f_D(x) = 0.5\delta(x) + 0.5\delta(x-1)$$

and

$$f_C(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$



Example 5

X is the random variable representing the life time of a device with the PDF $f_X(x)$ for $x \geq 0$. Define the following random variable

$$\begin{aligned} Y &= X \quad \text{if} \quad X \leq a \\ &= a \quad \text{if} \quad X > a \end{aligned}$$

Find $F_Y(y)$.

Solution: $R_D = \{a\}$

$$R_C = (0,\ a)$$

$$\begin{aligned} p &= P\{y \in D\} \\ &= P\{X > a\} \\ &= 1 - F_X(a) \end{aligned}$$

$$\therefore F_Y(y) = pF_D(x) + (1-p)F_C(x)$$

OTHER DISTRIBUTION AND DENSITY RVS

In the following, we shall discuss a few commonly-used discrete random variabes. The importance of **these random variables will be highlighted.**

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

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

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

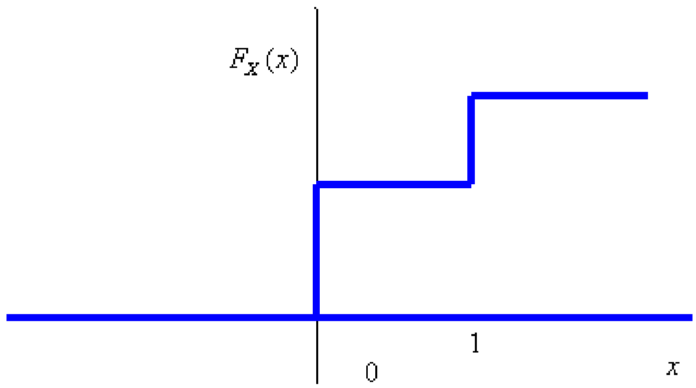


Figure 2

Remark

We can define the pdf of X with the help of Dirac delta function. Thus

$$f_X(x) = (1-p)\delta(x) + p\delta(x)$$

Example 2 Consider the experiment of tossing a *biased* coin. Suppose $P(\{H\}) = p$ and $P(\{T\}) = 1 - p$.

If we define the random variable $X(H) = 1$ and $X(T) = 0$ then X is a Bernoulli random variable.

Mean and variance of the Bernoulli random variable

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

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

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

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.

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

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

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.

Example 3 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 is shown in the Figure 3 below.

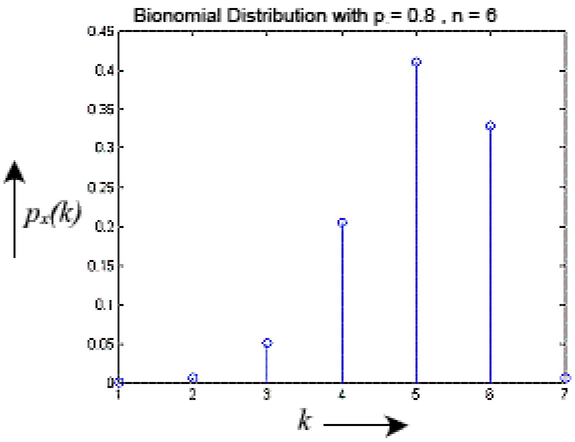


Figure 3

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 $B(n-1, p)$.

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

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

$$= np(1-p)$$

Poisson Random Variable

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

$$p_X(k) = \frac{e^{-\lambda} \lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

The plot of the pmf of the Poisson RV is shown in Figure 2

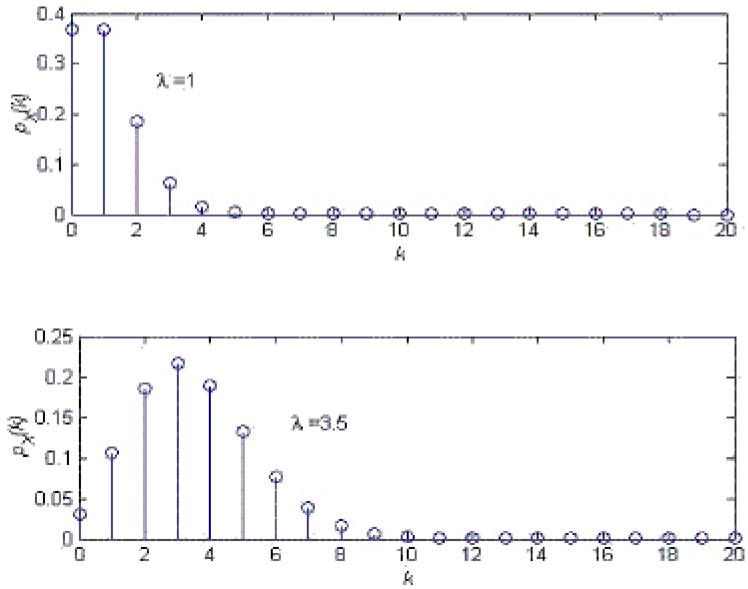


Figure 2

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 3 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

$$p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$$ 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

$$= 1-\sum_{k=0}^{5}p_X(k) = 1-e^{-10}(1+\frac{10}{1}+\frac{10^2}{2!}+\frac{10^3}{3!}) =$$ 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

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

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

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

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

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

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

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

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

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,

$$\begin{aligned}
&\therefore \lambda = 120 \times 0.01 = 0.12 \\
&\text{P(more than one errors)} \\
&\quad = 1 - p_X(0) - p_X(1) \\
&\quad \simeq 1 - e^{-\lambda} - \lambda e^{-\lambda} \\
&\quad = 0.0066
\end{aligned}$$

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

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

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

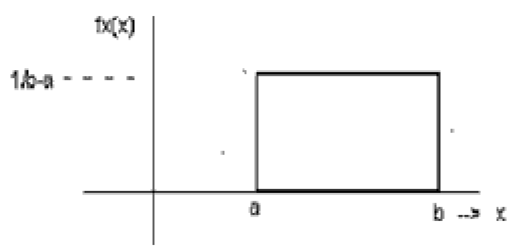


Figure 1

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

[*a*,*b*]. Also note that

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_a^b \frac{1}{b-a}dx = 1$$

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.

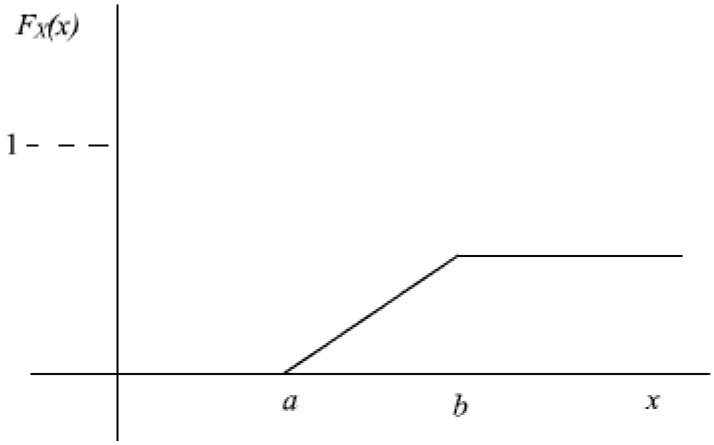


Figure 2

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

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

$$= \frac{e^{jwb} - e^{jwa}}{jw(b-a)}$$

Example 1

Suppose a random noise voltage X across an electronic circuit is uniformly distributed between -4 V and 5 V. What is the probability that the noise voltage will lie between 2 V and 3 V? What is the variance of the voltage?

$$P(\{2 < X \leq 3\}) = \int_2^3 \frac{dx}{5-(-4)} = \frac{1}{9}.$$

$$\sigma_X^2 = \frac{(5+4)^2}{12} = \frac{27}{4} \text{ V}^2.$$

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 modelled as a normal random variable.

A 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 $N\left(\mu_X, \sigma_X^2\right)$ distributed.

If $\mu_X = 0$ and $\sigma_X^2 = 1$,

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

and the random variable X is called the *standard normal variable*.

