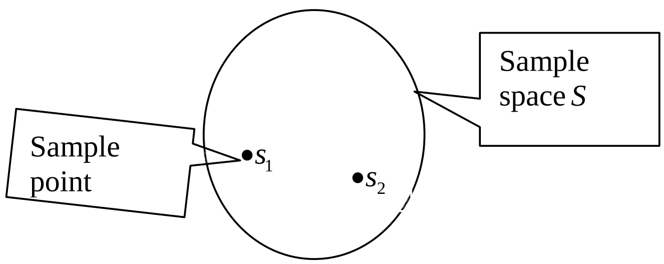
Before we give a definition of probability, let us examine the following concepts:

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.

2. *Sample Space*: The sample space S is the collection of all 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*



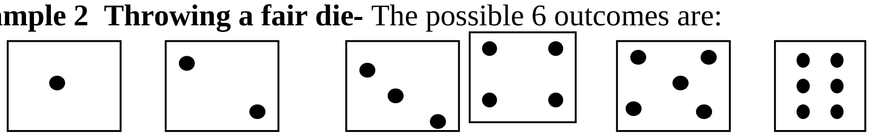
3. *Event:* An event A is a subset of the sample space such that probability can be assigned to it. Thus

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

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

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

$B =$ The event of getting a six=$\{6\}$

And so on.

Example 3 Tossing a fair coin until a head is obatined

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 $S=\{s \mid s\in\mathbb{R}, -1\le s\le 1\}=[-1,\ 1].$ Clearly S is a continuous sample space.

The probability of an event A is a number $P(A)$ assigned to the event . Let us see how we can define probability.

1. **Classical definition of probability (Laplace 1812)**

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 = Number of outcomes favourable to A.

Example 4 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 5 A fair coin is tossed twice. What is the probability of getting two 'heads'?

Here $S=\{HH,\ TH,\ TT,\ TT\}$ 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 example, 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.

2. Relative-frequency based definition of proability(von Mises, 1919)

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

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

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?

Example 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**

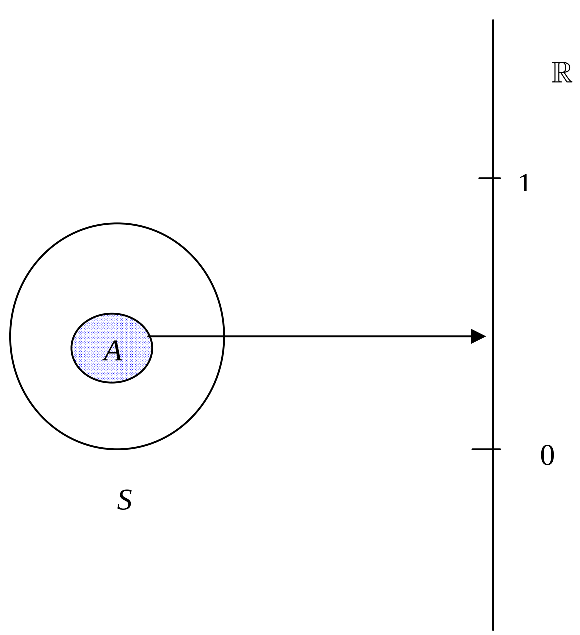
3. Axiometic definition of probability (Kolmogorov, 1933)

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 F a sigma field defined over it. Let $P: F \to \Re$ be a mapping from the sigma-algebra F into the real line such that for each $A \in F$, there exists a unique $P(A) \in \Re$. Clearly P is a set function and is called probability if it satisfies the following these axioms

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

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



Remark

- The triplet $(S, \mathbb{F}, P)$. is called the probability space.

- Any assignment of probability assignment must satisfy the above three axioms
- If *if* $A \cap B = \emptyset, \quad P(A \cup B) = P(A) + P(B)$

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. $P(\phi) = 0$

This is because,

$$S \cup \phi = S$$
$$\Rightarrow P(S \cup \phi) = P(S)$$
$$\Rightarrow P(S) + P(\phi) = P(S)$$
$$\therefore P(\phi) = 0$$

2. $P(A^c)$ = 1- P (A) where where A$\in$ F

We have

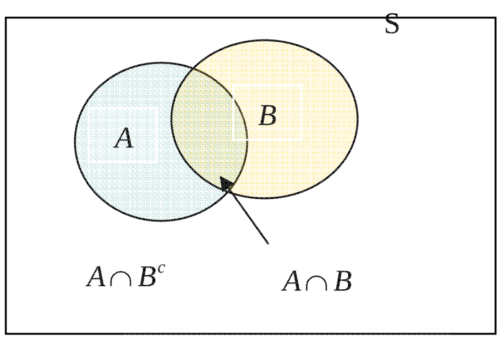
$$A \cup A^c = S$$
$$\Rightarrow P(A \cup A^c) = P(S)$$
$$\Rightarrow P(A) + P(A^c) = 1 \quad \because A \cap A^c = \phi$$
$$\therefore P(A) = 1 - P(A^c)$$

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

We have

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

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



$A = (A \cap B^c) \cup (A \cap B)$

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 $A, B \in \mathbb{F}, \quad P(AUB) = P(A) + P(B) - P(A \cap B)$

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.*

7. Principle of Inclusion-exclusion

Suppose $A_1, A_2, ..., A_n \in \mathbb{F}$ Then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i,j|i<j} P(A_i \cap A_j) + \sum_{i,j,k|i<j<k} P(A_i \cap A_j \cap A_K) - .... + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Discussion
We require some rules to assign probabilities to some basic events in $\Im$. For other events we can compute the probabilities in terms of the probabilities of these basic events.

Probability assignment in a discrete sample space

Consider a finite sample space $S = \{ s_1, s_2, ...s_n \}$. Then the sigma algebra $\Im$ is defined by the power set of S. For any e*lementary event* $\{s_i\} \in \Im$, we can assign a probability $P(s_i)$ such that,

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

For any event $A \in \Im$, 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.
$$\therefore \quad \sum_{i=1}^{n} p = 1$$
$$\Rightarrow \ p = 1/n$$
$$\therefore \quad P(A) = P\left(\bigcup_{S_i \in A} \{S_i\}\right)$$
$$= \quad n(A) \ \frac{1}{n} \quad = \frac{n(A)}{n}$$
Example 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 equiprobale and $S = A_1 \cup A_2 \cup .... \cup A_6$ we get

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

Su ppose 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 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 4th 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.

Suppose $S = \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ is a non-negative integrable function such that,

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

For any Borel set A,

$P(A) = \int_A f(x)\, dx$ defines the probability on the Borel sigma-algebra B.

We can similarly define probability on the continuous space of $\mathbb{R}^2, \mathbb{R}^3$ etc.

Example Suppose

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

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

$$P([a_1, b_1]) = \frac{1}{b - a}$$

Example Consider $S = \mathbb{R}^2$ the two-dimensional Euclidean space. Let $S_1 \subseteq \mathbb{R}^2$ and $|S_1|$ represents the area under S_1.

$$f_X(x) = \begin{cases} \frac{1}{|S_1|} & \text{for } x \in S_1 \\ 0 & \text{otherwise} \end{cases}$$

$$P(A) = \frac{|A|}{|S_1|}$$

This example interprets the geometrical definition of probability.

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 rules 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 $A \times B = \left\{ \left(a_i, b_j \right) | a_i \in A, b_j \in B \right\}$. Then $A \times B$ contains mn ordered pair of elements. This is illustrated in Fig for $m = 5$ and $n = 4$. In 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_5
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

Fig Illustration of the product rule

The above result can be generalized as follows:

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

Example
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 $6.6 = 36$ 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 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

(2) Sampling with replacement and ordering:

Suppose we have to choose k objects from a set of n objects. Further, after every choosing, the object is placed back in the set. In this case, the number of distinct ordered k-tupples = $n \times n \times ...... \times n \ (k - \text{times}) = n^k$.

(3) Sampling without replacement:

Suppose we have to choose k objects from a set of n objects by picking one object after another at random.

In this case the first object can be chosen from n objects, the second object can be chosen from n-1 objects, and so. Therefore, by applying the product rule, the number of distinct order k-tupples in this case is

$$n \times (n-1) .... (n-k+1)$$

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

The number $\frac{n!}{(n-k)!}$ is called the *permutation* of n objects taking k at a time and denoted by ${}^nP_k$. Thus

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

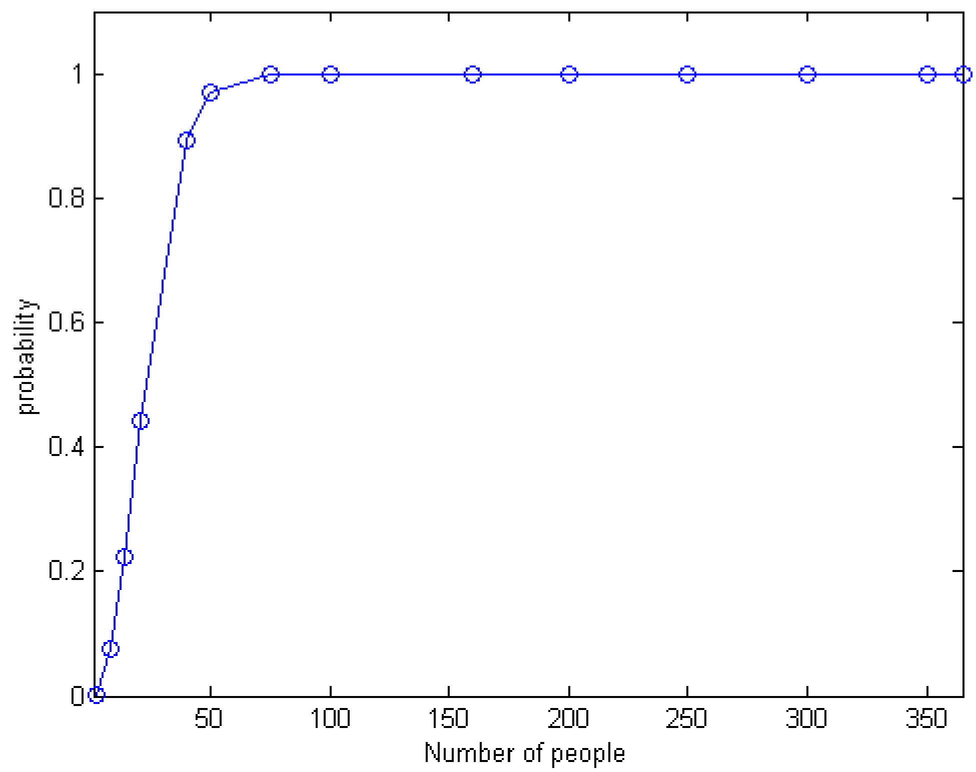
Clearly , ${}^nP_n = n!$

Example: *Birthday problem*- Given a class of students, what is the probability of two students in theclass having the same birthday? Plot this probability vs. number of people and be surprised! If the group has more than 365 people the probability of two people in the group having the same birthday is obviously 1.

Let k 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 Fig. . 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 stuents is exactly one.

(4) Sampling without replacing and without ordering

Suppose ${}^nC_k$ be the number of ways in which k objects can be chosen out of a set of n objects. In this case ordering of the objects in the set of k objects is not considered.

Note that k objects can be arranged among themselves in $k!$ ways. Therefore, if ordering of the k objects is considered, the number of ways in which k objects can be chosen out of n objects is ${}^nC_k k!$. This is the case of sampling with ordering.

$$\therefore {}^nC_k k! = {}^n p_k = \frac{\underline{|n}}{\underline{|n-k}}$$

$$\therefore {}^nC_k = \frac{\underline{|n}}{\underline{|k}\,\underline{|n-k}}$$

${}^nC_k$ is also called the binomial co-efficient.

Example:

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 $\frac{{}^6C_4 \times {}^5C_3 \times {}^4C_2}{{}^{15}C_9}$

(5) Arranging n objects into k specific groups

Suppose we want to partition a set of n distinct elements into k distinct subsets $A_1, A_2, \ldots\ldots, A_n$ of sizes $n_1, n_2, .., n_k$ respectively so that $n = n_1 + n_2 + \ldots\ldots + n_k$. Then the total number of distinct partitions is

$$\frac{n!}{n_1!n_2!\ldots.n_k!}$$

This can be proved by noting that the resulting number of partitions is ${}^nC_{n_1} \times {}^{n-n_1}C_{n_2} \times \ldots\ldots\ldots\ldots \times {}^{n-n_1-n_2\ldots n_{k-1}}C_{n_k}$

$$= \frac{n!}{n_1!(n-n_1)!} \times \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \times \ldots.. \times \frac{(n-n_1-n_2\ldots.-n_{k-1})!}{n_k!(n-n_1-n_2..-n_{k-1}-n_k)!}$$

$$= \frac{n!}{n_1!n_2!\ldots.n_k!}$$

Example: 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

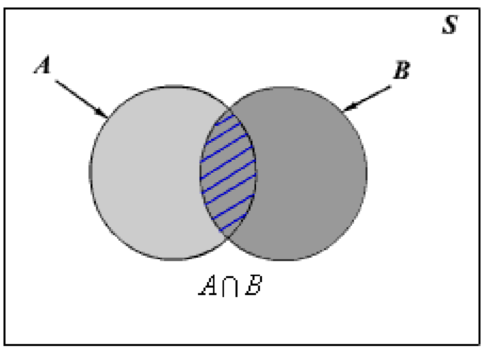
Consider the probability space $(S, \mathbb{F}, P)$. Let A and B two events in $\mathbb{F}$. We ask the following question –

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 $A \cap B$.



Clearly

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

$$= \frac{N_{AB}}{N_A}$$

$$= \frac{\frac{N_{AB}}{N}}{\frac{N_A}{N}} = \frac{P(A \cap B)}{P(A)}$$

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

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

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

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

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

Example 1

Consider the example tossing the fair die. Suppose

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

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

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

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

Example 2

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

A = event of at least one girl

B = event of two girls

Clearly

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

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

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

Conditional probability and the axioms of probability

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

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

Axiom 1

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

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

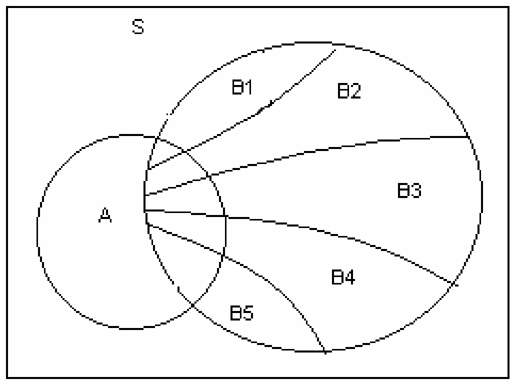
Axiom 2 We have S ∩ A = A

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

Axiom 3 Consider a sequence of disjoint events $B_1, B_2, ..., B_n, ...$ We have

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

(See the Venn diagram below for illustration of finite version of the result.)



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

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

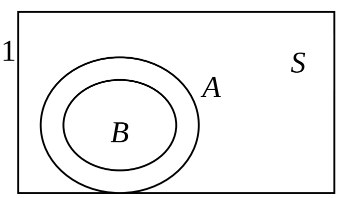
$$\therefore P\left(\bigcup_{i=1}^{\infty} B_i / A\right) = \frac{P\left(\bigcup_{i=1}^{\infty} B_i \cap A\right)}{P(A)} = \frac{\sum_{i=1}^{\infty} P(B_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} P(B_i / A)$$

Properties of Conditional Probabilities

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

We have $A \cap B = B$

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



and

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

(2) Chain rule of probability

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

We have

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

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

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

(3) Theorem of 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,

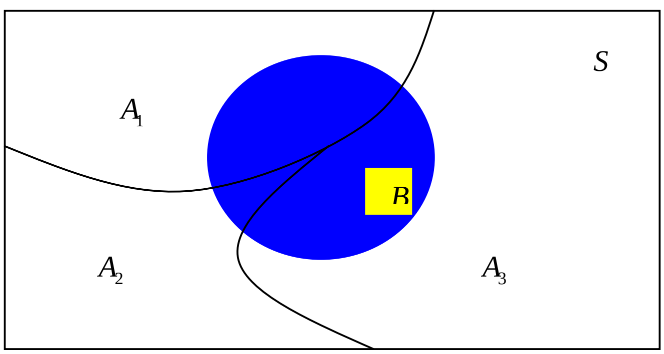
$$P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$$

Proof: We have $\bigcup_{i=1}^{n} B \cap A_i = B$ and the sequence $B \cap A_i$ is disjoint.

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

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

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



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

$S = A_1 \cup A_2 \ldots\ldots \cup A_n$ and $A_i \cap A_j = \phi$ for i ≠ j.

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

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

$$P(B) = P(A_1)P(B/A_1) + P(A_1^c)P(B/A_1^c)$$

$$= \frac{2}{5}\times\frac{1}{4} + \frac{3}{5}\times\frac{2}{4} = \frac{2}{5}$$

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

$$\boxed{P\left(A \cap B\right) = P(A)P(B)}$$

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 $A_1, A_2, ..., A_n$ are called independent if and only if

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

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

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

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

$P\left(A\right) = \frac{1}{2}$ and $P\left(B\right) = \frac{1}{2}$.

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

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

Hence the events A and B are independent.

Example 5 Consider the experiment of picking two balls at random discussed in example 3. 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.

Bayes' Theorem

Suppose $A_1, A_2, \ . \ . \ . \ A_n$ *are partitions on S such that* $S = A_1 \cup A_2 ..... \cup A_n$ and $A_i \cap A_j = \phi$ for i ≠ j.
Suppose the event B occurs if one of the events $A_1, A_2, \ . \ . \ . \ A_n$ occurs. Thus we have the information of 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 wo *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)

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

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

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 $P(A_k \mid B)$ 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) $P(R_1) = \text{Probabilty of receiving 'one'}$

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

(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 percentage 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

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

Repeated Trials

In our discussions so far, we considered the probability defined over a sample space corresponding to a random experiment. Often, we have to consider several random experiments in a sequence. For example, the experiment corresponding to sequential transmission of bits through a communication system may be considered as a sequence of experiments each representing transmission of single bit through the channel.

Suppose two experiments E_1 and E_2 with the corresponding sample space S_1 and S_2 are performed sequentially. Such a combined experiment is called the *product* of two experiments E_1 and E_2.

Clearly, the outcome of this combined experiment consists of the *ordered pair* (s_1, s_2) *where* $s_1 \in S_1$ *and* $s_2 \in S_2$. The sample space corresponding to the combined experiment is given by $S = S_1 \times S_2$. The events in S consist of all the Cartesian products of the form $A_1 \times A_2$ where A_1 is an event in S_1 and A_2 is an event in S_2. Our aim is to define the probability $P\left(A_1 \times A_2\right)$.

The sample space $S_1 \times S_2$ and the events $A_1 \times S_2, S_1 \times A_2$ and $A_1 \times A_2$ are illustrated in in Fig.

We can easily show that

$$P\left(S_1 \times A_2\right) = P_2(A_2)$$

and

$$P\left(A_1 \times S_2\right) = P_1(A_1)$$

where P_i is the probability defined on the events of $A_i \; i = 1, 2. A_i$,. This is because, the event $A_1 \times \mathrm{S}_2$ in S occurs whenever A_1 in S_1 occurs, irrespective of the event in S_2.

Also note that

$$A_1 \times A_2 = A_1 \times S_2 \cap S_1 \times A_2$$

$$\therefore P(A \times B) = P\left[\left(A_1 \times S_2\right) \cap \left(S_1 \times A_2\right)\right]$$

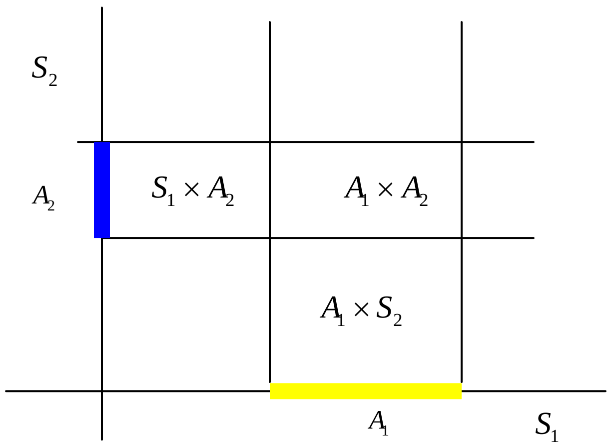


Fig. (Animate)

Independent Experiments:

In many experiments, the events $A_1 \times S_2$ and $S_1 \times A_2$ are independent for every selection of $A_1 \in S_1$ and $A_2 \in S_2$. Such experiments are called independent experiments. In this case can write

$$\begin{aligned} P(A \times B) &= P\left[\left(A_1 \times S_2\right) \cap \left(S_1 \times A_2\right)\right] \\ &= P\left(A_1 \times S_2\right) P\left(S_1 \times A_2\right) \\ &= P_1(A_1) P_2(A_2) \end{aligned}$$

Example 1

Suppose S_1 is the sample space of the experiment of rolling of a six-faced fair die and S_2 is the sample space of the experiment of tossing of a fair die.

Clearly,

$S_1 = \{1, 2, 3, 4, 5, 6\}$, $A_1 = \{2,3\}$

and

$S_2 = \{H, T\}$, $A_2 = \{H\}$

$\therefore S_1 \times S_2 = \{(1, H), (1, T), (2, H), (2, T), (3, H), (3, T), (4, H), (4, T), (5, H), (5, T), (6, H), (6, T)\}$

and

$$P(A_1 \times A_2) = \frac{1}{2} . \frac{2}{6} = \frac{1}{6}$$

Example 2

In a digital communication system transmitting 1 and 0, 1 is transmitted twice as often as 0. If two bits are transmitted in a sequence, what is the probability that both the bits will be 1?

$S_1 = \{0, 1\}$, $A_1 = \{1\}$

and

$S_2 = \{0, 1\}$, $A_2 = \{1\}$

$\therefore S_1 \times S_2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$

and

$$P(A_1 \times A_2) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2}$$

Generalization

We can similarly define the sample space $S = S_1 \times S_2 \times ....\times S_n$ corresponding to n experiments and the Cartesian product of events $A_1 \times A_2 \times ....\times A_n = \left\{ \left(s_1, s_2, ..., s_n \right) \middle| s_1 \in A_1, s_2 \in A_2 ...., s_n \in A_n \right\}$.

If the experiments are independent, we can write

$$P\left(A_1 \times A_2 \times ......\times A_n \right) = P_1\left(A_1 \right) P_2\left(A_2 \right) ....P_n\left(A_n \right)$$

where P_i is probability defined on the event of s_i.

Bernoulli trial

Suppose in an experiment, we are only concerned whether a particular event A has occurred or not. We call this event as the 'success' with probability $P(A) = p$ and the complementary event A^c as the 'failure' with probability $P(A^c) = 1 - p.$ Such a random experiment is called Bernoulli trial.

Binomial Law:

We are interested in finding the probability of k 'successes' in n independent Bernoulli trials. This probability $p_x(k)$ is given by,

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

Consider n independent repetitions of the Bernoulli trial. Let S_1 be the sample space associated with each trial and we are interested in a particular event $A_1 \in s$ and its complement A^c such that $P(A) = p$ and $P(A^c) = 1 - p$. If A occurs in a trial, then we have a 'success' otherwise a 'failure'.

Thus the sample space corresponding to the n repeated trials is $S = S_1 \times S_2 \times ....\times S_n$.

Any event in S is of the form $A_1 \times A_2 \times ....\times A_n$ where some A_i s are A and remaining A_i s are A^c.

Using the property of independent experiment we have,

$$P(A_1 \times A_2 \times ....\times A_n) = P\left(A_1 \right) P\left(A_2 \right) ....P\left(A_n \right).$$

If k A_i s are A and the remaining n - k A_i s are A^c, then

$$P\left(A_1 \times A_2 \times ....\times A_n\right) = p^k (1-p)^{n-k}$$

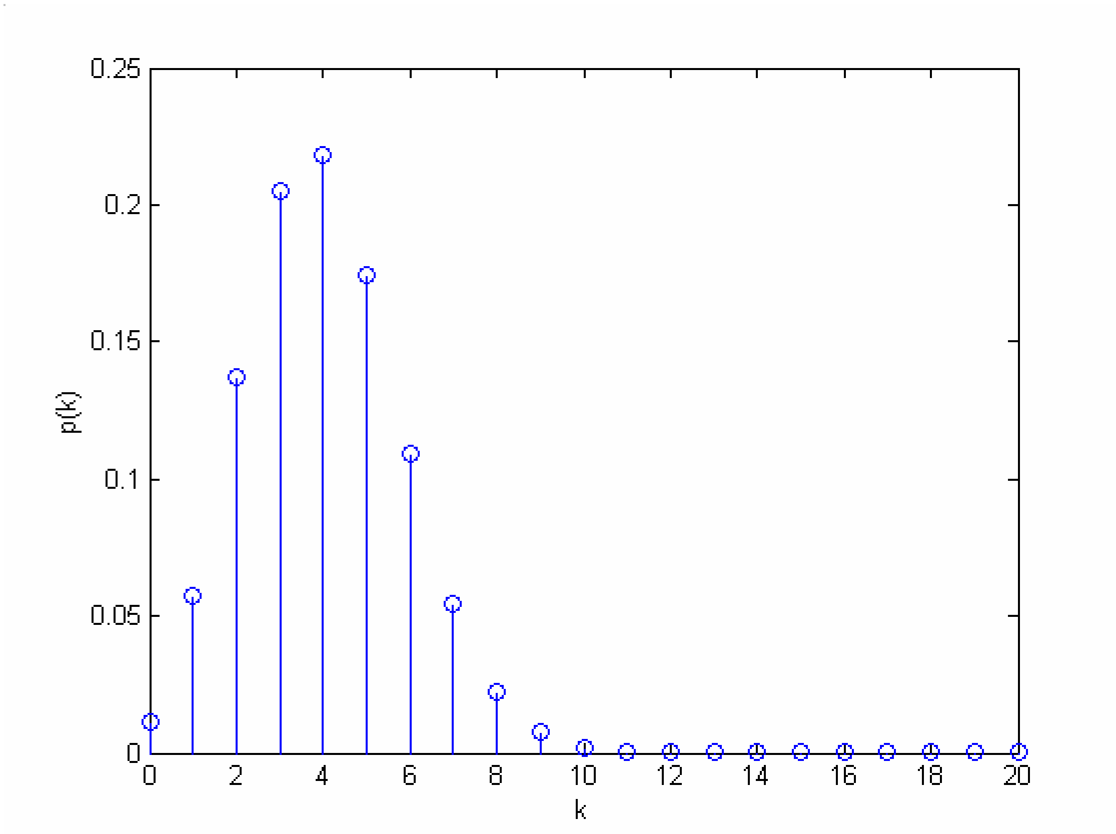
But the ${}^nC_k$ number of events in S with k number of As and $n-k$ number of A^cs. For example, if $n=4, k=2$, the possible events are

$$A \times A^c \times A^c \times A$$
$$A \times A \times A^c \times A^c$$
$$A \times A^c \times A \times A^c$$
$$A^c \times A \times A \times A^c$$
$$A^c \times A^c \times A \times A$$
$$A^c \times A \times A^c \times A$$

We also note that all the ${}^nC_k$ events are mutually exclusive.

Hence the probability of k successes in n independent repetitions of the Bernoulli trial is given by

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



Example1:

A fair dice is rolled 6 times. What is the probability that a 4 appears thrice?

Solution:

We have $s_1 = \{1, 2, 3, 4, 5, 6\}$

$$A = \{4\} \quad \text{with} \quad P(A) = p_i = \frac{1}{6}$$

And $A^c = \{1, 2, 3, 5, 6\}$ with $P(A_c) = 1 - p = 5/6$

$$\therefore P_6(4) = \frac{6 \times 5}{2} \times \left(\frac{1}{6}\right)^4 \times \left(\frac{5}{6}\right)^2 = 0.008$$

Example2:

A communication source emits binary symbols 1 and 0 with probability 0.6 and 0.4 respectively. What is the probability that there will be 5 1s in a message of 20 symbols?

Solution:

$S_1 = \{0,1\}$

$A = \{1\}$, $P(A) = p = 0.6$

$\therefore P_{20}(5) = {}^{20}C_5 \left(0.6\right)^5 \left(0.4\right)^{15}$ =0.0013

Example 3

In a binary communication system, bit error occurs with a probability of 10^{-5}. What is the probability of getting at least one error bit in a message of 8 bits?

Here we can consider the sample space

S_1 = {error in transmission of 1 bit } $\cup$ {*no* error in transmission in transmission of 1 bit}

with p={error in transmission of 1 bit }=10^{-5}

$\therefore$ Probability of no bit-error in transmission of 8 bits

$= P_8(0) = (1-10^{-5})^8 = 0.9999$

$\therefore$ Probability of at least bit-error in transmission of 8 bits=1-$P_8(0) = 0.0001$

Typical plots of binomial probabilities are shown in the figure.

Approximations of the Binomial probabilities

Two interesting approximations of the binomial probabilities are very important.

Case 1

Suppose n is very large and p is very small and $np = \lambda$ a constant.

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

$$= {}^nC_k \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

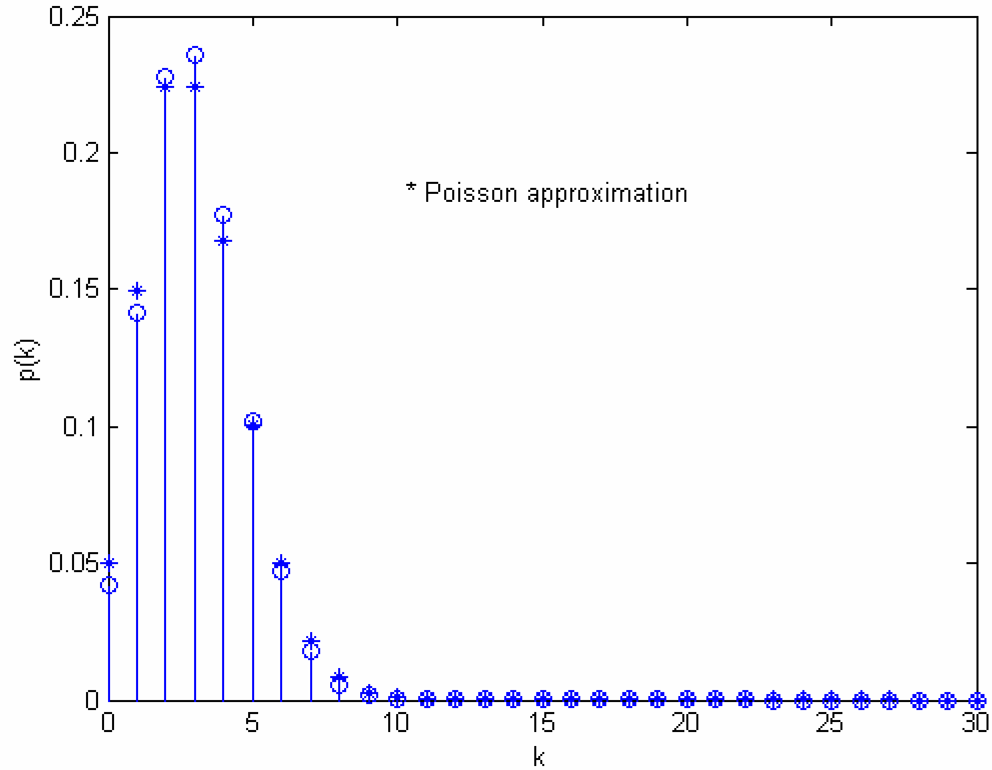
$$= \frac{n(n-1)....(n-k+1)}{\lfloor k}.\frac{\lambda^k\left(1-\frac{\lambda}{n}\right)^n}{n^k\left(1-\frac{\lambda}{n}\right)^k}$$

$$= \frac{n^k\left(1-1/n\right)\left(1-2/n\right).......\left(1-\frac{k-1}{n}\right)}{n^k\left(1-\frac{\lambda}{n}\right)^k}\frac{\lambda^k\left(1-\frac{\lambda}{n}\right)^n}{\lfloor k}$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^k = 0 \text{ and } \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$\therefore\ p_n(k) = \frac{\lambda^k e^{-\lambda}}{\lfloor k}$$

This distribution is known as Poisson probability and widely used in engineering and other fields. We shall discuss more about this distribution in a later class.



Case 2

When n is sufficiently large and $np(1-p) \gg 1$, $p_n(k)$ may be approximated as

$$p_n(k) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}$$

The right hand side is an expression for *normal distribution* to be discussed in a later class.

Example: Consider the problem in example 2.

Here $p_{20}(5) \approx \dfrac{1}{\sqrt{2\pi \times 20 \times 0.6 \times 0.4}} e^{-\frac{(5-3)^2}{2n \times 0.6 \times 0.4}} = 0.0128$

Random Variables

In application of probabilities, we are often concerned with numerical values which are random in nature. 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.

Mathematical Preliminaries
Real-valued point function on a set

Recall that a real-valued function $f: S \rightarrow \mathbb{R}$ maps each element $s \in S$, a unique element $f(s) \in \mathbb{R}$. The set S is called the *domain* of f and the set $R_f = \{f(x) \mid x \in S\}$ is called the *range* of f. Clearly $R_f \subseteq \mathbb{R}$. The range and domain of f are shown in Fig, .

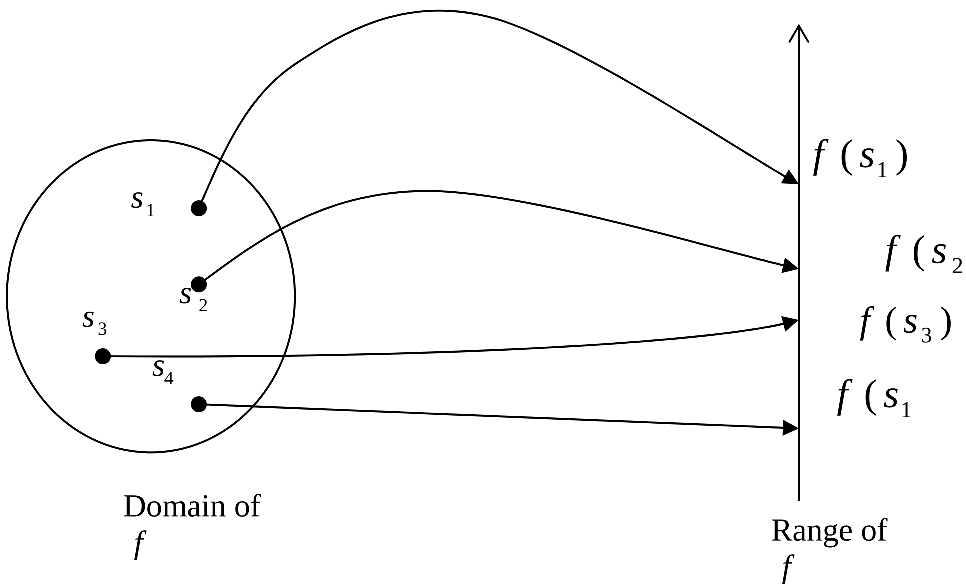


Image and Inverse image
For a point $s \in S$, the functional value $f(s) \in \mathbb{R}$ is called the *image* of the point s. If $A \subseteq S$, then the set of the images of elements of A is called *the image of* A and denoted by $f(A)$. Thus

$f(A) = \{f(s) \mid s \in A\}$

Clearly $f(A) \subseteq R_f$

Suppose $\mathrm{B} \subseteq \mathbb{R}$. The set $\{x \mid f(x) \in \mathrm{B}\}$ is called the inverse image of B under f and is denoted by $f^{-1}(\mathrm{B})$.