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

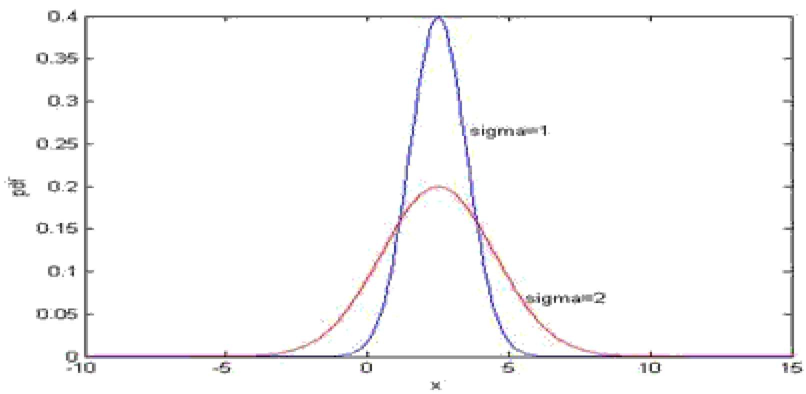


Figure 3

- $f_X(x)$ 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 random variable

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

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

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

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

$$= \Phi\left(\frac{x-\mu_X}{\sigma_X}\right)$$

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

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

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

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

$$= \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{u^2}{2}} du$$

Note that $Q(0) = \frac{1}{2}, \quad Q(-x) = Q(x)$ 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, $erf(x) = \dfrac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$

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

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

Mean and Variance of a Gaussian Random Variable

If X is $N\left(\mu_X, \sigma_X^{\ 2}\right)$ distributed, then

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

Proof:

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

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

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

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

Exponential Random Variable

A continuous random variable X is called exponentially distributed with the parameter $\lambda > 0$ if the probability density function is of the form

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

The corresponding probability distribution function is

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

$$= \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

We have $\mu_X = EX = \int_0^\infty x\lambda e^{-\lambda x}dx$

$$= \frac{1}{\lambda}\int_0^\infty ue^{-u}du \quad (\textit{substituting } u = \lambda x)$$

$$= \frac{1}{\lambda}\overline{|2}$$

$$= \frac{1}{\lambda}$$

Similarly Ex^2

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

$$= \frac{1}{\lambda^2}\int_0^\infty u^2 e^{-u}du$$

$$= \frac{1}{\lambda^2}\overline{|2}$$

$$= \frac{2}{\lambda^2}$$

$$\therefore \sigma_x{}^2 = Ex^2 - \mu x^2$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2}$$

$$= \frac{1}{\lambda^2}$$

Figure 1 shows the typical pdf of an exponential RV.

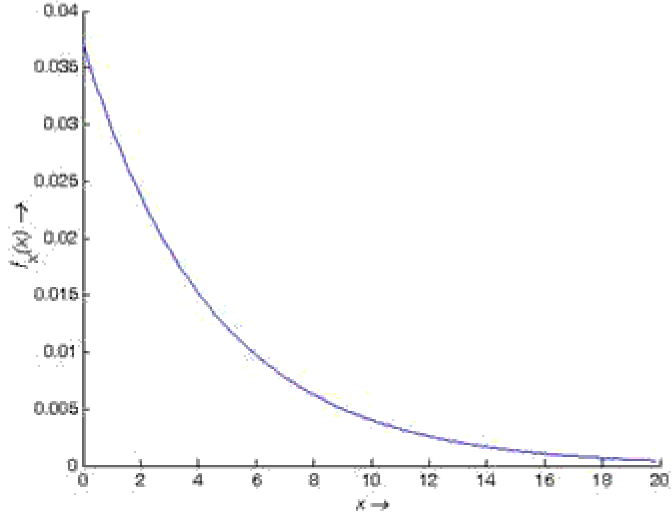


Figure 1

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

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

where σ is the parameter of the random variable.

The probability density functions for the Rayleigh RVs are illustrated in Figure 6.

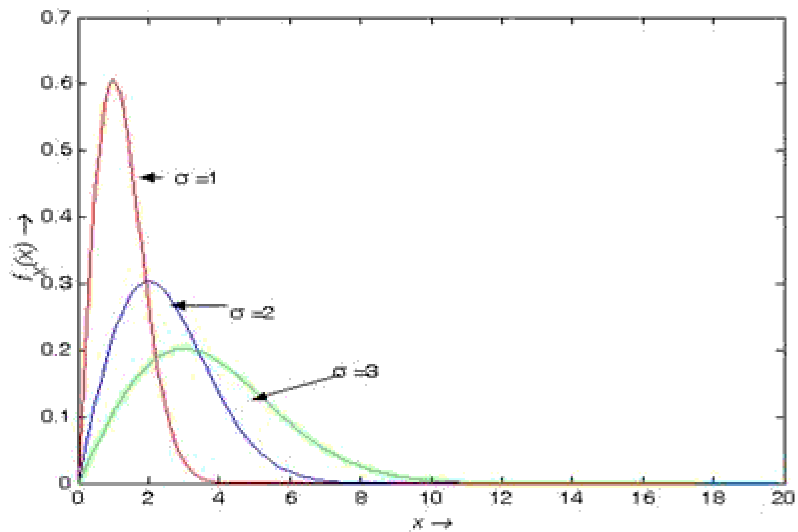


Figure 6

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,

$$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx$$
$$= \int_0^{\infty} x^2 \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} dx$$
$$= 2\sigma^2 \int_0^{\infty} ue^{-u} du \quad (\textbf{Substituting } u = \frac{x^2}{2\sigma^2})$$
$$= 2\sigma^2 \quad (\textbf{Noting that } \int_0^{\infty} ue^{-u}du \textbf{ is the mean of the exponential RV with } \lambda\textbf{=1})$$
$$\therefore \sigma_X^2 = 2\sigma^2 - \left(\sqrt{\frac{\pi}{2}}\sigma\right)^2$$
$$= (2 - \frac{\pi}{2})\sigma^2$$

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

$$F_X(x/B) = P\left[\{X \le x\}/B\right]$$
$$= \frac{P\left[\{X \le x\} \cap B\right]}{P(B)} \qquad P(B) \ne 0$$

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

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

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

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

$$P(\{x_1 < X \le x_2\}/B) = F_X(x_2/B) - F_X(x_1/B)$$
$$= \int_{x_1}^{x_2} f_X(x/B)dx$$