Example Suppose $S = \{H, T\}$ and $f: S \rightarrow \mathbb{R}$ is defined by $f(H) = 1$ and $f(T) = -1$. Therefore,

- $R_f = \{1, -1\} \subseteq \mathbb{R}$

- Image of H is 1 and that of T is -1.
- For a subset of $\mathbb{R}$ say $\mathrm{B}_1 = (-\infty,\ 1.5]$,

  $f^{-1}(\mathrm{B}_1) = \{s \mid f(s) \in \mathrm{B}_1\} = \{H, T\}$.

  For another subset $B_2 = [5,\ 9], f^{-1}(B_2) = \phi$.

Random variable

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 $X : S \to \mathbb{R}$ 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
$X^{-1}(\mathrm{B}) = \{s \mid X(s) \in \mathrm{B}\} \in \mathbb{F}$.

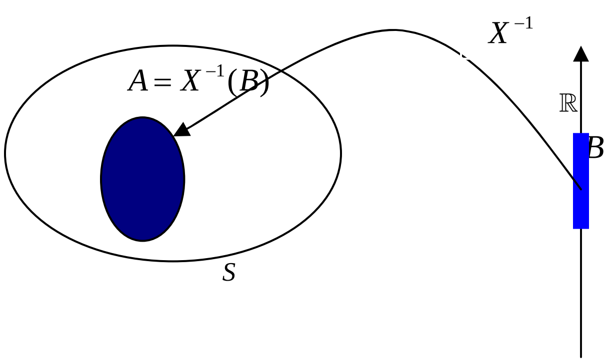


Figure Random Variable

(To be animated)

Notations:

- Random variables are represented by upper-case letters.
- Values of a random variable are denoted by lower case letters
- $X(s) = x$ means that x is the value of a random variable X at the sample point s.
- Usually, the argument s is omitted and we simply write $X = x$.

Remark

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

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

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

- The above definition of the random variable requires that the mapping X is such that $X^{-1}(\mathrm{B})$ 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 1: Consider the example of tossing a fair coin twice. The psample 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 $X = x$
HH	0
HT	1
TH	2
TT	3

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

Example 2: 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 in the face of the die, then

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

Probability Space induced by a Random Variable

The random variable X induces a probability measure P_X on $\mathbb{B}$ defined by

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

The probability measure P_X satisfies the three axioms of probability:

Axiom 1

. $$P_X(B) = P(X^{-1}(B)) \leq 1$$

Axiom 2

$$P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(S) = 1$$

Axiom 3

Suppose $B_1, B_2, ...$ are disjoint Borel sets. Then $X^{-1}(B_1), X^{-1}(B_2), ...$ are distinct events in $\mathbb{F}$. Therefore,

$$\begin{aligned} P_X(\bigcup_{i=1}^{\infty} B_i) &= P(\bigcup_{i=1}^{\infty} X^{-1}(B_i)) \\ &= \sum_{i=1}^{\infty} P(X^{-1}(B_i)) \\ &= \sum_{i=1}^{\infty} P_X(B_i) \end{aligned}$$

Thus the random variable X induces a probability space $(S, \mathbb{B}, P_X)$

Probability Distribution Function

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 $P(\{X \in B\})$ in stead 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 $X^{-1}((-\infty, x]) = \{s \mid X(s) \leq x, \ s \in S\}$ is denoted as $\{X \leq x\}$. The event $\{X \leq x\}$ 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, $\{X > x\} = \{X \leq x\}^c, \{x_1 < X \leq x_2\} = \{X \leq x_2\} \setminus \{X \leq x_1\}$,

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

and so on.

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

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

Value of the random variable

$$F_X(x)$$

Random variable

Eaxmple 3 Consider the random variable X in Example 1

We have

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

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 $1 \le x < 2$,

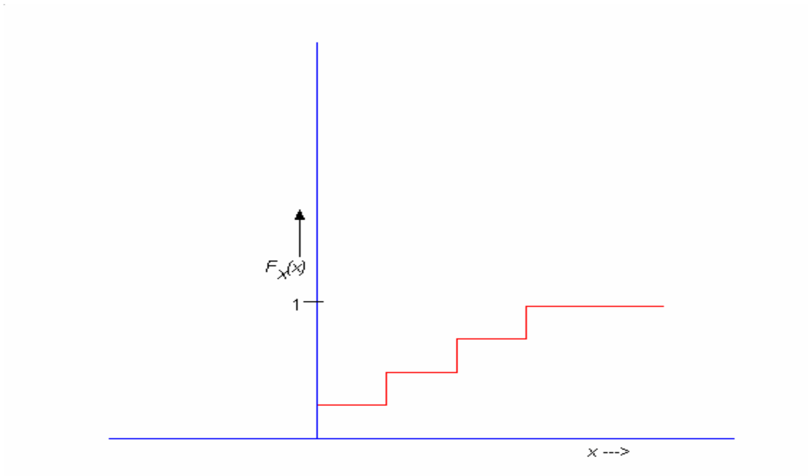
$$\begin{aligned} F_X(x) &= P(\{X \le x\}) \\ &= P(\{X = 0\} \cup \{X = 1\}) \\ &= P(\{X = 0\}) + P(\{X = 1\}) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$

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 $x \ge 3$,

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



Properties of 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, $x_1 < x_2$, then $F_X(x_1) < F_X(x_2)$

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

A real function $f(x)$ is said to be continuous at a point a if and only if

(i) $f(a)$ is defined

(ii) $\lim_{x \to a+} f(x) = \lim_{x \to a-} f(x) = f(a)$

Thefunction $f(x)$ is said to be right-continuous at a point a if and only if

(iii) $f(a)$ is defined

$\lim_{x \to a+} f(x) = f(a)$

- $F_X(x)$ is right continuous

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

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

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

- $F_X(-\infty) = 0$

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

- $F_X(\infty) = 1$

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

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

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

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

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

$$P\{x_1 \le X \le x_2\} = F_X(x_2) - F_X(x_1) + P(X = x_1)$$
$$P(\{x_1 \le X < x_2\}) = F_X(x_2) - F_X(x_1) + P(X = x_1) - P(X = x_2)$$
$$P(\{X > x\}) = P(\{x < X < \infty\}) = 1 - F_X(x)$$

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

Example 4 Consider the random variable X defined by

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

Find a) $P(X = 0)$

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

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

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

Solution:

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

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

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

$$= 1$$

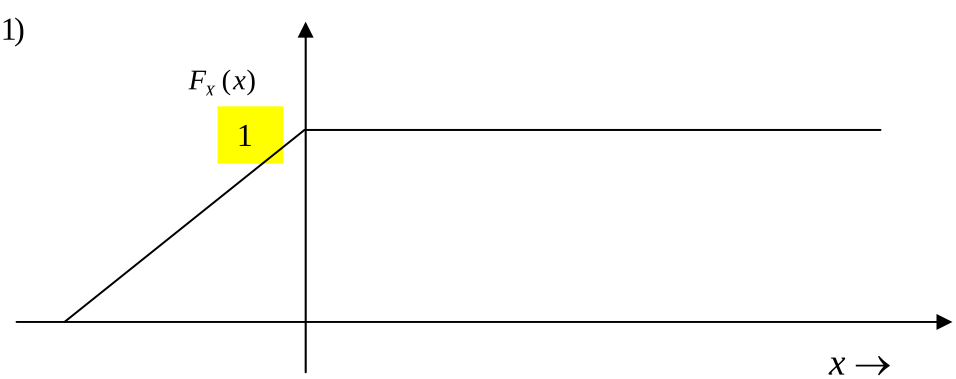
c) $P\{X > 2\} = 1 - F_X(2)$

$$= 1 - 1 = 0$$

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

$$= F_X(1) - F_X(-1)$$

$$= 1 - \frac{1}{8} = \frac{7}{8}$$



Conditional Distribution and Density function:

We discussed conditional probability in an earlier class. 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*.

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

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

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

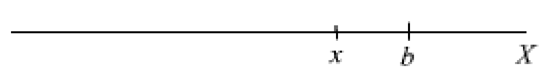
In a similar manner, we can define the conditional density function $f_X(x/B)$ of the random variable *X* given the event *B* as

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

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

Then

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



Case 1: *x<b*

Then

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

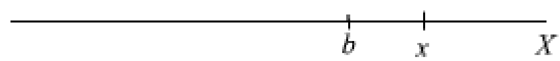
$$=\frac{P\{X\le x\}}{F_X\left(b\right)}=\frac{F_X\left(x\right)}{F_X\left(b\right)}$$

And $f_X\left(x/B\right)=\dfrac{d}{dx}\dfrac{F_X\left(x\right)}{F_X\left(b\right)}=\dfrac{f_X\left(x\right)}{f_X\left(b\right)}$

Case 2: $x\ge b$

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

$$=\frac{P\{X\le b\}}{F_X\left(x\right)}=\frac{F_X\left(b\right)}{F_X\left(x\right)}$$



and $f_X\left(x/B\right)=\dfrac{d}{dx}F_X\left(x/B\right)=\dfrac{d}{dx}\dfrac{F_X\left(b\right)}{F_X\left(x\right)}=0$

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

Remark: We can define the Bayo rule in a similar manner.

Suppose the interval $\{X\le x\}$ is portioned into non overlapping subsets such that $\{X\le x\}=\bigcup_{i=1}^{n}B_i$.

Then $F_X\left(x\right)=\sum_{i=1}^{n}P\left(B_i\right)F_X\left(x/B_i\right)$

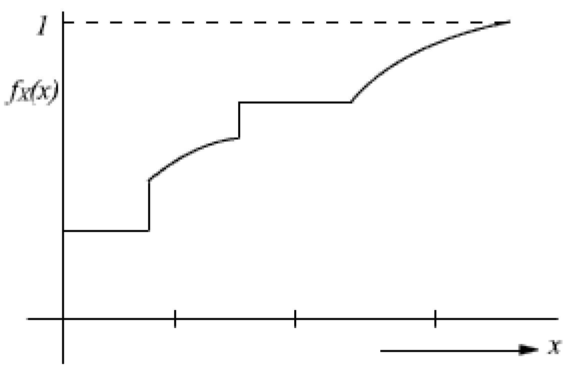
$$\therefore P\left(B/\{X\le x\}\right)=\frac{P\left[B\cup\{X\le x\}\right]}{F_X\left(x\right)}$$

$$=\frac{P\left[B\cup\{X\le x\}\right]}{\sum_{i=1}^{n}P\left(B_i\right)F_X\left(x/B_i\right)}$$

Mixed Type Random variable:

A random variable *X* is said to be mixed type if its distribution function $F_X(x)$ has discontinuous finite number of points and increases strictly with respect to over at least one interval of values of the random variable *X*.

Thus for a mixed type random variable X, $F_X(x)$ has discontinuous, but is not of stair case type as the in the case of discrete random variable. A typical plot of the distribution functions of a mixed type random variable as shown in Figure.



Suppose SD denotes the countable subset of points on SX such that the RV X is characterized by the probability mass function $p_X(x), x \in S_D$. Similarly let SC be a continuous subset of points on SX such that RV is characterized by the probability density function $f_X(x), x \in S_C$.

Clearly the subsets SD and SC partition the set SX. If $P(S_D)$, then $P(S_C) = 1 - p$.

Thus the probability of the event $\{X \leq x\}$ can be expressed as

$$\begin{aligned} P\{X \leq x\} &= P(S_D)P(\{X \leq x\} \mid S_D) + P(S_C)P(\{X \leq x\} \mid S_C) \\ &= pF_D(x) + (1-p)F_C(x) \end{aligned}$$

Where $F_D(x)$ is the conditional distribution function of X given X is discrete and $F_C(x)$ is the conditional distribution function given that X is continuous.

Clearly $F_D(x)$ is a staircase of function and $F_C(x)$ is a continuous function.

Also note that,

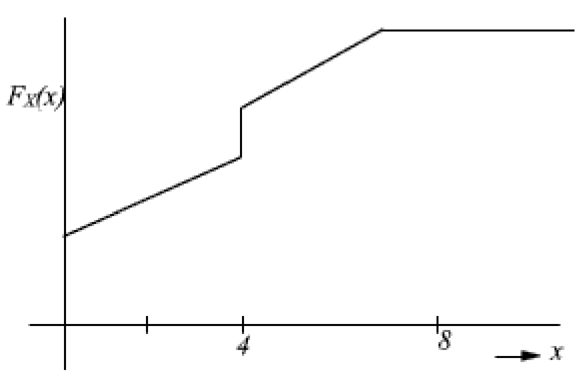
$$
\begin{aligned}
p &= P(S_D) \\
&= \sum_{x \in X_D} p_X(x)
\end{aligned}
$$

Example: Consider the distribution function of a random variable X given by

$$
\begin{aligned}
F_X(x) &= 0, && x < 0 \\
&= \frac{1}{4} + \frac{1}{16}x && 0 \le x < 4 \\
&= \frac{3}{4} + \frac{1}{16}(x-4) && 4 \le x \le 8 \\
&= 1 && x > 8
\end{aligned}
$$

Express $F_X(x)$ as the distribution function of a mixed type random variable.

Solution:



The distribution function $F_X(x)$ is as shown in the figure.

Clearly $F_X(x)$ has jumps at $x = 0$ and $x = 4$.

$$\therefore p = p_X(0) + p_X(4) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

And

$$
\begin{aligned}
F_D(x) &= p_X(0)u(x) + p_X(4)u(x-4) \\
&= \frac{1}{2}u(x) + \frac{1}{2}u(x-4)
\end{aligned}
$$

$$F_C(x) = \begin{cases} 0 & \infty < x < 0 \\ x & 0 \le x \le 8 \\ 1 & x > 8 \end{cases}$$

Example 2:

X is the RV representing the life time of a device with the PDF $f_X(x)$ for x>0.

Define the following random variable

$$\begin{aligned} y &= X && \text{if} \quad X \le a \\ &= a && \text{if} \quad X > a \end{aligned}$$

$$S_D = a$$
$$S_C = (0, a)$$

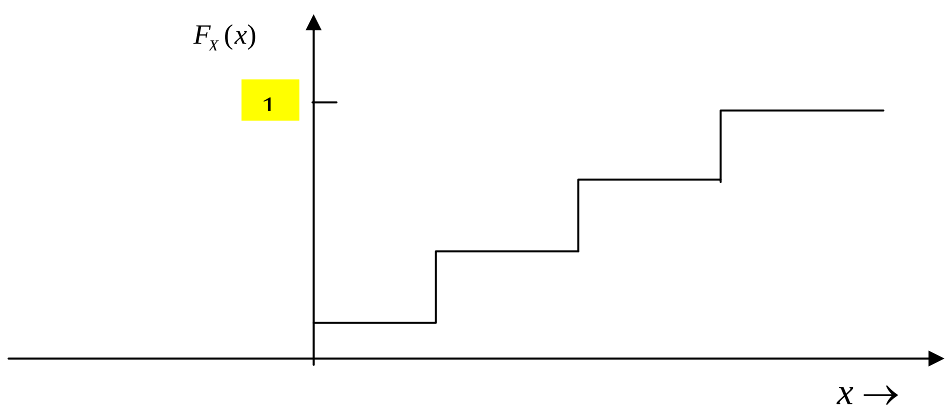
$$\begin{aligned} p &= P\{y \in D\} \\ &= P\{X > a\} \\ &= 1 - F_X(a) \end{aligned}$$

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

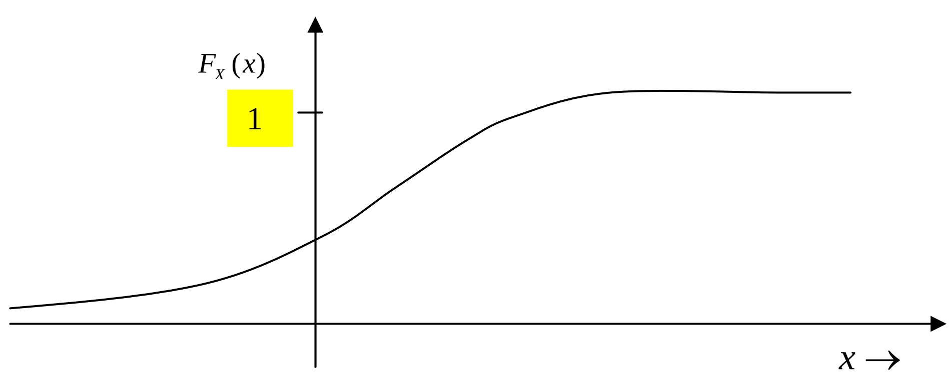
Discrete, Continuous and Mixed-type random variables

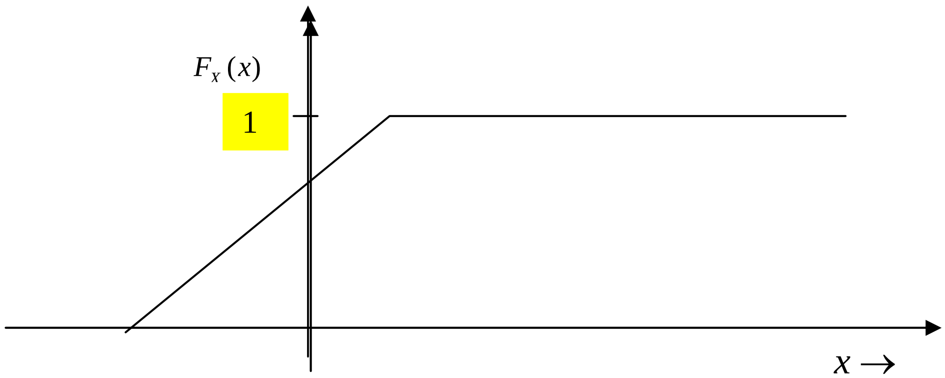
- A random variable X is called a discrete random variable $F_X(x)$ is piece-wise constant. Thus $F_X(x)$ is flat except at the points of jump discontinuity. If the sample space S is discrete the random variable X defined on it is always discrete.
- X is called a *continuous random variable* if $F_X(x)$ 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 it increases continuously at least at one interval of values of x. For a such type RV *X*,
  $F_X(x) = pF_{X_d}(x) + (1-p)F_{X_c}(x)$

where $F_{X_d}(x)$ is the distribution function of a discrete RV and $F_{X_c}(x)$ is the distribution function of a continuous RV. Typical plots of $F_X(x)$ for discrete, continuous and mixed-random variables are shown in Fig below.