-

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

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

Case 1: $x < b$

Then

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

And

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

Case 2: $x \ge b$

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

and

$$f_X(x/B) = \frac{d}{dx} F_X(x/B) = 0$$

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

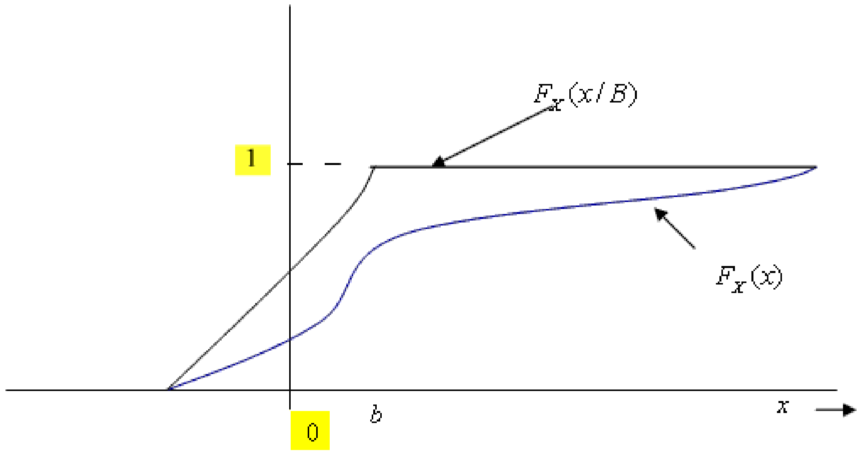
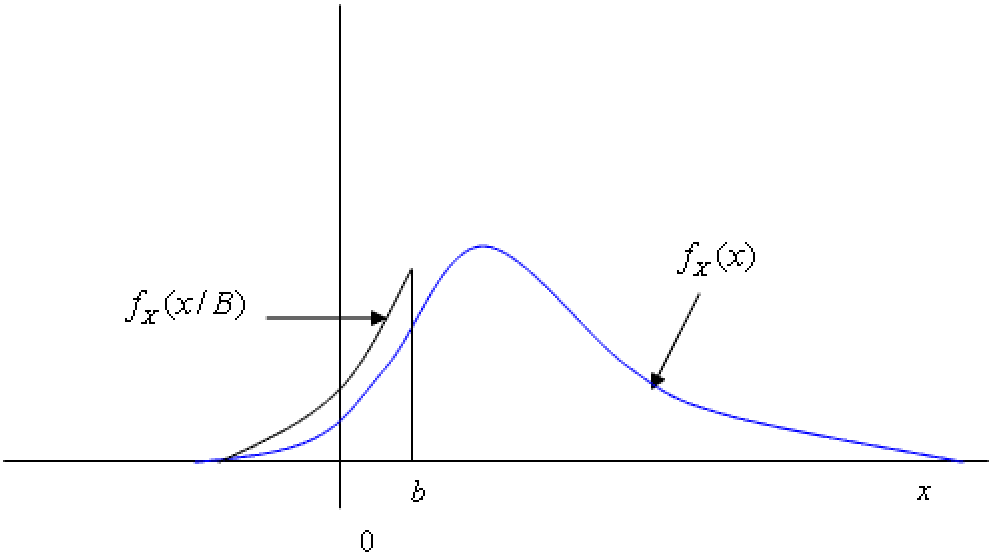


Figure 1



Example 2 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

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

Example 3 Suppose *X* is a random variable with the probability density function

$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $B = \{-1 < X < 1\}$. Then

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

$$= \frac{P(\{X \le x\} \cap \{-1 \le X \le 1\})}{P(\{-1 \le X \le 1\})}$$

$$= \frac{P(\{X \le x\} \cap \{-1 \le X \le 1\})}{\int_{-1}^{1} f_X(x)dx}$$

$$\therefore \quad F_X(x/B) = \begin{cases} 0\,, & x \le -1 \\ \dfrac{F_X(x) - F_X(1)}{\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx}\,, & -1 < x < 1 \\ 1\,, & x \ge 1 \end{cases}$$

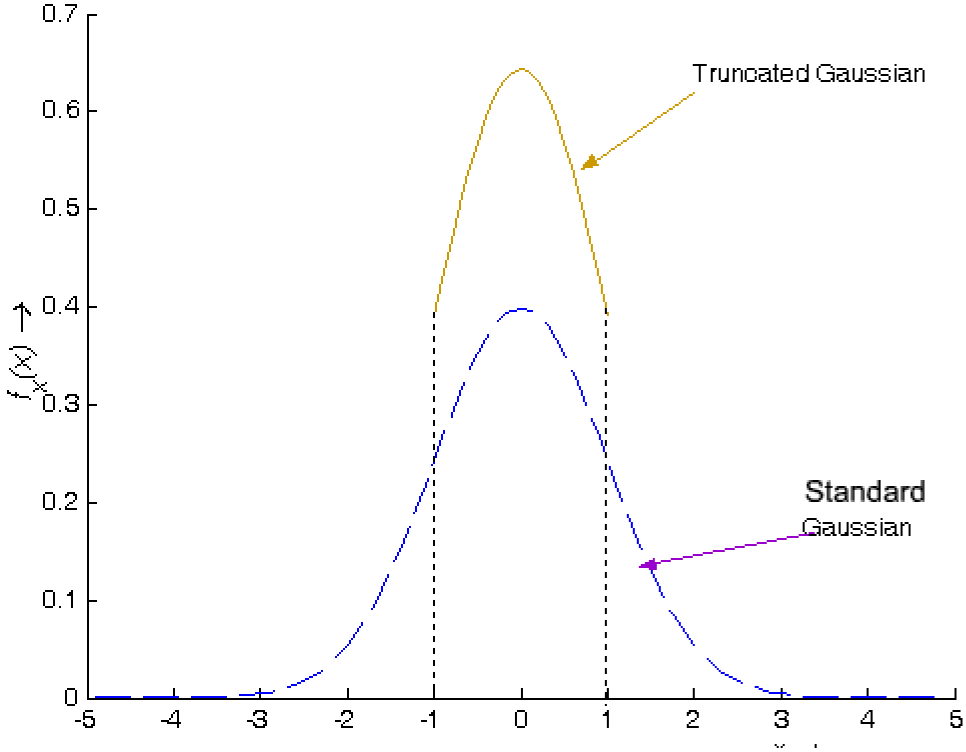
$$\therefore \quad f_X(x/B) = \begin{cases} \dfrac{f_X(x)}{erf(\frac{1}{\sqrt{2}})}\,, & -1 < x < 1 \\ 0\,, & \text{otherwise} \end{cases}$$

where $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$

Remark

$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the *standard Gaussian distribution.*

$f_X(x/B)$ is called the *truncated Gaussian* and plotted in Figure 3 on next page.



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 expection 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 $p_X(x_i), i = 1, 2, \ldots, N,$

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

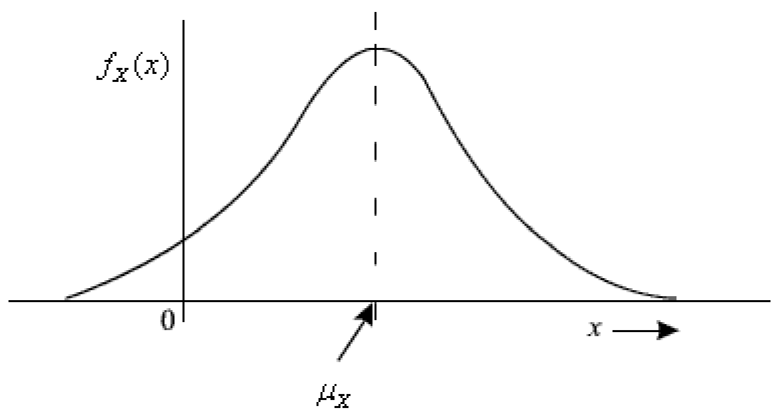


Figure1 Mean of a random variable

Example 1

Suppose X is a random variable defined by the pdf

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

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
$p_X(x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$

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,

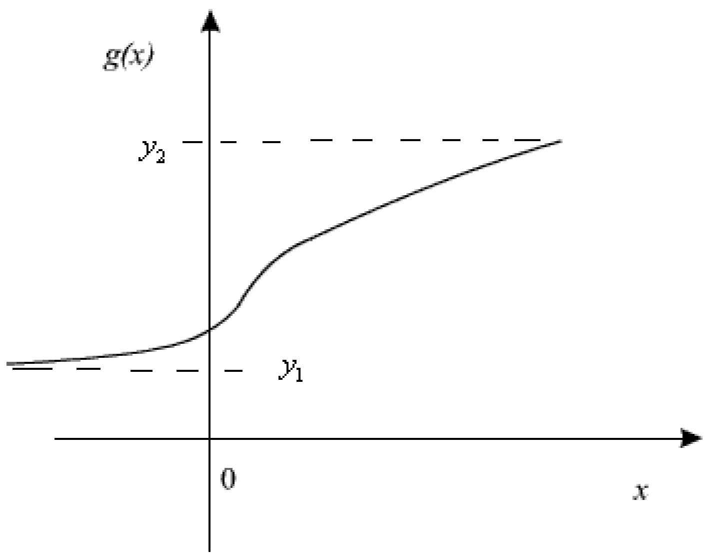


Figure 2

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

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

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

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

(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,

$$E[c_1 g_1(X) + c_2 g_2(X)] = c_1 Eg_1(X) + c_2 Eg_2(X)$$

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

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

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

$$= c_1 Eg_1(X) + c_2 Eg_2(X)$$

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

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

$$= \frac{(b-a)^2}{12}$$

Example 5

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

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

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

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

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

***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.

Inequalities based on expectations

The mean and variance also give some quantitative information about the bounds of RVs. Following inequalities are extremely useful in many practical problems.