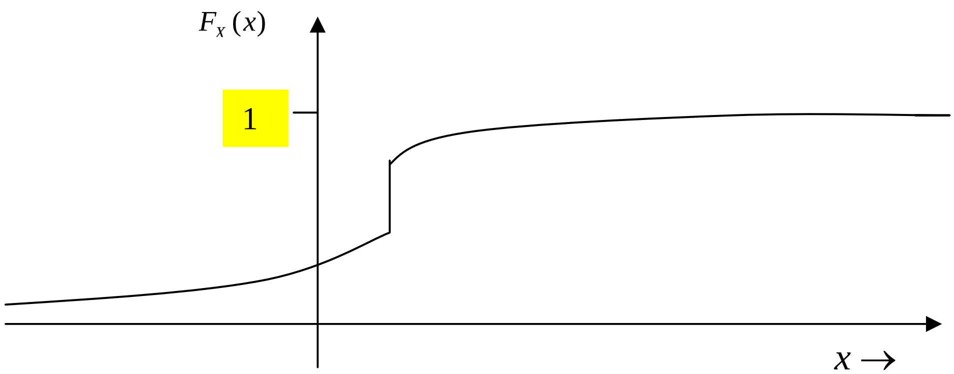


Plot $F_X(x)$ vs. x for a discrete random variable (to be animated)





Plot $F_X(x)$ vs. x **for a continuous random variable (to be animated)**



Plot $F_X(x)$ vs. x **for a mixed-type random variable (to be animated)**

Discrete Random Variables and Probability mass functions

A random variable is said to be *discrete* if the number of elements in the range of R_X is finite or countably infinite. Examples 1 and 2 are discrete random variables.
Assume R_X to be countably finite. Let $x_1, x_2, x_3 ..., x_N$ be the elements in R_X. Here the mapping $X(s)$ partitions S into N subsets $\{s \mid X(s) = x_i\}, i = 1, 2, ..., N.$
The discrete random variable in this case is completely specified by the *probability mass function* (pmf) $p_X(x_i) = P(s \mid X(s) = x_i), i = 1, 2, ..., N.$

Clearly,

- $p_X(x_i) \geq 0 \quad \forall x_i \in R_X$ and
- $\sum_{i \in R_X} p_X(x_i) = 1$
- Suppose $D \subseteq R_X$. Then

$$P(\{x \in D\}) = \sum_{x_i \in D} p_X(x_i)$$

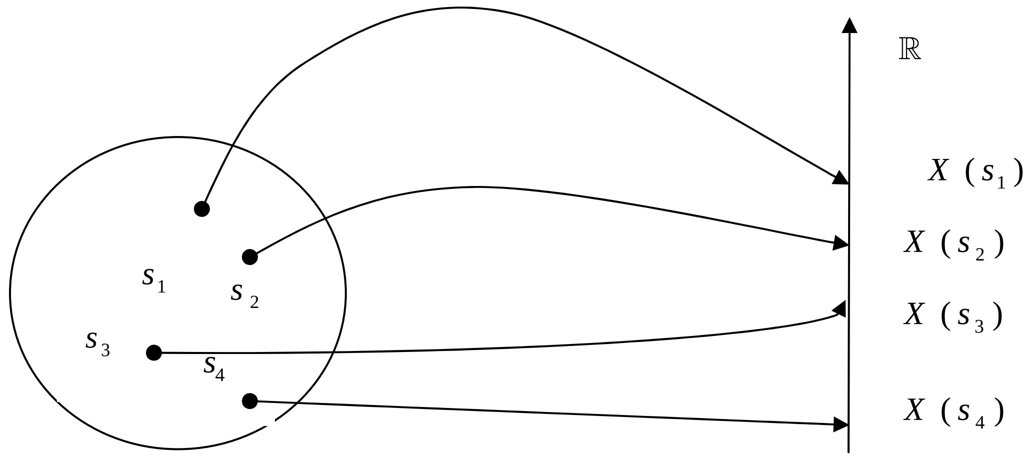


Figure Discrete Random Variable

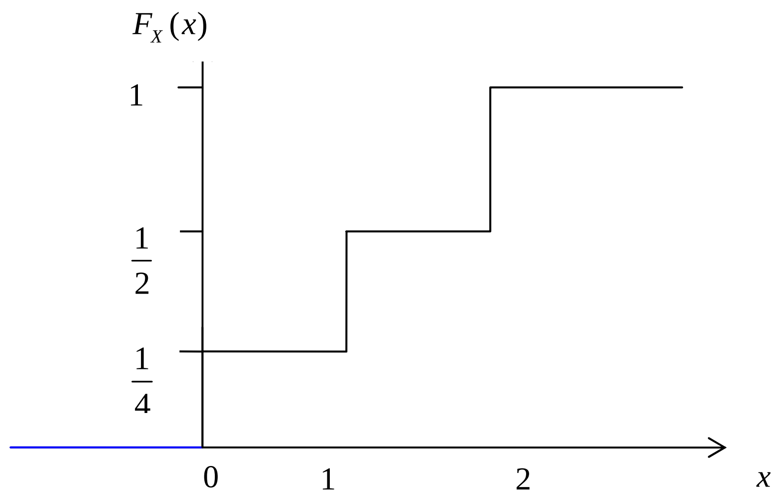
(To be animated)

Example

Consider the random variable X with the distribution function

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

The plot of the $F_X(x)$ is shown in Fig.



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	$\frac{1}{2}$

We shall describe about some useful discrete probability mass functions in a later class.

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

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

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 $f_X(x)$, us 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

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

Thus the probability of X lying in the 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

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

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

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

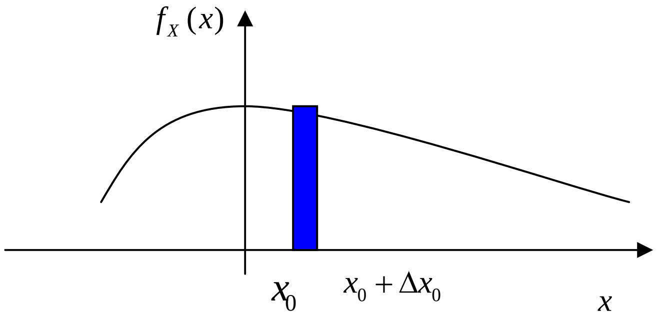


Fig. $P(\{x_0 < X \le x_0 + \Delta x_0\}) \simeq f_X(x_0)\Delta x_0$

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

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

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

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

Consider the random variable defined in Example 1 and Example 3. The distribution function $F_X(x)$ can be written as

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

and

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

Probability Density Function of Mixed-type Random Variable

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

$$F_X(x) = pF_{X_d}(x) + (1-p)F_{X_c}(x)$$

where $F_{X_d}(x)$ is the distribution function of a discrete RV and $F_{X_c}(x)$ is the distribution function of a discrete RV. Therefore,

$$f_X(x) = pf_{X_d}(x) + (1-p)f_{X_c}(x)$$

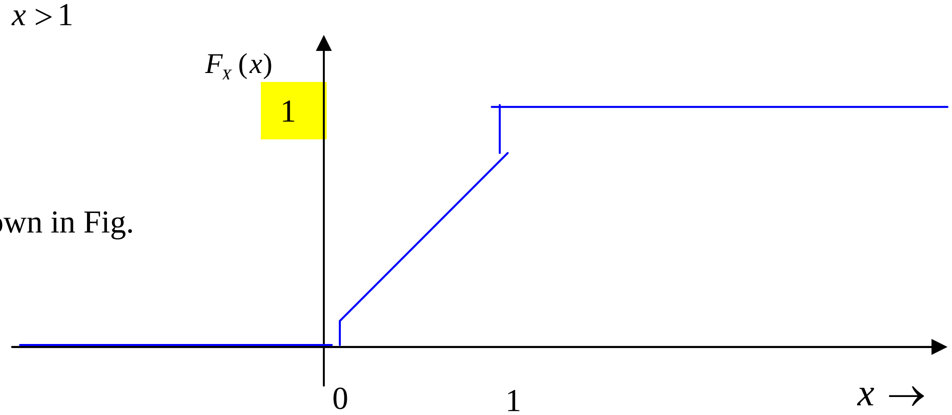
where

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

Example Consider the random variable X with the distribution function

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

The plot of $F_X(x)$ is shown in Fig.



$F_X(x)$ can be expressed as

$$F_X(x) = 0.2F_{X_d}(x) + 0.8F_{X_c}(x)$$

where

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

and

$$F_{X_c}(x) = \begin{cases} 0 & x < 0 \\ \text{x} & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

The pdf is given by

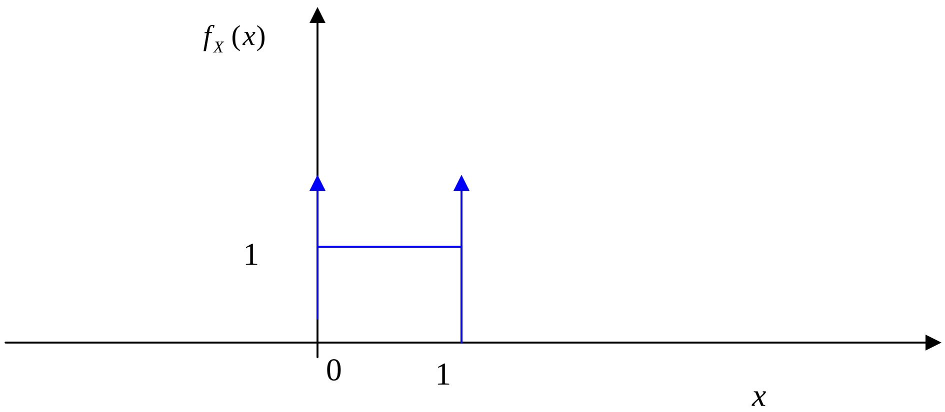
$$f_X(x) = 0.2f_{X_d}(x) + 0.8f_{X_c}(x)$$

where

$$f_{X_d}(x) = 0.5\delta(x) + 0.5\delta(x-1)$$

and

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



Functions of Random Variables

Often we have to consider random variables which are functions of other random variables. Let X be a random variable and $g(.)$ is function $\mathbb{R}$. Then $Y = g(X)$ is a random variable. We are interested to find the pdf of Y. For example, suppose X represents the random voltage input to a full-wave rectifier. Then the rectifier output Y is given by $Y = |X|$. We have to find the probability description of the random variable Y. We consider the following cases:

(a) X is a discrete random variable with probability mass function $p_X(x)$

The probability mass function of Y is given by

$$\begin{aligned} p_Y(y) &= P(\{Y = y\}) \\ &= P(\{x \mid g(x) = y\}) \\ &= \sum_{\{x \mid g(x)=y\}} P(\{X = x\}) \\ &= \sum_{\{x \mid g(x)=y\}} p_X(x) \end{aligned}$$

(b) X is a continuous random variable with probability density function $f_X(x)$ and $y = g(x)$ is one-to-one and monotonically increasing

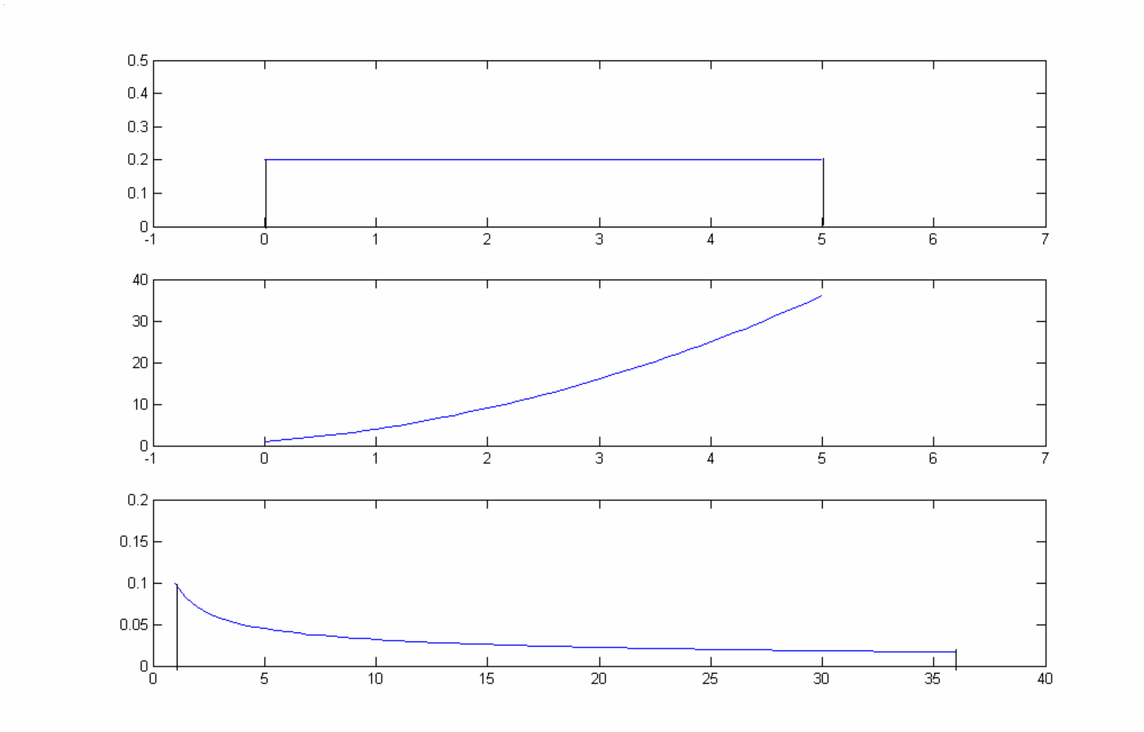
The probability distribution function of Y is given by

$$
\begin{aligned}
F_Y(y) &= P\{Y \le y\} \\
&= P\{g(X) \le y\} \\
&= P\{X \le g^{-1}(y)\} \\
&= P(\{X \le x\})\Big]_{x=g^{-}}(y) \\
&= F_X(x)\Big]_{x=g^{-1}(y)}
\end{aligned}
$$

$$
\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} \\
&= \frac{dF_X(x)}{dy}\Bigg]_{x=g^{-1}(y)} \\
&= \frac{dF_X(x)}{dx}\frac{dx}{dy}\Bigg]_{x=g^{-1}(y)} \\
&= \frac{\frac{dF_X(x)}{dx}}{\frac{dy}{dx}}\Bigg]_{x=g^{-1}(y)} \\
&= \frac{f_X(x)}{g'(x)}\Bigg]_{x=g^{-1}(y)}
\end{aligned}
$$

$$
f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}} = \frac{f_X(x)}{g'(x)}\Bigg]_{x=g^{-1}(y)}
$$

This is illustrated in Fig.



Example 1: Probability density function of a linear function of random variable

Suppose $Y = aX + b,\ a > 0.$

Then $x = \dfrac{y-b}{a}$ and $\dfrac{dy}{dx} = a$

$$\therefore\ f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}} = \frac{f_X(\frac{y-b}{a})}{a}$$

Example 2: Probability density function of the distribution function of a random variable

Suppose the distribution function $F_X(x)$ of a continuous random variable X is monotonically increasing and one-to-one and define the random variable $Y = F_X(X).$ Then, $f_Y(y) = 1\ \ 0 \le y \le 1.$

$y = F_X(x)$

Clearly $0 \le y \le 1$

$$\frac{dy}{dx} = \frac{dF_X(x)}{dx} = f_X(x)$$

$$\therefore\ f_Y(y) = \frac{f_X(x)}{\frac{dy}{dx}} = \frac{f_X(x)}{f_X(x)} = 1$$

$$\therefore\ f_Y(y) = 1\ \ 0 \le y \le 1.$$

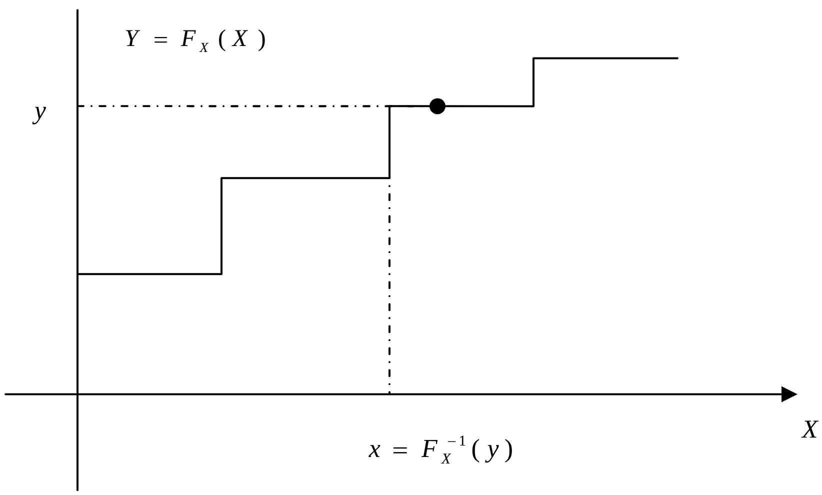
Remark

(1) The distribution given by $f_Y(y) = 1 \quad 0 \le y \le 1$ is called a uniform distribution over the interval [0,1].