Chebychev Inequality

Suppose X a parameter of a manufactured item with known mean μ_X and variance σ_X^2. The quality control department rejects the item if the absolute deviation of X from μ_X is greater than $2\sigma_X$.

The standard deviation gives us an intuitive idea how the random variable is distributed about the mean. This idea is more precisely expressed in the remarkable *Chebysev Inequality* stated below. For a random variable X with mean μ_X and variance σ_X^2.

$$P\{|X-\mu_X| \geq \varepsilon\} \leq \frac{\sigma_X^2}{\varepsilon^2}$$

Proof:

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx$$
$$\geq \int_{|X-\mu_X|\geq\varepsilon} (x-\mu_X)^2 f_X(x)dx$$
$$\geq \int_{|X-\mu_X|\geq\varepsilon} \varepsilon^2 f_X(x)dx$$
$$= \varepsilon^2 P\{|X-\mu_X| \geq \varepsilon\}$$

$\therefore$

$$P\{|X-\mu_X| \geq \varepsilon\} \leq \frac{\sigma_X^2}{\varepsilon^2}$$

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.
- The interpretation that $\phi_X(\omega),$ is the expectation of $e^{j\omega X}$ helps in calculating moments with the help of the characteristics function. In a simple case ,
  if $Y = aX + b$
  $$\phi_y(\omega) = Ee^{ja\omega X + bj\omega}$$
  $$= e^{jb\omega}\phi_x(a\omega)$$
- As $f_X(x)$ always non-negative and $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$, $\phi_X(\omega),$ always exists. 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

$f_X(x) = \frac{1}{b-a} \quad a \le x \le b$ = 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-III
MULTIPLE RANDOM VARIABLES AND OPERATIONS

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 $S \rightarrow \mathbb{R}^2$ such that for $s \in S,\ (X(s), Y(s)) \in \mathbb{R}^2$ is called a joint random variable.

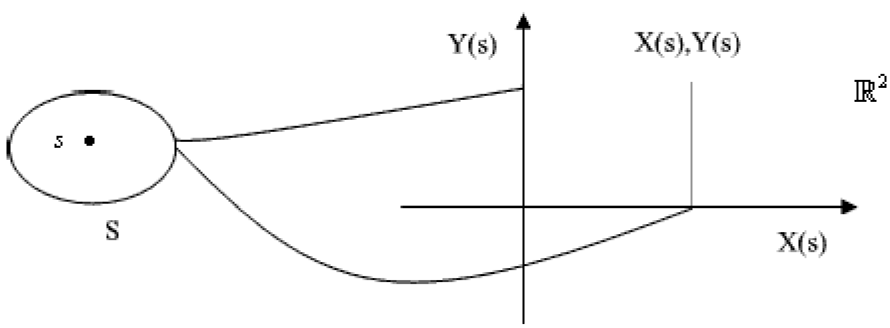


Figure 1

Joint Probability Distribution Function

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

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

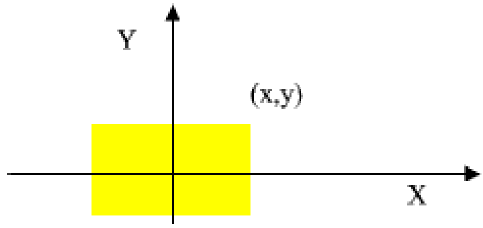


Figure 2

Properties of JPDF

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

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

2) If $x_1 < x_2$ and $y_1 < y_2$,
$\{X \le x_1, Y \le y_1\} \subseteq \{X \le x_2, Y \le y_2\}$
$\therefore P\{X \le x_1, Y \le y_1\} \le P\{X \le x_2, Y \le y_2\}$
$\therefore F_{X,Y}(x_1,y_1) \le F_{X,Y}(x_2,y_2)$

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

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

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 $y_1 < y_2$

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

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*

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

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

Remark

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

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

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

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

This is because

$$=P(S) = 1$$

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

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

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

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.

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	

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

Properties of Joint Probability Density Function

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

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

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

- The probability of any Borel set can be obtained by

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

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

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

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

$$F_{Y/X}(y/x) = P(Y \leq y / X = x)$$

$$= \frac{P(Y \leq y, X = x)}{P(X = x)}$$

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

$$F_{Y/X}(y/x) = lim_{\Delta x \to 0} P(Y \leq y / x < X \leq x + \Delta x)$$

$$= lim_{\Delta x \to 0} \frac{P(Y \leq y, x < X \leq x + \Delta x)}{P(x < X \leq x + \Delta x)}$$

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

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

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

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

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

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

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

The right hand side of the highlighted equation is

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

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

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

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

find,

(a) k

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

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

Solution:

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

We get

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

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

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

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

Independent Random Variables (or) Statistical Independence

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

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

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

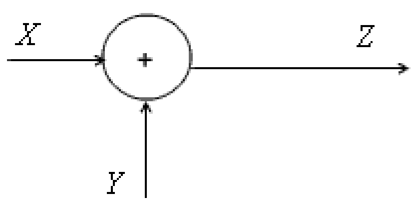
Sum of Two Random Variables

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 $D_Z = \{(x,y) \mid g(x,y) \le z\}$.

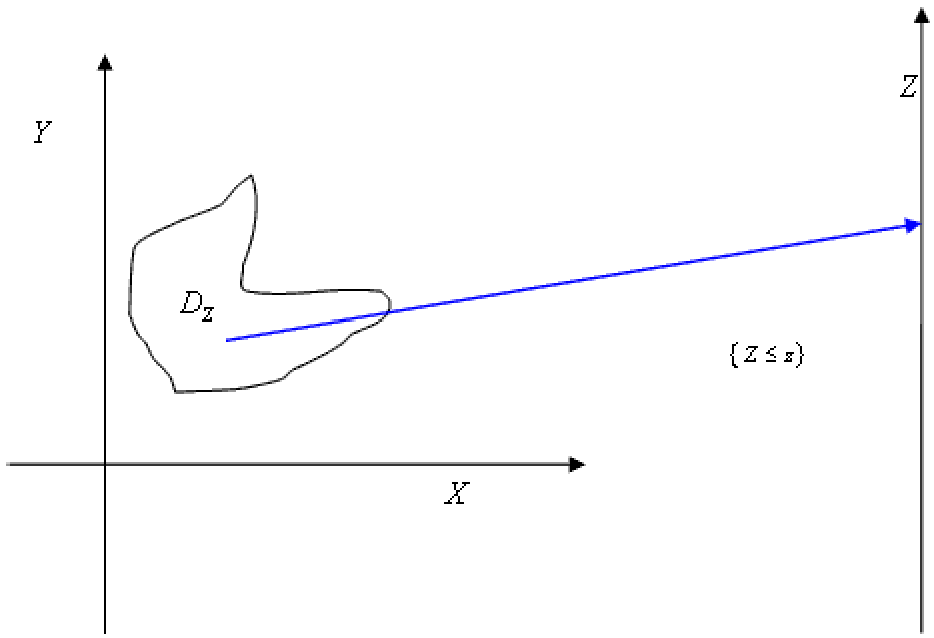


Figure 1

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

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

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

Consider Figure 2

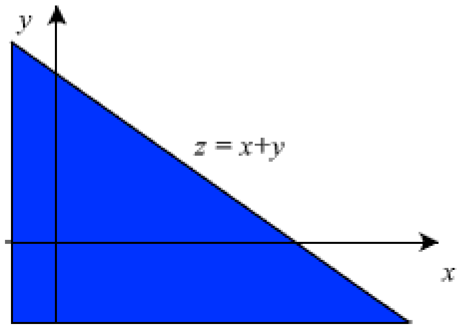


Figure 2

We have

$Z \le z$

$\Rightarrow X + Y \le z$

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

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

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

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

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

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

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

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

If *X* and *Y* are independent

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

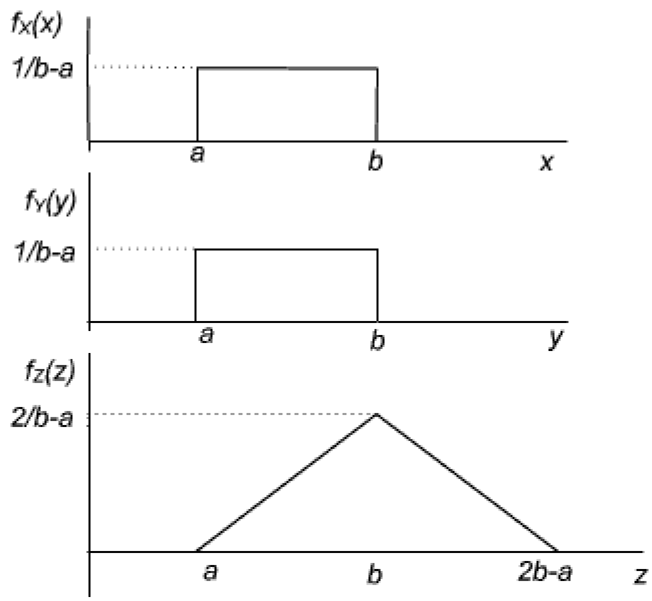
$$\therefore f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx$$

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

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 $Y \sim N(\mu_Y, \sigma^2_Y)$ 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,

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

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

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 $n = 3$.

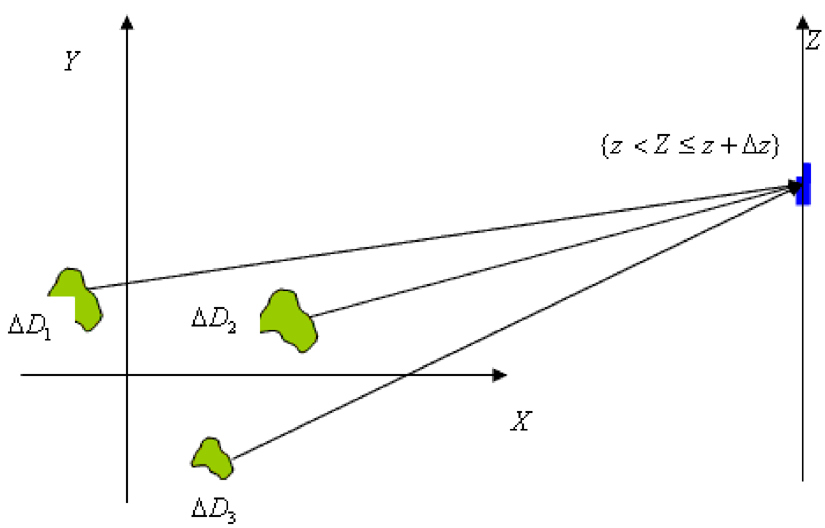


Figure 1

Note that

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

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

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

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,

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

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

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

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

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

(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

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

Since the minimum is nonnegative,

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

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 $f_{X,Y}(x,y) = f_X(x) f_Y(y)$.

then

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

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

$$= EXEY$$

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:

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

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

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

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

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

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

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

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

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

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

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

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.

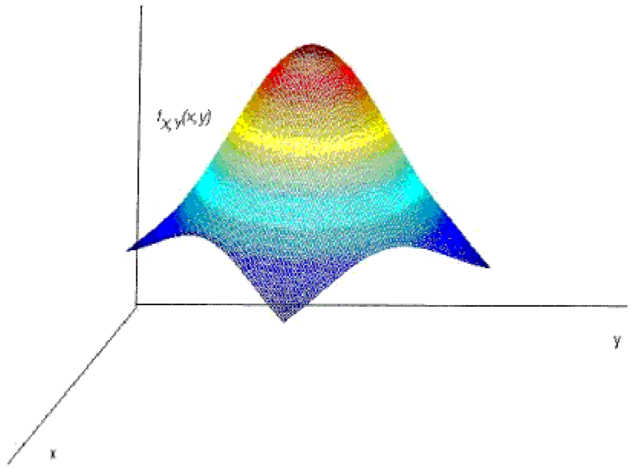


Figure 1 Jointly Gaussian PDF surface

Properties of jointly Gaussian random variables

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

We have

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

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

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

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

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

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

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

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

$$\phi_Z(\omega) = \phi_{X,Y}(\omega, \omega)$$
$$= \phi_X(\omega)\phi_Y(\omega)$$

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 $g(\boldsymbol{z})$, 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 $g(\boldsymbol{z})$ 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 $\frac{1}{\det A}f\left(A^{-1}(y-B)\right)$.

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

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

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

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 $N\left(\mu,\sigma^2\right)$. 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 –IV
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 $X(t,s)$.

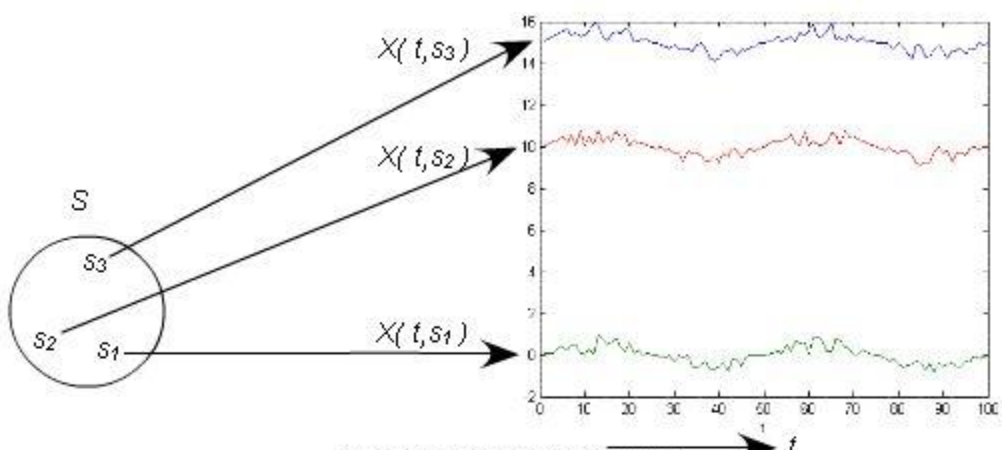


Figure : Random Process

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 $n-\mathrm{th\ order}$ 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 $n-$ 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 $t = E(X(t))$
- $R_X(t_1,t_2) = autocorrelation\ function$ **of the process at times** $t_1, t_2 = E(X(t_1)X(t_2))$.

Note that

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

- The *autocovariance function* $C_X(t_1,t_2)$ of the random process at time t_1 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 $\rho_X(t_1,t_2) = \frac{C_X(t_1,t_2)}{\sqrt{C_X(t_1,t_1)CX(t_2,t_2)}}$ 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

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

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 characteristics 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

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

$\forall n \in N,\ \ \forall t_0 \in \Gamma$ 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 $n = 1, 2, .., k$, 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

  $F_{X(t_1)}(x_1) = F_{X(0)}(x_1)$ 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

$$
\begin{aligned}
R_X(t_1,t_2) &= E(X(t_1)X(t_2)) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X(0),X(t_1-t_2)}(x_1,x_2)\,dx_1 dx_2 \\
&= R_X(t_1-t_2)
\end{aligned}
$$

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 $\forall n,$

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

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 $\{X(t)\}$

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

(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

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

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

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

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:

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

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

We Have

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

Further,

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

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, $R_{XY}(\tau) = EX\left(t+\tau\right)Y\left(t\right) = 0$

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 $\Phi \sim U[0\ 2\pi]$

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

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

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

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

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

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

$$E\left(\langle\mu_X\rangle_T - \mu_X\right)^2 = E\left(\frac{1}{2T}\int_{-T}^{T} X(t)dt - \mu_X\right)^2$$
$$= E\left(\frac{1}{2T}\int_{-T}^{T} (X(t) - \mu_X)dt\right)^2$$
$$= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} E(X(t_1) - \mu_X)(X(t_2) - \mu_X)dt_1dt_2$$
$$= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C_X(t_1 - t_2)dt_1dt_2$$