(2) The above result is particularly important in simulating a random variable with a particular distribution function. We assumed $F_X(x)$ to be one-to-one function for invariability. However, the result is more general- *the random variable defined by the distribution function of any random variable is uniformly distributed over* [0,1]. For example, if X is a discrete RV,

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(F_X(x) \le y) \\
&= P(X \le F_X^{-1}(y)) \\
&= F_X(F_X^{-1}(y)) \\
&= y \text{ (Assigning } F_X^{-1}(y) \text{ to the left-most point of the interval for which } F_X(x) = y).
\end{aligned}
$$

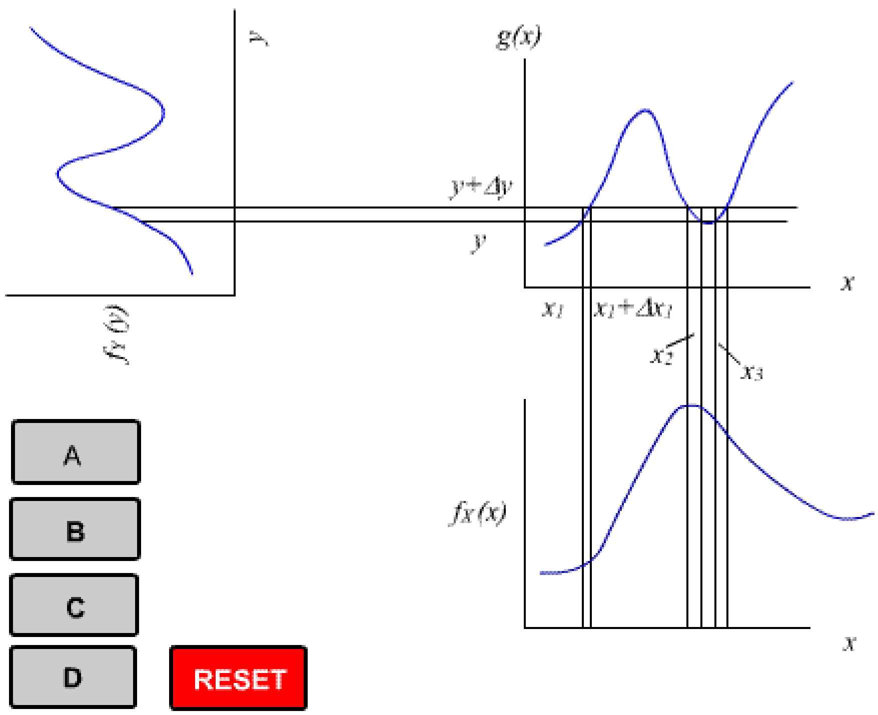
$$\therefore f_Y(y) = \frac{dF_Y(y)}{dy} = 1 \quad 0 \le y \le 1.$$



(c) X is a continuous random variable with probability density function $f_X(x)$ and $y = g(x)$ has multiple solutions for x

Suppose for $y \in Y, y = g(x)$ has solutions $x_i, i = 1, 2, 3, ..., n$. Then

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x)}{\left|\frac{dy}{dx}\right|_{x=x_i}}$$



Proof:

Consider the plot of $Y = g(X)$. Suppose at a point $y = g(x)$, we have three distinct roots as shown. Consider the event $\{y < Y \le y + dy\}$. This event will be equivalent to union events

$\{x_1 < X \le x_1 + dx_1\}, \{x_2 - dx_2 < X \le x_2\}$ and $\{x_3 < X \le x_3 + dx_3\}$

$$\therefore P\{y < Y \le y + dy\} = P\{x_1 < X \le x_1 + dx_1\} + P\{x_2 - dx_2 < X \le x_2\} + P\{x_3 < X \le x_3 + dx_3\}$$

$$\therefore f_Y(y)dy = f_X(x_1)dx_1 + f_X(x_2)(-dx_2) + f_X(x_3)dx_3$$

Where the negative sign in $-dx_2$ is used to account for positive probability.

Therefore, dividing by *dy* and taking the limit, we get

$$f_Y(y) = f_X(x_1)\left(\frac{dx_1}{dy}\right) + f_X(x_2)\left(-\frac{dx_2}{dy}\right) + f_X(x_3)\left(\frac{dx_3}{dy}\right)$$

$$= f_X(x_1)\left|\frac{dx_1}{dy}\right| + f_X(x_2)\left|\frac{dx_2}{dy}\right| + f_X(x_3)\left|\frac{dx_3}{dy}\right|$$

$$= \sum_{i=1}^{3} \frac{f_X(x_i)}{\left|\frac{dy}{dx}\right|_{x=x_i}}$$

In the above we assumed $y = g(x)$ to have three roots. In general, if $y = g(x)$ has n roots, then

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{\left|\frac{dy}{dx}\right|_{x=x_i}}$$

Example 3: Probability density function of a linear function of random variable

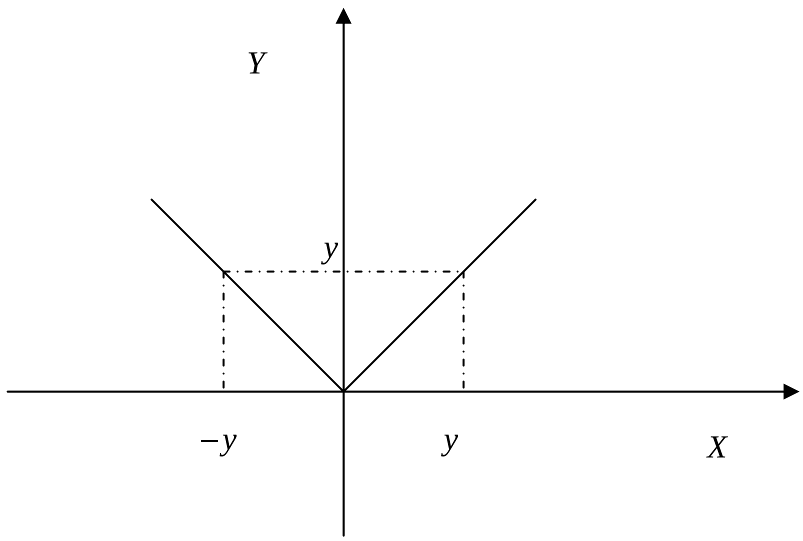
Suppose $Y = aX + b,\ a \neq 0.$

Then $x = \frac{y-b}{a}$ and $\frac{dy}{dx} = a$

$$\therefore\ f_Y(y) = \frac{f_X(x)}{\left|\frac{dy}{dx}\right|} = \frac{f_X(\frac{y-b}{a})}{|a|}$$

Example 4: Probability density function of the output of a full-wave rectifier

Suppose $Y = |X|, \quad -a \leq X \leq a, \qquad a > 0$



$y = |x|$ has two solutions $x_1 = y$ and $x_2 = -y$ and $\left|\frac{dy}{dx}\right| = 1$ at each solution point.

$$\therefore f_Y(y) = \frac{f_X(x)\big]_{x=y}}{1} + \frac{f_X(x)\big]_{x=-y}}{1}$$
$$= f_X(x) + f_X(-x)$$

Example 5: Probability density function of the output a square-law device

$Y = CX^2, C \geq 0$

$\therefore y = cx^2 \qquad \Rightarrow x = \pm\sqrt{\frac{y}{c}} \qquad y \geq 0$

And $\frac{dy}{dx} = 2cx$ so that $\left|\frac{dy}{dx}\right| = 2c\sqrt{y/c} = 2\sqrt{cy}$

$$\therefore f_Y(y) = \frac{f_X\left(\sqrt{y/c}\right) + f_X\left(\sqrt{-y/c}\right)}{2\sqrt{cy}} \qquad y > 0$$

$= 0$ otherwise

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.

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 denoted by μ_X.

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

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

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

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

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

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

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

Interpretation of the mean

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

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

$$= \frac{\int_{-\infty}^{\infty} x f_X(x)dx}{\int_{-\infty}^{\infty} f_X(x)dx} \qquad \because \int_{-\infty}^{\infty} f_X(x)dx = 1$$

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

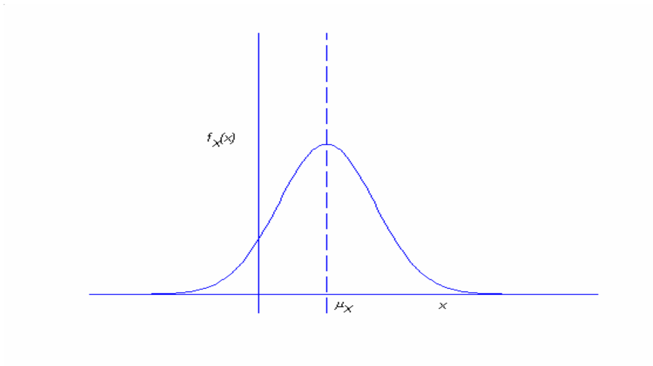


Fig. Mean of a random variable

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

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

Then

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

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

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

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

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

Expected value of a function of a random variable

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

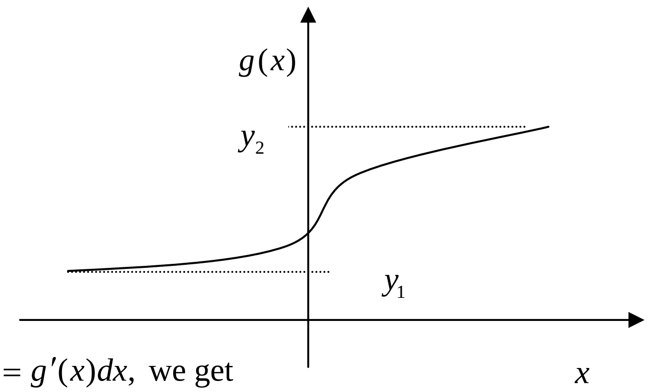
Then, $EY = Eg(X) = \int_{-\infty}^{\infty} g(x) f_X(x)dx$

We shall illustrate the theorem in the special case $g(X)$ when $y = g(x)$ is one-to-one and monotonically increasing function of x. In this case,

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

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

$$= \int_{y_1}^{y2} y \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} dy$$



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,

$$\boxed{Ec = c}$$

Clearly

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

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

$$\boxed{E[c_1 g_1(X) + c_2 g_2(X)] = c_1 Eg_1(X) + c_2 Eg_2(X)}$$

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

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

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

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

The above property means that E is a linear operator.

Mean-square value

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

Variance

For a random variable X with the pdf $f_X(x)$ and men 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 $p_X(x_i), i = 1, 2, ..., N,$

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

The standard deviation of X is defined as

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

Example 3 Find the variance of the random variable discussed in Example 1.

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

Example 4 Find the variance of the random variable discussed in Example 2.

As already computed

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

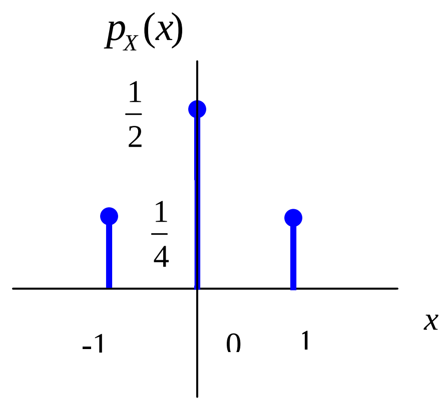
$$\begin{aligned}
\sigma_X^2 &= E(X-\mu_X)^2 \\
&= (0-\frac{17}{8})^2\times\frac{1}{8} + (1-\frac{17}{8})^2\times\frac{1}{8} + (2-\frac{17}{8})^2\times\frac{1}{4} + (3-\frac{17}{8})^2\times\frac{1}{2} \\
&= \frac{117}{128}
\end{aligned}$$

Remark

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

For example, consider two random variables X_1 and X_2 with pmf as shown below. Note that each of X_1 and X_2 has zero means. $\sigma_{X_1}^2=\frac{1}{2}$ and $\sigma_{X_2}^2=\frac{5}{3}$ implying that X_2 has more spread about the mean.

x	-1	0	1
$p_{X_1}(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$



x	-2	-1	0	1	2
$p_{X_2}(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

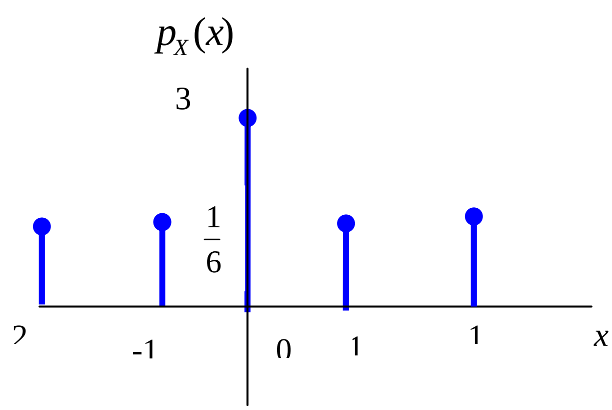
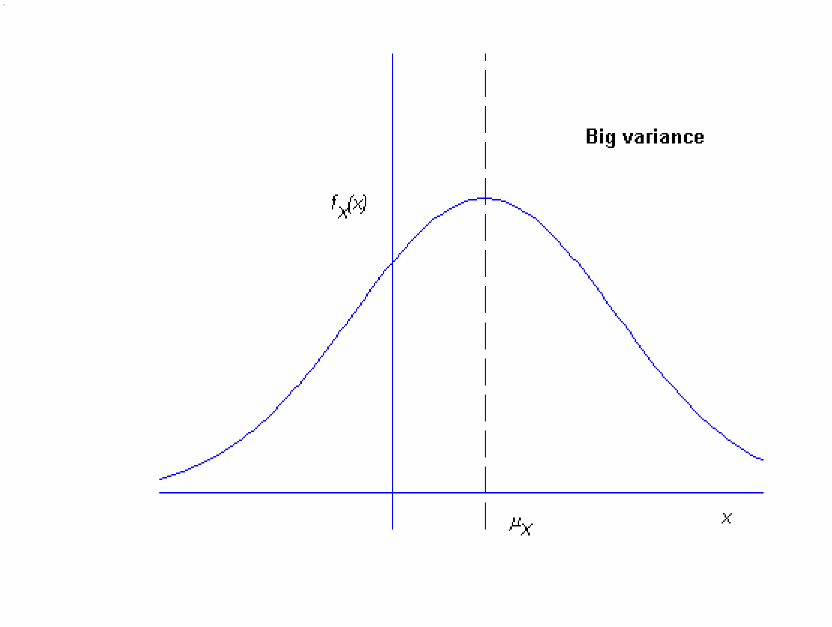
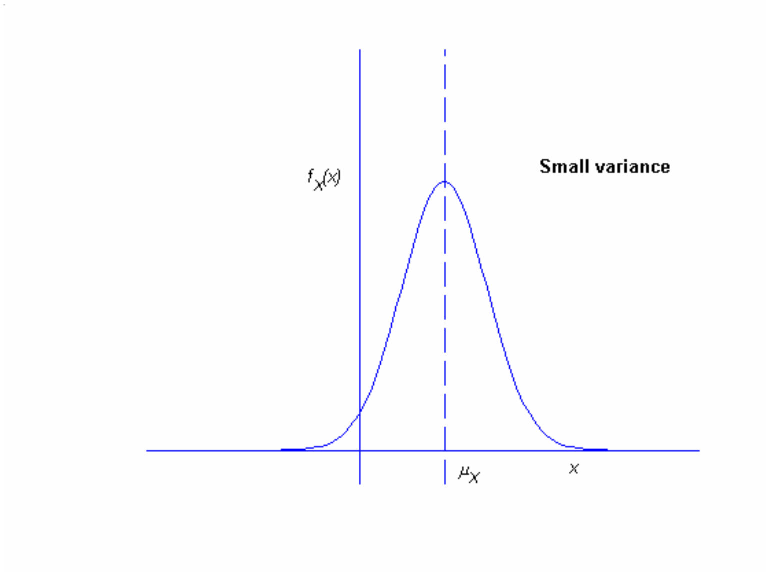


Fig. shows the pdfs of two continuous random variables with same mean but different variances





- We could have used the *mean absolute deviation* $E|X - \mu_X|$ for the same purpose. But it is more difficult both for analysis and numerical calculations.

Properties of variance

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

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

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

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

(3) If c is a constant,

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

nth **moment of a random variable**

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

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

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

Note that

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

Chebysev Inequality

Suppose X is 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 $2\sigma_X$. What fraction of the manufacturing item does the quality control department reject? Can you roughly guess it?

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

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

<u>Proof:</u>

$$\begin{aligned}
\sigma_x^2 &= \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx \\
&\geq \int_{|X-\mu_X|\geq\varepsilon} (x-\mu_X)^2 f_X(x)dx \\
&\geq \int_{|X-\mu_X|\geq\varepsilon} \varepsilon^2 f_X(x)dx \\
&= \varepsilon^2 P\{|X-\mu_X| \geq \varepsilon\}
\end{aligned}$$

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

Markov Inequality

For a random variable X which take only nonnegative values

$$P\{X \geq a\} \leq \frac{E(X)}{a} \qquad \text{where } a > 0.$$

$$E(X)=\int_0^{\infty} xf_X(x)dx$$
$$\geq \int_a^{\infty} xf_X(x)dx$$
$$\geq \int_a^{\infty} af_X(x)dx$$
$$= aP\{X \geq a\}$$

$$\therefore \quad P\{X \geq a\} \leq \frac{E(X)}{a}$$

Remark: $P\{(X-k)^2 \geq a\} \leq \dfrac{E(X-k)^2}{a}$

Example

Example A nonnegative RV X has the mean $\mu_X = 1$. Find an upper bound of the probability $P(X \geq 3\})$.

By Markov's inequality

$$P(X \geq 3\}) \leq \frac{E(X)}{3} = \frac{1}{3}.$$

Hence the required upper bound $= \dfrac{1}{3}$.

Characteristic Functions of Random Variables

Just as the frequency-domain charcterisations of discrete-time and continuous-time signals, the probability mass function and the probability density function can also be characterized in the frequency-domain by means of the *charcteristic function* of a random variable. These functions are particularly important in

- calculating of moments of a random variable
- evaluating the PDF of combinations of multiple RVs.

Characteristic function

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

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

$$= \int_{-\infty}^{\infty} e^{j\omega x} f_X(x)dx$$

where $j = \sqrt{-1}.$

Note the following:

- $\phi_X(\omega)$ is a complex quantity, representing the Fourier transform of $f(x)$ 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.
- As $\phi_X(\omega)$ always +ve and $\int_{-\infty}^{\infty} f_X(x)dx = 1$, $\phi_X(\omega)$ always exists.

[Recall that the Fourier transform of a function *f(t)* exists if $\int_{-\infty}^{\infty} \left[f(t) \right] dt < \infty$, i.e., *f(t)* is absolutely integrable.]

We can get $f_X(x)$ from $\phi_X(\omega)$ by the inverse transform

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

Example 1 Consider the random variable *X* with pdf $f_X(x)$ given by

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

= 0 otherwise. The characteristics function is given by

Solution:

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

$$= \frac{1}{b-a} \quad \frac{e^{j\omega x}}{j\omega}\Big]_a^b$$

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

Example 2 The characteristic function of the random variable X with $f_X(x) = \lambda e^{-\lambda x} \quad \lambda > 0, x > 0$ 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 $R_X = \{x_1, x_2, ...\}$ with corresponding probability mass function $p_X(x_i)$ for the value x_i.

Then

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

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

Note that $\phi_X(\omega)$ can be interpreted as the discrete-time Fourier transform with $e^{j\omega x_i}$ substituting $e^{-j\omega x_i}$ in the original discrete-time Fourier transform. The inverse relation is

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

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

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

Then $\phi_X(\omega) = \sum_{k=0}^{n} {}^nC_k\, p^k\,(1-p)^{n-k}\, e^{j\omega k}$

$$= \sum_{k=0}^{n} {}^nC_k \left(pe^{j\omega}\right)^k (1-p)^{n-k}$$

$$= \left[pe^{j\omega} + (-p)\right]^n$$ (Using the Binomial theorem)

Example 4 The characteristic function of the discrete random variable X with

$$p_X(k) = p(1-p)^k, \qquad k = 0,1,....$$

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

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

$$= \frac{p}{1-(1-p)e^{j\omega}}$$

Moments and the characteristic function

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

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

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

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

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

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

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

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

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

Example 3 First two moments of random variable in Example 2

$$\phi_X(\omega) = \frac{\lambda}{\lambda - j\omega}$$

$$\frac{d}{d\omega}\phi_X(\omega) = \frac{\lambda j}{(\lambda - j\omega)^2}$$

$$\frac{d^2}{d\omega^2}\phi_X(\omega) = \frac{2\lambda j^2}{(\lambda - j\omega)^3}$$

$$EX = \left.\frac{1}{j}\frac{\lambda j}{(\lambda - j\omega)^2}\right|_{\omega=0} = \frac{1}{\lambda}$$

$$EX^2 = \frac{1}{j^2}\frac{2\lambda j^2}{(\lambda - j\omega)^3}_{\omega=0} = \frac{2}{\lambda^2}$$

Probability generating function:

If the random variable under consideration takes non negative integer values only, it is convenient to characterize the random variable in terms of the probability generating function G (z) defined by

$$G_X(z) = Ez^X$$

$$= \sum_{k=0}^{\infty} p_X(k) z$$

Note that

- $G_X(z)$ is related to z-transform, in actual z-transform, z^{-k} is used instead of z^k .
- The characteristic function of *X* is given by $\phi_X(\omega) = G_X(e^{j\omega})$
- $G_X(1) = \sum_{k=0}^{\infty} p_X(k) = 1$

- $G_X\,'(z) = \sum_{k=0}^{\infty} kp_X(k) z^{k-1}$
- $G\,'(1) = \sum_{k=0}^{\infty} kp_X(k) = EX$
- $G_X\,''(z) = \sum_{k=0}^{\infty} k(k-1)p_X(k)\,z^{k-2} = \sum_{k=0}^{\infty} k^2 p_x(k)\,z^{k-2} - \sum_{k=0}^{\infty} kp_x(k)\,z^{k-2}$

$$\therefore G_X\,''(1) = \sum_{k=0}^{\infty} k^2 p_X(k) - \sum_{k=0}^{\infty} kp_X(k) = EX^2 - EX$$

$$\therefore \sigma_X{}^2 = EX^2 - (EX)^2 = G_X\,''(1) + G_X\,'(1) - (G_X\,'(1))^2$$

Ex: Binomial distribution:

$$p_m(x) = {}^nC_x p^x (1-p)^x$$
$$\therefore \phi_x(z) = \sum_x {}^nC_x p^x (1-p)^{n-x} z^x$$
$$= \sum_x {}^nC_x (pz)^x (1-p)^{n-x}$$
$$= (1-p+pz)^n$$

$$\phi'_X(1) = EX = np$$

$$\phi_X{}''(1) = EX^2 - EX = n(n-1)p^2$$

$$\therefore EX^2 = \phi_X{}''(1) + EX$$
$$= n(n-1)p^2 + np$$
$$= np^2 + npq$$

Example 2: Geometric distribution

$$p_X(x) = p(1-p)^x$$

$$\phi_X(z) = \sum_x p(1-p)^x z^x$$
$$= p\sum_x ((1-p)z)^x$$
$$= p\frac{1}{1-(1-p)z}$$

$$\phi_X{}'(z) = +\frac{p(1-p)}{(1-(1-p)z)^2}$$

$$\phi_X{}'(1) = \frac{p(1-p)}{(1-1+p)^2} = \frac{p(1-p)}{p^2} = \frac{q}{p}$$

$$\phi_X{}''(z) = \frac{2p(1-p)(1-p)}{(1-(1-p)z)^3}$$

$$\phi_X{}''(1) = \frac{2pq^2}{p^3} = 2\left(\frac{q}{p}\right)^2$$

$$EX^2 = \phi''(1) + \frac{q}{p} = 2\left(\frac{q}{p}\right)^2 + \frac{q}{p}$$

$$Var(X) = 2\left(\frac{q}{p}\right)^2 + \left(\frac{q}{p}\right) - \left(\frac{q}{p}\right)^2$$

Moment Generating Function:

Sometimes it is convenient to work with a function similarly to the Laplace transform and known as the moment generating function.

For a random variable X, the moment generating function $M_X(s)$ is defined by

$$M_X(s) = Ee^{SX}$$
$$= \int_{R_X} f_X(x) e^{SX} dx$$

Where R_X is the range of the random variable X.

If *X* is a non negative continuous random variable, we can write

$$M_X(s) = \int_0^\infty f_X(x) e^{SX} dx$$

Note the following:

- $$M_x{}'(s) = \int_0^\infty x f_x(x) e^{sx} dx$$

$$\therefore M'(0) = EX$$

- $$\frac{d^k}{ds^k} M_X(s) = \int_0^\infty x^k f_X(x) e^{sx} dx$$

$$= EX^k$$

Example Let *X* be a continuous random variable with

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

Then

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

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

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

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

$$=\left(\frac{q}{p}\right)^2+\left(\frac{q}{p}\right)$$

The joint characteristic function of two random variables X and Y is defined by,

$$\phi_{X,Y}(\omega_1,\omega_2)=Ee^{j\omega_1x+j\omega_2y}$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,y}(x,y)e^{j\omega_1x+j\omega_2y}dydx$$

And the joint moment generating function of $\phi(s_1,s_2)$ is defined by,

$$\phi_{X,Y}(s_1,s_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)e^{xs_1+ys_2}dxdy$$

$$=Ee^{s_1x+s_2y}$$

IF $Z=ax+by$, then

$$\phi_Z(s)=Ee^{Zs}=Ee^{(ax+by)s}=\phi_{X,Y}(as,bs)$$

Suppose *X* and *Y* are independent. Then

$$\phi_{X,Y}(s_1,s_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{s_1x+s_2x} f_{X,Y}(x,y)dydx$$

$$= \int_{-\infty}^{\infty} e^{s_1x} f_X(x)dx \int_{-\infty}^{\infty} e^{s_2y} f_Y(y)dy$$

$$= \phi_1(s_1)\phi_2(s_2)$$

Particularly if *Z=X+Y* and *X* and *Y* are independent, then

$$\phi_Z(s) = \phi_{X,Y}(s,s)$$

$$= \phi_X(s)\phi_Y(s)$$

Using the property of Laplace transformation we get,

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

Let us recall the MGF of a Gaussian random variable

$$X = N(\mu_X, \sigma_X{}^2)$$

$$\phi_X(s) = Ee^{Xs}$$

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

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

$$= \frac{1}{\sqrt{2\pi}\sigma_X}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\mu_X-\sigma_X{}^2s)} dx + e^{+\frac{1}{2}\frac{(\sigma_X{}^4s^2+2\mu_X\sigma_X{}^2s)}{\sigma_X{}^2}} dx$$

We have,

$$\phi_{X,Y}(s_1,s_2)$$

$$= Ee^{Xs_1+Ys_2}$$

$$= E(1 + Xs_1 + Ys_2 + \frac{(Xs_1+Ys_2)}{2} + ...)$$

$$= 1 + s_1EX + s_2EY + \frac{s_1{}^2EX^2}{2} + \frac{s_2{}^2EY^2}{2} + s_1s_2EXY$$

Hence, $EX = \frac{\partial}{\partial s_1}\phi_{X,Y}(s_1,s_2)]_{s_1=0}$

$$EY = \frac{\partial}{\partial s_2}\phi_{X,Y}(s_1,s_2)]_{s_2=0}$$

$$EXY = \frac{\partial^2}{\partial s_1 \partial s_2} \phi_{X,Y}(s_1, s_2)]_{s_1=0, s_2=0}$$

We can generate the joint moments of the RVS from the moment generating function.

Important Discrete Random Variables

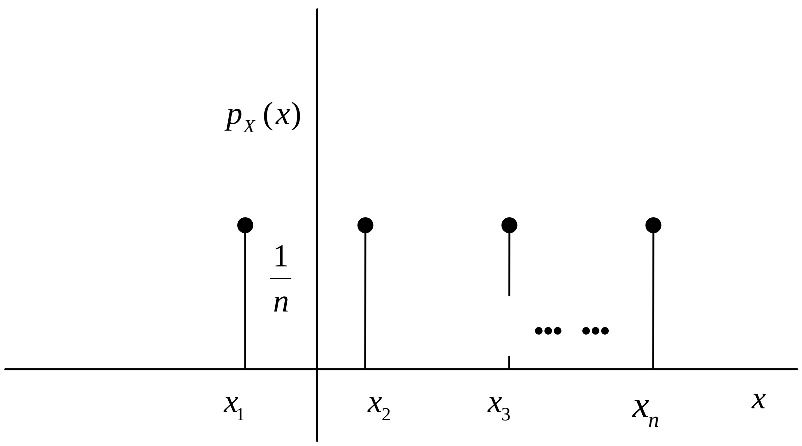
Discrete uniform random variable

A discrete random variable X is said to be a uniform random variable if it assumes each of the values $x_1, x_2, ...x_n$ with equal probability. The probability mass function of the uniform random variable X is given by

$$p_X(x_i) = \frac{1}{n}, \quad i = 1, 2, ...n$$

Its CDF is

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



…

Mean and variance of the Discrete uniform random variable

$$\mu_X = EX = \sum_{i=0}^{n} x_i p_X(x_i)$$
$$= \frac{1}{n}\sum_{i=0}^{n} x_i$$
$$EX^2 == \sum_{i=0}^{n} x_i^2 p_X(x_i)$$
$$= \frac{1}{n}\sum_{i=0}^{n} x_i^2$$
$$\therefore \sigma_X^2 = EX^2 - \mu_X^2$$
$$= \frac{1}{n}\sum_{i=0}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=0}^{n} x_i\right)^2$$

Example: Suppose X is the random variable representing the outcome of a single roll of a fair dice. Then X can assume any of the 6 values in the set {1, 2, 3, 4, 5, 6} with the probability mass function

$$p_X(x) = \frac{1}{6} \qquad x = 1,2,3,4,5,6$$

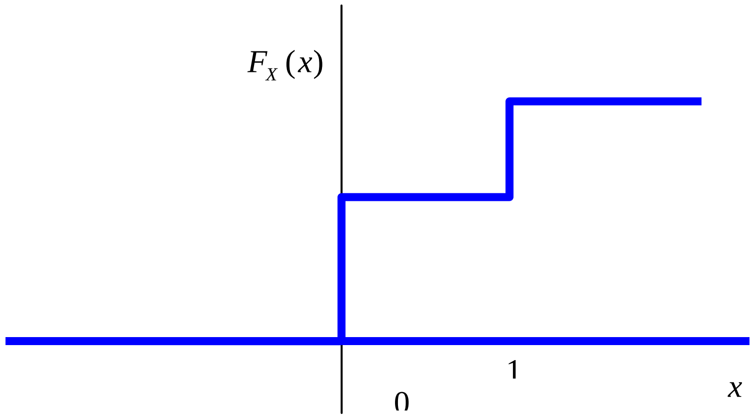
Bernoulli random variable

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 Fig.



Remark We can define the pdf of X with the help of delta function. Thus
$f_X(x) = (1-p)\delta(x) + p\delta(x)$

Example 1: 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, $EX^m = p \quad m = 1, 2, 3....$ 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 $\{0, 1, ..., n\}$. X is called a binomial random variable with parameters n and $0 \le p \le 1$ if

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

where

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

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.
- $p_X(k)$, k=0,1,...,n k), defines a valid probability mass function. This is because $\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} {}^nC_k\, p^k (1-p)^{n-k} = [p + (1-p)]^n = 1.$
- 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: In a binary communication system, the probability of bit error is 0.001. 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) all the bits will bit 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) P(exactly 2 bit errors will occur) $= p_X(2)$

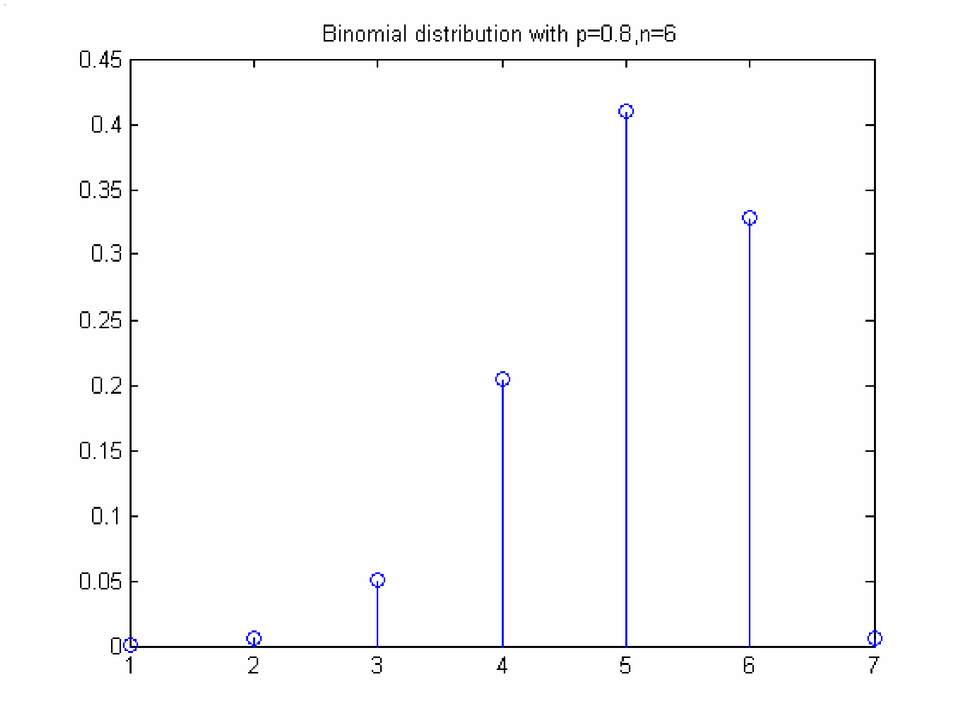
$$= {}^8C_2 \times 0.01^2 \times 0.99^6$$

(b) P(at least 2 bit errors will occur)=$p_X(0) + p_X(1) + p_X(2)$

$$=0.99^8 + {}^8C_1 \times 0.01^1 \times 0.99^7 + {}^8C_2 \times 0.01^2 \times 0.99^6$$

(c) P(all 8 bits will be erroneous)=$p_X(8) = 0.01^8 = 10^{-6}$

The probability mass function for a binomial random variable with n = 6 and p =0.8 is shown in the figure below.