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:

$$E\left(\langle\mu_X\rangle_T - \mu_X\right)^2 = \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C_X(t_1 - t_2)dt_1dt_2$$
$$= \frac{1}{4T^2}\int_{-2T}^{2T}(2T - |\tau|)C_X(\tau)d\tau$$
$$= \frac{1}{2T}\int_{-2T}^{2T}\left(1 - \frac{|\tau|}{2T}\right)C_X(\tau)d\tau$$

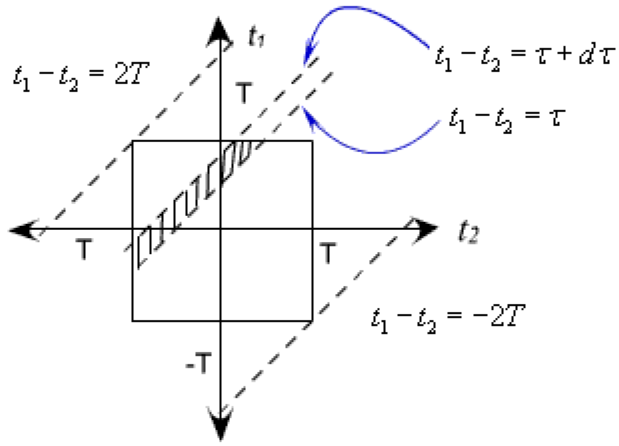


Figure 1

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

$$\begin{aligned} \int_{-2T}^{2T} |C_X(\tau)| d\tau &= 2\int_{0}^{2T} |C_X(\tau)| d\tau \\ &= 2\int_{0}^{T_p} \left(1 - \frac{\tau}{T_p}\right) d\tau \\ &= 2\left(T_p + \frac{T_p^3}{3T_p^2} - \frac{T_p^2}{T_p}\right) \\ &= \frac{2T_p}{3} \end{aligned}$$

$$\int_{-2T}^{2T} |C_X(\tau)| d\tau < \infty$$

hence $\{X(t)\}$ is mean ergodic.

Autocorrelation ergodicity

$$\langle R_X(\tau) \rangle_T = \frac{1}{2T} \int_{-T}^{T} X(t) X(t+\tau) \mathrm{d}t$$

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

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\tau_1|}{2T}\right) C_Z(\tau_1) \mathrm{d}\tau_1 = 0$$

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

$$\langle R_x(\tau) \rangle_T = \frac{1}{2T}\int_{-T}^{T} A\cos(w_0 t + \phi_1) A\cos(w_0(t+\tau) + \phi_1)\,dt$$
$$= \frac{A^2}{4T}\int_{-T}^{T} [\cos w_0\tau + A\cos(w_0(2t+\tau) + 2\phi_1)]\,dt$$
$$= \frac{A^2\cos w_0\tau}{2} + \frac{A^2\sin(w_0(2T+\tau))}{4w_0 T}$$

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 -V
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

$$X_T(t) = X(t) \qquad -T < t < T$$
$$= 0 \qquad \text{otherwise}$$
$$= X(t)rect(\frac{t}{2T})$$

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

$$E\frac{|FTX_T(\omega)|^2}{2T} = E\frac{FTX_T(\omega)FTX_T^*(\omega)}{2T}$$

$$= \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} EX_T(t_1)X_T(t_2)e^{-j\omega t_1}e^{+j\omega t_2}dt_1dt_2$$

$$= \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} R_X(t_1-t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2$$

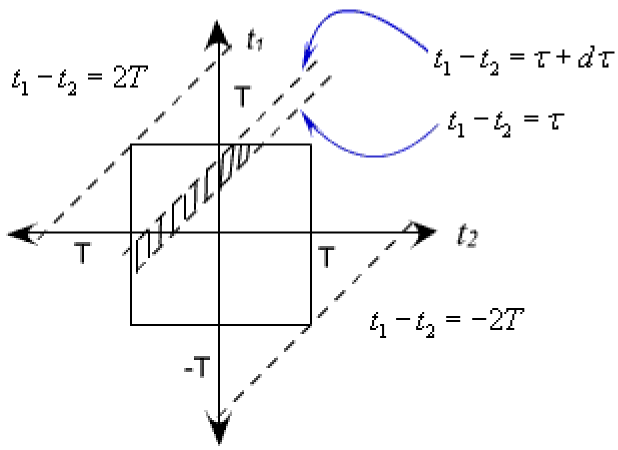


Figure 1

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,

$$E\frac{FTX_\tau(\omega)X_\tau^*(\omega)}{2T} = \frac{1}{2T}\int_{-2T}^{2T} R_x(\tau)e^{-j\omega\tau}(2T-|\tau|)d\tau$$

$$= \int_{-2T}^{2T} R_X(\tau)e^{-j\omega\tau}(1-\frac{|\tau|}{2T})d\tau$$

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

$$R_X(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)e^{j\omega\tau}d\omega$$

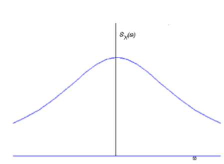
Example 1 The autocorrelation function of a WSS process $\{X(t)\}$ is given by

$$R_X(\tau) = a^2 e^{-b|\tau|} \quad b > 0$$

Find the power spectral density of the process.

$$\begin{aligned}
S_X(\omega) &= \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau \\
&= \int_{-\infty}^{\infty} a^2 e^{-b|\tau|}e^{-j\omega\tau}d\tau \\
&= \int_{-\infty}^{0} a^2 e^{b\tau}e^{-j\omega\tau}d\tau + \int_{0}^{\infty} a^2 e^{-b\tau}e^{-j\omega\tau}d\tau \\
&= \frac{a^2}{b-j\omega} + \frac{a^2}{b+j\omega} \\
&= \frac{2a^2 b}{b^2+\omega^2}
\end{aligned}$$

The autocorrelation function and the PSD are shown in 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.

$$\begin{aligned} R_X(\tau) &= E\, M(t+\tau)\ \cos(\omega_c(t+\tau)+\Phi)\ \ M(t)\cos(\omega_c\tau+\Phi) \\ &= E\, M(t+\tau)\, M(t)\ E\cos(\omega_c(t+\tau)+\Phi)\ \cos(\omega_c\tau+\Phi) \\ &\qquad\qquad (\text{Using the independence of } M(t) \text{ and the sinusoid}) \\ &= R_M(\tau)\,\frac{A^2}{2}\cos\omega_c\tau \end{aligned}$$

$$\therefore\ S_X(\omega) = \frac{A^2}{4}\left(S_M(\omega+\omega_c) + S_M(\omega-\omega_c)\right)$$

where $S_M(\omega)$ is the PSD of $M(t)$.

Figure 4 illustrates the above result.

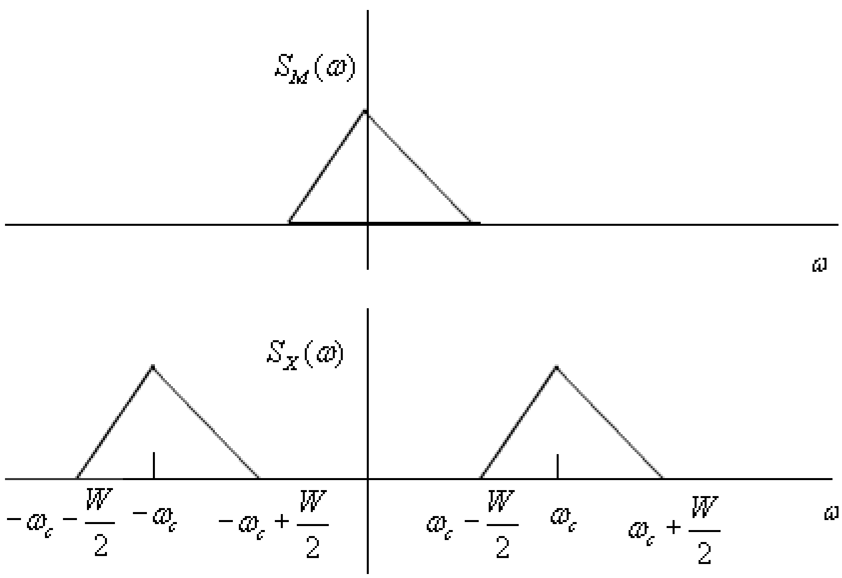


Figure 4

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 $R_{XY}(\tau) = R_{YX}(-\tau)$