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

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

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

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

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

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

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

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

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

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

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

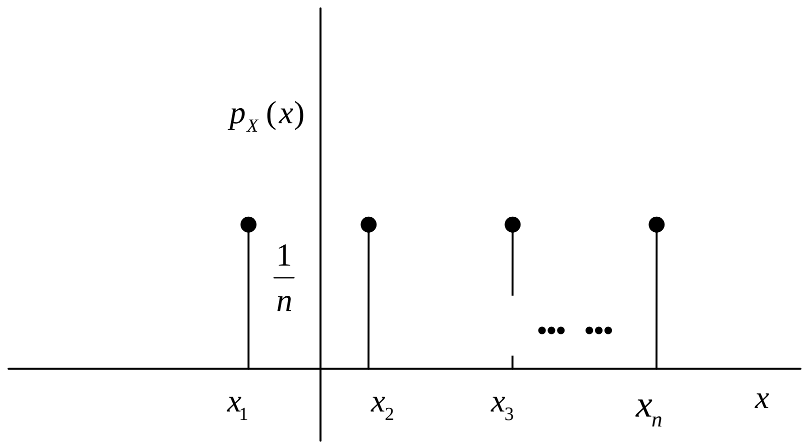
Mean of $B(n-1, p)$

Important Random Variables

Discrete uniform random variable

A discrete random variable X is said to be a uniform random variable if it assumes each of the values $x_1, x_2, ...x_n$ with equal probability. The probability mass function of the uniform random variable X is given by

$$p_X(x_i) = \frac{1}{n}, \qquad i = 1, 2, ...n$$



…

Example: Suppose X is the random variable representing the outcome of a single roll of a fair dice. Then X can assume any of the 6 values in the set {1, 2, 3, 4, 5, 6} with the probability mass function

$$p_X(x) = \frac{1}{6} \qquad x = 1, 2, 3, 4, 5, 6$$

Bernoulli random variable

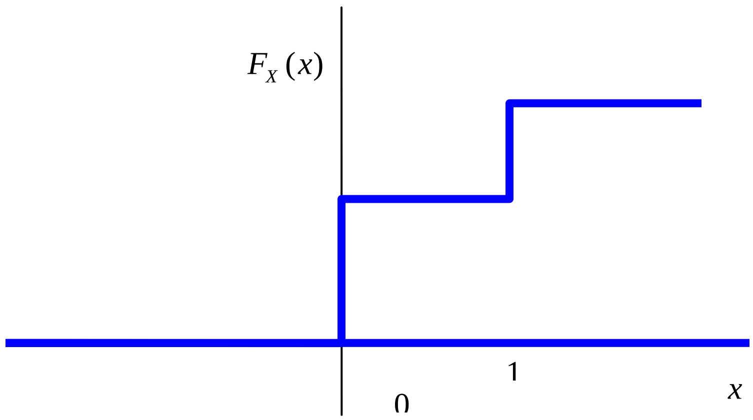
Suppose X is a random variable that takes two values 0 and 1, with probability mass functions

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

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 Fig.



Remark We can define the pdf of X with the help of delta function. Thus
$f_X(x) = (1-p)\delta(x) + p\delta(x)$

Example 1: 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, $EX^m = p \quad m = 1, 2, 3....$ 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 $\{0, 1, ..., n\}$. X is called a binomial random variable with parameters n and $0 \leq p \leq 1$ if

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

where

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

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.
- $p_X(k)$, k=0,1,...,n k), defines a valid probability mass function. This is because $\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} {}^nC_k\, p^k (1-p)^{n-k} = [p + (1-p)]^n = 1.$
- 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: In a binary communication system, the probability of bit error is 0.001. 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) all the bits will bit 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) P(exactly 2 bit errors will occur) = $p_X(2)$

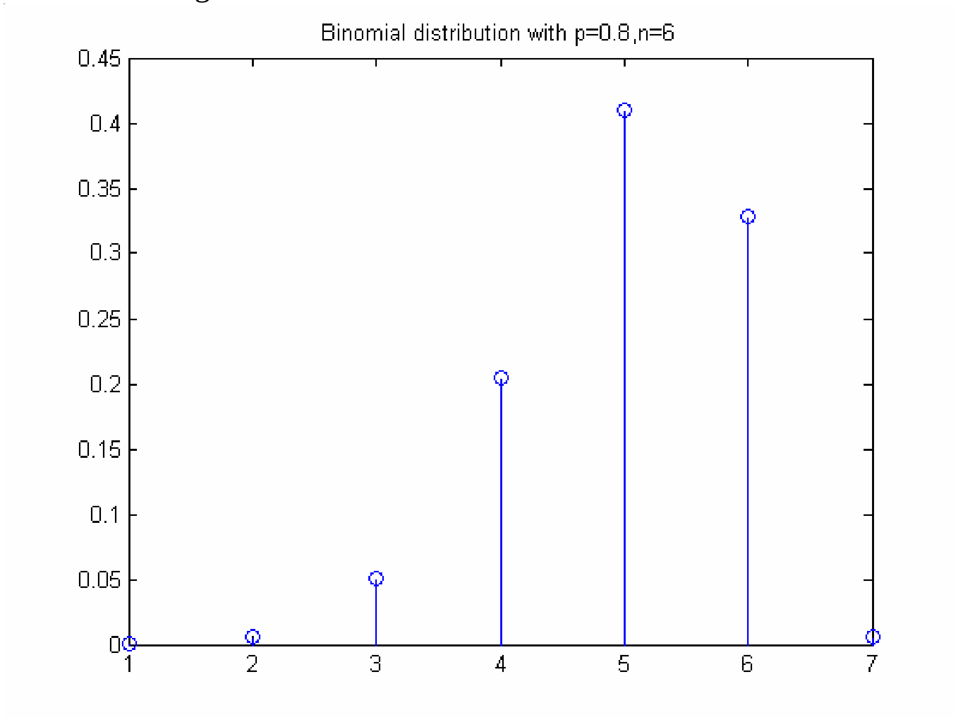
$$= {}^8C_2 \times 0.01^2 \times 0.99^6$$

(b) P(at least 2 bit errors will occur)=$p_X(0) + p_X(1) + p_X(2)$

$$=0.99^8 + {}^8C_1 \times 0.01^1 \times 0.99^7 + {}^8C_2 \times 0.01^2 \times 0.99^6$$

(c) P(all 8 bits will be erroneous)=$p_X(8) = 0.01^8 = 10^{-6}$

The probability mass function for a binomial random variable with n = 6 and p =0.8 is shown in the figure below.



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

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

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

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

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

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

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

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

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

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

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

Mean of $B(n-1, p)$

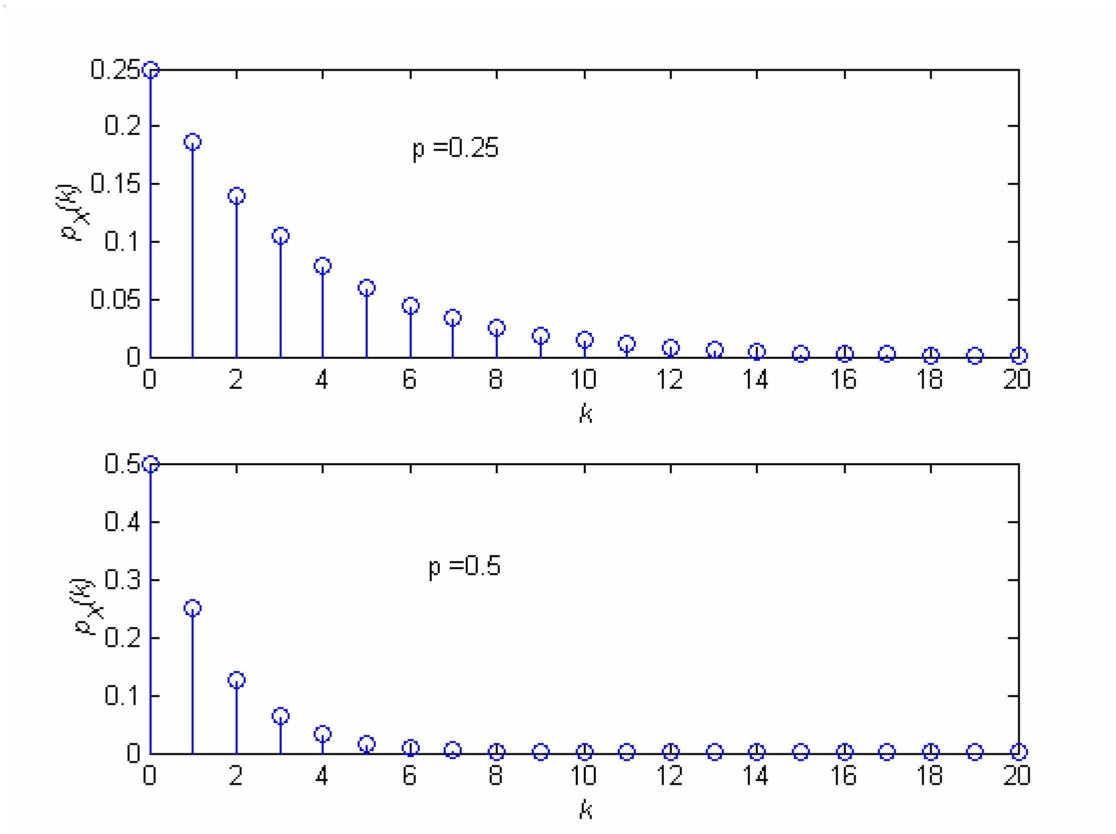
Geometric random variable:

X as a discrete random variable with range $R_X = \{1, 2, ...\}$. X is called a geometric random variable of $p_X(k) = p(1-p)^{k-1} \qquad 0 \le p \le 1$

- X describes the outcomes of independent Bernoulli trials each with probability of success p, before a 'success' occurs. If the first success occurs at the kth trial, then the outputs of the Bernoulli trials are
  $F\underbrace{ailiure, Failure, .., Failure}_{k-1 \text{ times}}, \ Success$
  $\therefore p_X(k) = (1-p)^{k-1} p = p(1-p)^{k-1}$
- R_X is countablly infinite, because we may have to wait infinitely long before the first success occurs.
- The geometric random variable X with the parameter p is denoted by $X \sim geo(p)$
- The CDF of $X \sim geo(p)$ is given by
  $F_X(k) = \sum_{i=1}^{k} (1-p)^{i-1} p = 1-(1-p)^k$

which gives the probability that the first 'success' will occur before the $(k+1)th$ trial.

Following Figure shows the pmf of a random variable $X \sim geo(p)$ for $p = 0.25$ and $p = 0.5$ respectively. Observe that the plots have a mode at $k = 1$.



Example:
Suppose X is the random variable representing the number of independent tossing of a coin before a head shows up. Clearly X will be a geometric random variable.
Example: A fair dice is rolled repeatedly. What is the probability that a 6 will be shown before the fourth roll.

Suppose X is the random variable representing the number of independent rolling of the dice before a '6' shows up. Clearly X will be a geometric random variable with $p = \frac{1}{6}$..

P**(a '6' will be shown before the 4th roll)**=$P(X = 1$ **or** $X = 2$ **or** $X = 3)$

$$
\begin{aligned}
&= p_X(1) + p_X(2) + p_X(3) \\
&= p + p(1-p) + p(1-p)^2 \\
&= p(3 - 3p + p^2) \\
&= \frac{91}{196}
\end{aligned}
$$

Mean and variance of the Geometric random variable

$$\text{Mean}=EX=\sum_{k=0}^{\infty} k p_X(k)$$

$$=\sum_{k=0}^{\infty} k p(1-p)^{k-1}$$

$$=p\sum_{k=0}^{\infty} k(1-p)^{k-1}$$

$$=-p\sum_{k=0}^{\infty}\frac{d}{dk}(1-p)^{k}$$

$$=-p\frac{d}{dk}\sum_{k=0}^{\infty}(1-p)^{k} \qquad \text{(Sum of the geometric series)}$$

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

$$=\frac{1}{p}$$

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

$$=\sum_{k=0}^{\infty} k^2 p(1-p)^{k-1}$$

$$=p\sum_{k=0}^{\infty}(k(k-1)+k)(1-p)^{k-1}$$

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

$$=p(1-p)\frac{d^2}{dp^2}\sum_{k=0}^{\infty}(1-p)^{k}+p\sum_{k=0}^{\infty} k(1-p)^{k-1}$$

$$=2\frac{(1-p)}{p^2}+\frac{1}{p}$$

$$\therefore \sigma_X^2=2\frac{(1-p)}{p^2}+\frac{1}{p}-\frac{1}{p^2}$$

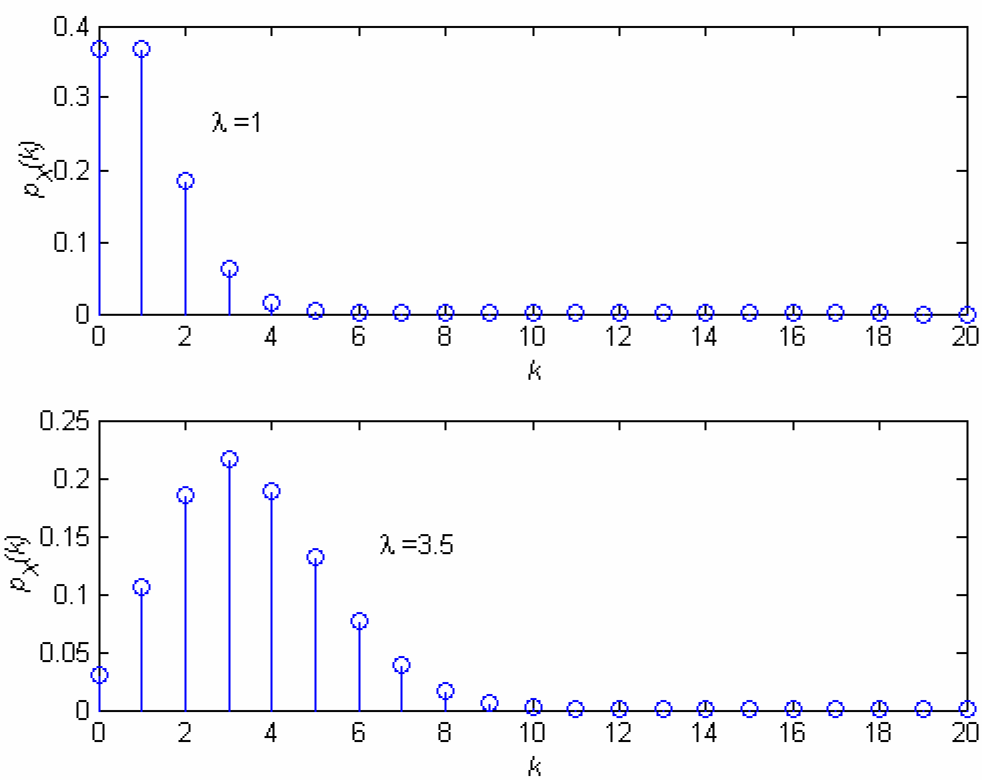
$$=\frac{(1-p)}{p^2}$$

$$\boxed{\begin{array}{l}\text{Mean} \quad \mu_X=\dfrac{1}{p}\\[2ex] \text{Variance } \sigma_X^2=\dfrac{(1-p)}{p^2}\end{array}}$$

Poisson Random Variable:

X is a Poisson distribution with the parameter λ such that $\lambda > 0$ and $p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0,1,2,....$

The plot of the pmfof the Poisson RV is shown in Fig.



Remark

- $p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$ satisfies to be a pmf, because
$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$
- Named after the French mathematician S.D. Poisson

Mean and Variance of the Poisson RV

The mean of the Poisson RV X is given by

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

$$\begin{aligned}
EX^2 &= \sum_{k=0}^{\infty} k^2 p_X(k) \\
&= 0 + \sum_{k=1}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^k}{k-1!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{(k-1+1)\lambda^k}{k-1!} \\
&= e^{-\lambda}\left(0 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k-2!}\right) + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k-1!} \\
&= e^{-\lambda}\lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{k-2!} + e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{k-1!} \\
&= e^{-\lambda}\lambda^2 e^{\lambda} + e^{-\lambda}\lambda e\lambda \\
&= \lambda^2 + \lambda
\end{aligned}$$

$$\therefore \sigma_X^2 = EX^2 - \mu_X^2 = \lambda$$

Example: The number of calls received ina 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 callsare received
(iii) more than 3 calls are received

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 $= p_X(0) = e^{-10} =$

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

(iii) probability that more the 3 calls are received $= 1 - \sum_{k=0}^{5} p_X(k) = 1 - e^{-10}(1 + \frac{10}{1} + \frac{10^2}{2!} + \frac{10^3}{3!}) =$

The Poisson distribution is used to model many practical problems. It is used in many counting applications to count events that take place independently of one another. Thus it is used to model the *count during a particular length of time* of:

- customers arriving at a service station
- telephone calls coming to a telephone exchange packets arriving at a particular server
- particles decaying from a radioactive specimen

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 a binomial RV $X \sim B(n, p)$ with $n \to \infty,\ p \to 0$ so that $EX = np = \lambda$ remains constant. Then

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

$$
\begin{aligned}
p_X(k) &= {}^nC_k\, p^k (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)(n-2).....(n-k+1)}{k!} p^k (1-p)^{n-k} \\
&= \frac{n^k (1-\frac{1}{n})(1-\frac{2}{n}).....(1-\frac{k-1}{n})}{k!} p^k (1-p)^{n-k} \\
&= \frac{(1-\frac{1}{n})(1-\frac{2}{n}).....(1-\frac{k-1}{n})}{k!} (np)^k (1-p)^{n-k} \\
&= \frac{(1-\frac{1}{n})(1-\frac{2}{n}).....(1-\frac{k-1}{n})(\lambda)^k (1-\frac{\lambda}{n})^n}{k!(1-\frac{\lambda}{n})^k}
\end{aligned}
$$

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

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

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

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

Example

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 120 words.