$$\therefore S_{XY}(\omega) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j\omega\tau} d\tau$$

$$= \int_{-\infty}^{\infty} R_{YX}(-\tau) e^{-j\omega\tau} d\tau$$

$$= \int_{-\infty}^{\infty} R_{YX}(\tau) e^{j\omega\tau} d\tau$$

$$= S_{YX}^*(\omega)$$

(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

$$S_{XY}(\omega) = \int_{-\infty}^{\infty} R_{XY}(\tau)(\cos\omega\tau + j\sin\omega\tau) d\tau$$

$$= \int_{-\infty}^{\infty} R_{XY}(\tau)\cos\omega\tau d\tau + j\int_{-\infty}^{\infty} R_{XY}(\tau)\sin\omega\tau) d\tau$$

$$= \mathrm{Re}(S_{XY}(\omega)) + j\,\mathrm{Im}(S_{XY}(\omega))$$

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

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

$$\therefore S_{XY}(\omega) = S_{YX}(\omega) = \mu_X \mu_Y \delta(\omega)$$

(4) If $\{X(t)\}$ and $\{Y(t)\}$ are orthogonal, then

$$S_{XY}(\omega) = S_{YX}(\omega) = 0$$

If $\{X(t)\}$ and $\{Y(t)\}$ are orthogonal, we have

$$
\begin{aligned}
R_{XY}(\tau) &= EX(t+\tau)Y(t) \\
&= 0 \\
&= R_{XY}(\tau)
\end{aligned}
$$

$$\therefore S_{XY}(\omega) = S_{YX}(\omega) = 0$$

(5) the *cross power* P_{XY} between $\{X(t)\}$ and $\{Y(t)\}$ is defined by

$$P_{XY} = \lim_{T \to \infty} \frac{1}{2T} E \int_{-T}^{T} X(t)Y(t)dt$$

Applying Parseval's theorem, we get

$$P_{XY} = \lim_{T\to\infty} \frac{1}{2T} E\int_{-T}^{T} X(t)Y(t)dt$$
$$= \lim_{T\to\infty} \frac{1}{2T} E\int_{-\infty}^{\infty} X_T(t)Y_T(t)dt$$
$$= \lim_{T\to\infty} \frac{1}{2T} E\frac{1}{2\pi}\int_{-\infty}^{\infty} FTX_T^*(\omega)FTY_T(\omega)d\omega$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \lim_{T\to\infty} \frac{EFTX_T^*(\omega)FTY_T(\omega)}{2T}d\omega$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XY}(\omega)d\omega$$

$$\therefore\ P_{XY} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XY}(\omega)d\omega$$

Similarly,

$$P_{YX} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{YX}(\omega)d\omega$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XY}^*(\omega)d\omega$$
$$= P_{XY}^*$$

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,

$$y(t) = T[x(t)]$$

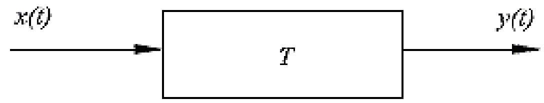


Figure 1

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

$$T\left[a_1x_1(t)+a_2x_2(t)\right] = a_1T\left[x_1(t)\right]+a_2T\left[x_2(t)\right]$$

Example 1 Consider the output of a differentiator, given by

$$y(t) = \frac{d\,x(t)}{dx}$$

Then, $\frac{d}{dt}\left(a_1x_1(t) + a_2x_2(t) \right)$

$$= a_1\frac{d}{dt}x_1(t) + a_2\frac{d}{dt}x_2(t)$$

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.*

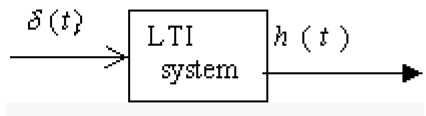


Figure 2

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

$$\begin{aligned} y(t) &= T\int_{-\infty}^{\infty} x(s)\,\delta(t-s)\,ds \\ &= \int_{-\infty}^{\infty} x(s)\,T\delta(t-s)\,ds \quad [\text{ Using the linearity property }] \\ &= \int_{-\infty}^{\infty} x(s)\,h(t,s)\,ds \end{aligned}$$

Where $h(t,s) = T\delta(t-s)$ is the response at time *t* due to the shifted impulse? $\delta(t-s)$

If the system is time invariant,

$$h(t,s) = h(t-s)$$

Therefore for a linear-time invariant system,

$$y(t) = \int_{-\infty}^{\infty} x(s)\,h(t-s)\,ds = x(t) * h(t)$$

where * denotes the convolution operation.

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.

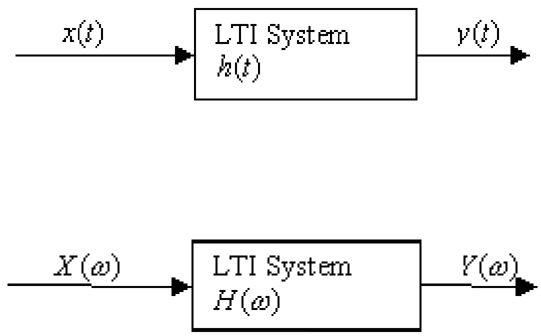


Figure 3

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 $\{Y(t)\}$

$$
\begin{aligned}
EY(t) &= E\int_{-\infty}^{\infty} h(s)X(t-s)\,ds \\
&= \int_{-\infty}^{\infty} h(s)EX(t-s)\,ds \\
&= \int_{-\infty}^{\infty} h(s)\mu_X\,ds \\
&= \mu_X \int_{-\infty}^{\infty} h(s)\,ds \\
&= \mu_X H(0)
\end{aligned}
$$

Where $H(0)$ is the frequency response 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

$$
\begin{aligned}
E\big(X(t+\tau)Y(t)\big) &= E\ X(t+\tau)\int_{-\infty}^{\infty} h(s)\ X(t-s)\ ds \\
&= \int_{-\infty}^{\infty} h(s)\ \ E\ X(t+\tau)\ X(t-s)\ ds \\
&= \int_{-\infty}^{\infty} h(s)\ \ R_X(\tau+s)\,ds \\
&= \int_{-\infty}^{\infty} h(-u)\ \ R_X(\tau-u)\,du \qquad [\text{ Put } s = -u\ ] \\
&= h(-\tau) * R_X(\tau)
\end{aligned}
$$

$$
\begin{aligned}
\therefore \quad & R_{XY}(\tau) = h(-\tau) * R_X(\tau) \\
also \quad & R_{YX}(\tau) = R_{XY}(-\tau) = h(\tau) * R_X(-\tau) \\
& \qquad\qquad = h(\tau) * R_X(\tau)
\end{aligned}
$$

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

$$E\left(X(t+\tau)Y(t)\right) = E\, X(t+\tau) \int_{-\infty}^{\infty} h(s)\, X(t-s)\, ds$$

$$= \int_{-\infty}^{\infty} h(s) \quad E\, X(t+\tau)\, X(t-s)\, ds$$

$$= \int_{-\infty}^{\infty} h(s) \quad R_X(\tau+s) ds$$

$$= \int_{-\infty}^{\infty} h(-u) \quad R_X(\tau-u) du \qquad [\text{ Put } s = -u \text{ }]$$

$$= h(-\tau) \;*\; R_X(\tau)$$

$$\therefore \quad R_{XY}(\tau) = h(-\tau) \;*\; R_X(\tau)$$

$$also \quad R_{YX}(\tau) = R_{XY}(-\tau) = h(\tau) \;*\; R_X(-\tau)$$

$$= h(\tau) \;*\; R_X(\tau)$$

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.,

$$E Y(t) Y(t+\tau) = R_Y(\tau)$$

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

$$\begin{aligned} R_{XY}(\tau) &= h(-\tau) * R_X(\tau) \\ \text{and } R_{YX}(\tau) &= h(\tau) * R_X(\tau) \end{aligned}$$

Taking the Fourier transform of $R_{XY}(\tau)$ we get the cross power spectral density $S_{XY}(\omega)$ given by

$$\begin{aligned} &S_{XY}(\omega) = H^*(\omega) S_X(\omega) \\ &\text{and} \\ &S_{YX}(\omega) = H(\omega) S_X(\omega) \end{aligned}$$

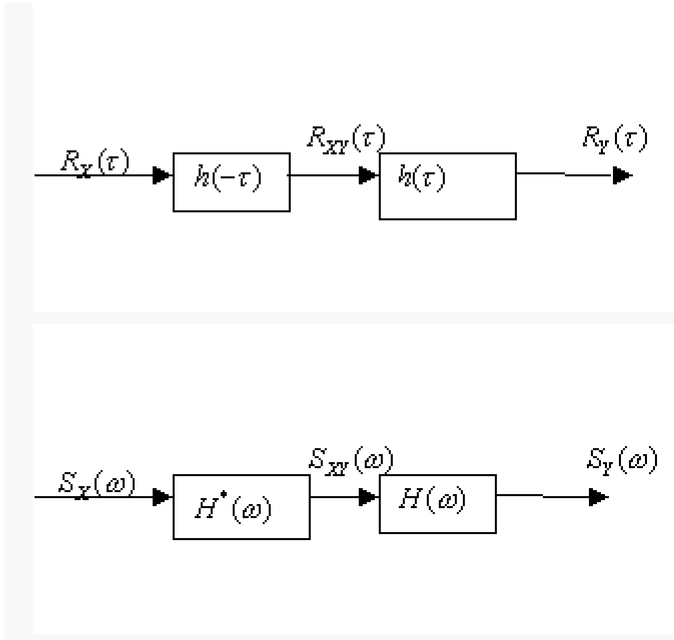


Figure 4

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 $EY^2(t)$

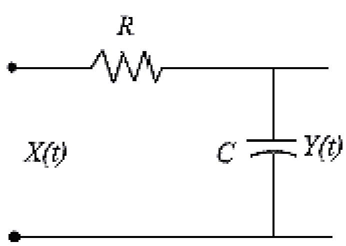


Figure 7

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 $\{(\omega_0 - \frac{B}{2} \le |\omega| \le \omega_0 + \frac{B}{2})\}$. $X_c(t)$ 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

$$R_{X_cX_s}(\tau) = R_X(\tau)\cos\omega_0(t+\tau)\cos\omega_0 t - R_Y(\tau)\sin\omega_0(t+\tau)\sin\omega_0 t$$
$$- R_{XY}(\tau)\cos\omega_0(t+\tau)\sin\omega_0 t + R_{YX}(\tau)\sin\omega_0(t+\tau)\cos\omega_0 t$$
$$= R_X(\tau)[\cos\omega_0(t+\tau)\cos\omega_0 t - \sin\omega_0(t+\tau)\sin\omega_0 t]$$

$$- R_{XY}(\tau)[\cos\omega_0(t+\tau)\sin\omega_0 t - \sin\omega_0(t+\tau)\cos\omega_0 t]$$
$$= R_X(\tau)\cos\omega_0\tau \;\; - R_{XY}(\tau)\sin(-\omega_0\tau)$$
$$= R_X(\tau)\cos\omega_0\tau \;\; - R_{YX}(\tau)\sin\omega_0\tau$$

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 $h(t) = \dfrac{1}{\pi t}$ and the integral is defined in the mean-square sense. See the illustration in Figure 2.

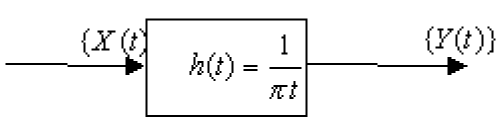


Figure 2

The frequency response $H(\omega)$ of the Hilbert transform is given by

$$H(\omega) = \begin{cases} -j, & \text{if } \omega>0 \\ j, & \text{if } \omega<0 \\ 0, & \text{if } \omega=0 \end{cases}$$

$$\therefore H(\omega) = -j\,\mathrm{sgn}(\omega)$$

$$\text{and } |H(\omega)|^2 = 1$$

$$\therefore S_Y(\omega) = |H(\omega)|^2 S_X(\omega) = S_X(\omega)$$

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

$$S_Y(\omega) = S_X(\omega)$$
and
$$S_{XY}(\omega) = -S_{YX}(\omega)$$

From the above two relations, we get

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

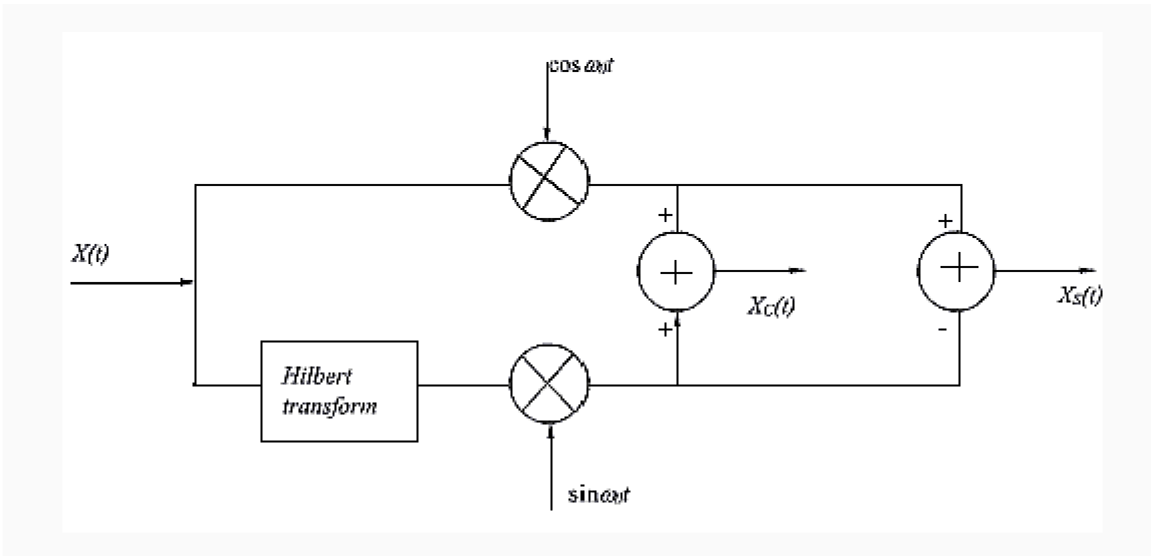


Figure 3

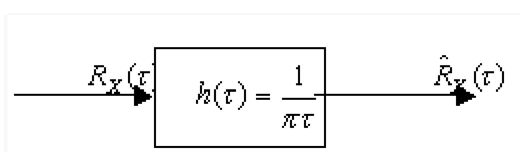
From the above analysis, we can summarize the following expressions for the autocorrelation functions

$$R_{X_c}(\tau) = R_{X_s}(\tau)$$
$$= R_X(\tau)\cos\omega_0\tau + R_{YX}(\tau)\sin\omega_0\tau$$
$$= R_X(\tau)\cos\omega_0\tau + h(t) * R_X(\tau)\sin\omega_0\tau \qquad \because R_{YX}(\tau) = h(\tau) * R_X(\tau)$$
$$= R_X(\tau)\cos\omega_0\tau + \hat{R}_X(\tau)\sin\omega_0\tau$$

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,

$$\begin{aligned} R_{X_c X_s}(\tau) &= R_X(\tau)\sin\omega_0\tau - R_{YX}(\tau)\cos\omega_0\tau \\ &= R_X(\tau)\sin\omega_0\tau - \hat{R}_X(\tau)\cos\omega_0\tau \end{aligned}$$

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