Suppose X is the RV representing the number of errors per page of 120 words. $X \sim B(120, p)$ where $p = 0.01$. Therefore,

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

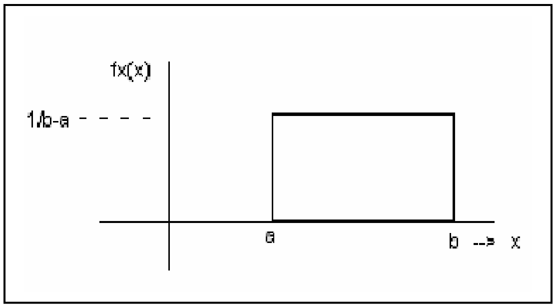
P(more than one errors)

$$
\begin{aligned}
&= 1 - p_X(0) - p_X(1) \\
&\simeq 1 - e^{-\lambda} - \lambda e^{-\lambda}
\end{aligned}
$$

Uniform Random Variable

A continuous random variable X is called uniformly distributed over the interval [a, b], if its probability density function is given by

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



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 $a \le x \le b$

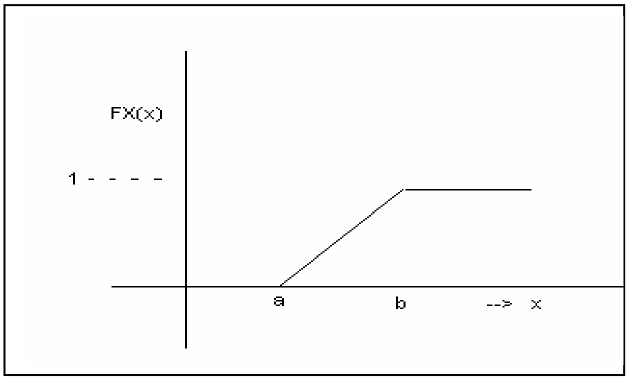
$\int_{-\infty}^{x} f_X(u)dx$

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

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

For $x > b$,

$F_X(x) = 1$



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
$$\phi_X(w) = Ee^{jwx} = \int_a^b \frac{e^{jwx}}{b-a}dx$$
$$= \frac{e^{jwb} - e^{jwa}}{b-a}$$
Example:
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 4 V? What is the variance of the voltage?
$$P(2 < X \le 3) = \int_2^3 \frac{dx}{5-(-4)} = \frac{1}{9}.$$
$$\sigma_X^2 = \frac{(5+4)^2}{12} = \frac{27}{4}.$$
Remark

- The uniform distribution is the simplest continuous distribution
- Used, for example, to to model *quantization errors*. If a signal is discritized into steps of Δ, then the quantization error is uniformly distributed between $\frac{-\Delta}{2}$ and $\frac{\Delta}{2}$.
- The unknown phase of a sinusoid is assumed to be uniformly distributed over $[0, 2\pi]$ in many applications. For example, in studying the noise performance of a communication receiver, the carrier signal is modeled as
  $X(t) = A\cos(w_c t + \Phi)$
  where $\Phi \sim U(0, 2\pi)$.
- A random variable of arbitrary distribution can be generated with the help of a routine to generate uniformly distributed random numbers. This follows from the fact that the distribution function of a random variable is uniformly distributed over [0,1]. (See Example)
  Thus if X is a contnuous raqndom variable, then $F_X(X) \sim U(0,1)$.

Normal or Gaussian Random Variable

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

A continuous random variable X is called a normal or a Gaussian random variable with parameters μ_X and $\sigma_X{}^2$ if its probability density function is given by,

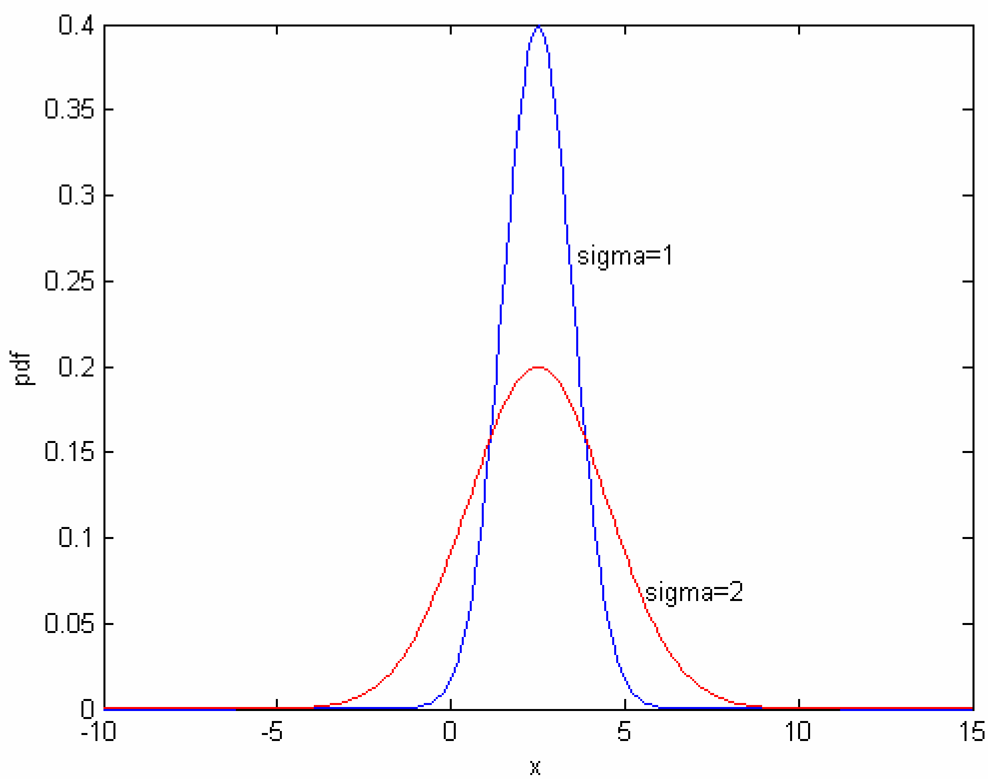
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}, \; -\infty < x < \infty$$

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 natural variable.



- $f_X(x)$, is a bell-shaped function, symmetrical about $x = \mu_X$.

- σ_X , determines the spread of the random variable X. If $\sigma_X{}^2$ is small X is more concentrated around the mean μ_X .

-

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
&EX = \mu_X \\
&\operatorname{var}(X) = \sigma_X{}^2
\end{aligned}
$$

Proof:

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

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

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

$$= 0 + \mu_X \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du$$

$$= \frac{\mu_{XX}}{\sqrt{2\pi}} 2\int_{0}^{\infty} e^{-\frac{u^2}{2}} du = \mu_X$$

Substituting,

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

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

Evaluation of $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$

Suppose $I = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$

Then

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx\right)^2$$

$$= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dydx$$

Substituting $x = r\cos\theta$ and $y = r\sin\theta$ we get

$$I^2 = \int_{-\pi}^{\pi}\int_{0}^{\infty} e^{-\frac{r^2}{2}} r\, d\theta\, dr$$

$$= 2\pi\int_{0}^{\infty} e^{-r^2} r\, dr$$

$$= 2\pi\int_{0}^{\infty} e^{-s} ds \qquad (r^2 = s)$$

$$= 2\pi \times 1 = 2\pi$$

$$\therefore I = \sqrt{2\pi}$$

$$Var(X) = E\left(X - \mu_X\right)^2$$

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

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

Put $\dfrac{x - \mu_X}{\sigma_X} = u$ So that $dx = \sigma_X du$

$$= 2 \times \frac{\sigma_X{}^2}{\sqrt{2\pi}} \int_0^{\infty} u^2 e^{-\frac{1}{2}u^2} du$$

Put

$t = \dfrac{1}{2}u^2$ so that $dt = udu$

$$= 2 \times \frac{\sigma_X{}^2}{\sqrt{2\pi}} \sqrt{2} \int_0^{\infty} t^{\frac{1}{2}} e^{-t} dt$$

$$= 2 \times \frac{\sigma_X{}^2}{\sqrt{\pi}} \sqrt{\frac{3}{2}}$$

$$= 2 \times \frac{\sigma_X{}^2}{\sqrt{\pi}} \frac{1}{2} \sqrt{\frac{1}{2}}$$

$$= \frac{\sigma_X{}^2}{\sqrt{\pi}} \times \sqrt{\pi}$$

$$= \sigma_X{}^2$$

Note the definition and properties of the gamma function

$$\Gamma n = \int_0^{\infty} t^{n-1} e^{-t} dt$$

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

$$\Gamma \frac{1}{2} = \sqrt{\pi}$$

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 probabilty distribution function is

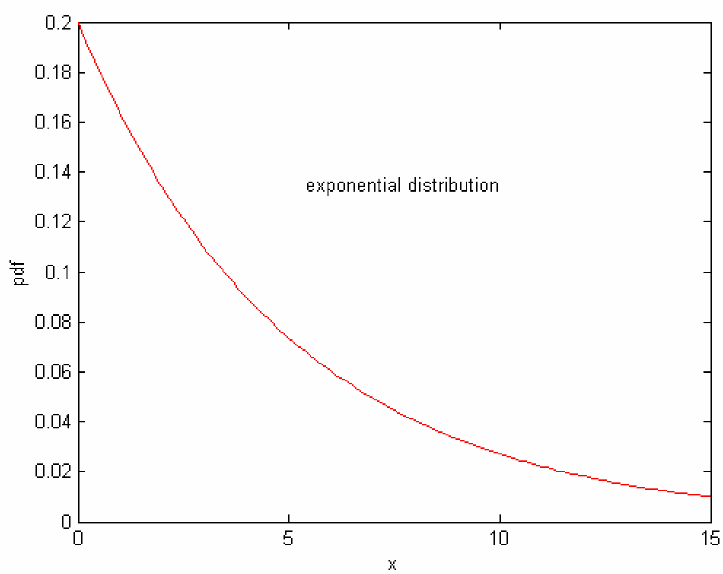
$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

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

We have $\mu_X = EX = \int_0^{\infty} x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$

$$\sigma_X^2 = E(X - \mu_X)^2 = \int_0^{\infty} (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

The following figure shows the pdf of an exponential RV.



- The time between two consecutive occurrences of independent events can be modeled by the exponential RV. For example, the exponential distribution gives probability density function of the time between two successive counts in Poisson distribution
- Used to model the service time in a queing system.
- In *reliability studies*, the expected life-time of a part, the average time between successive failures of a system etc.,are determined using the exponential distribution.

Memoryless Property of the Exponential Distribution

For an exponential RV X with parameter λ,

$$P(X > t + t_0 / X > t_0) = P(X > t) \text{ for } t > 0,\ t_0 > 0$$

Proof:

$$\begin{aligned}
P(X > t + t_0 / X > t_0) &= \frac{P[(X > t + t_0) \cap PX > t_0)]}{P(X > t_0)} \\
&= \frac{P(X > t + t_0)}{P(X > t_0)} \\
&= \frac{1 - F_X(t + t_0)}{1 - F_X(t_0)} \\
&= \frac{e^{-\lambda(t+t_0)}}{e^{-\lambda t_0}} \\
&= e^{-\lambda t} = P(X > t)
\end{aligned}$$

Hence if X represents the life of a component in hours, the probability that the component will last more than $t + t_0$ hours given that it has lasted t_0 hours, is same as the probability that the component will last t hours. The information that the component has already lasted for t_0 hours is not used. Thus the life expectancy of an used component is same as that for a new component. Such a model cannot represent a real-world situation, but used for its simplicity.

Laplace Distribution

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

$$f_X(x) = \frac{\lambda}{2} e^{-\lambda|x|} \qquad \lambda > 0, \text{-}\infty < x < \infty$$

We have $\mu_X = EX = \int_{-\infty}^{\infty} x \frac{\lambda}{2} e^{-\lambda|x|} \, dx = 0$

$$\sigma_X^2 = E(X - \mu_X)^2 = \int_{\infty}^{\infty} x^2 \frac{\lambda}{2} e^{-\lambda|x|} \, dx = \lambda$$

Chi-square random variable

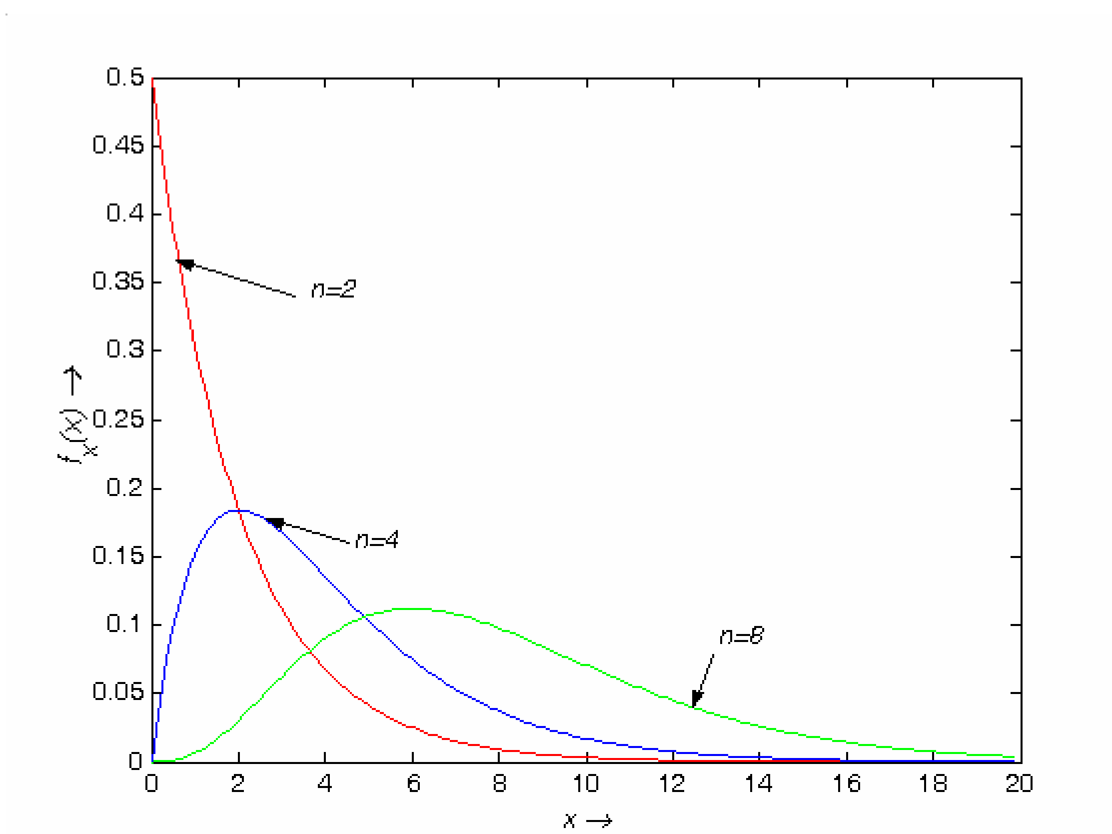
A random variable is called a Chi-square random variable with n degrees of freedom if its PDF is given by

$$f_X(x) = \begin{cases} \dfrac{x^{n/2-1}}{2^{n/2}\sigma^n\Gamma(n/2)} e^{-x/2\sigma^2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

with the parameter $\sigma > 0$ and $\Gamma(.)$ denoting the gamma function. A Chi-square random variable with n degrees of freedom is denoted by χ_n^2.

Note that a χ_2^2 RV is the exponential RV.

The pdf of χ_n^2 RVs with different degrees of freedom is shown in Fig. below:



Mean and variance of the chi-square random variable

$$\begin{aligned}
\mu_X &= \int_{-\infty}^{\infty} x f_X(x)dx \\
&= \int_0^{\infty} x \frac{x^{n/2-1}}{2^{n/2}\sigma^n\Gamma(n/2)} e^{-x/2\sigma^2} dx \\
&= \int_0^{\infty} \frac{x^{n/2}}{2^{n/2}\sigma^n\Gamma(n/2)} e^{-x/2\sigma^2} dx \\
&= \int_0^{\infty} \frac{(2\sigma^2)^{n/2}u^{n/2}}{2^{n/2}\sigma^n\Gamma(n/2)} e^{-u}(2\sigma^2)du \quad (\text{ Substituting } u = x/2\sigma^2) \\
&= \frac{2\sigma^2\Gamma[(n+2)/2]}{\Gamma(n/2)} \\
&= \frac{2\sigma^2 n/2\Gamma(n/2)}{\Gamma(n/2)} \\
&= n\sigma^2
\end{aligned}$$

Similarly,

$$\begin{aligned}
EX^2 &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= \int_0^{\infty} x^2 \frac{x^{n/2-1}}{2^{n/2}\sigma^n\Gamma(n/2)} e^{-x/2\sigma^2} dx \\
&= \int_0^{\infty} \frac{x^{(n+2)/2}}{2^{n/2}\sigma^n\Gamma(n/2)} e^{-x/2\sigma^2} dx \\
&= \int_0^{\infty} \frac{(2\sigma^2)^{(n+2)/2}u^{n/2}}{2^{n/2}\sigma^n\Gamma(n/2)} e^{-u}(2\sigma^2)du \quad (\text{ Substituting } u = x/2\sigma^2) \\
&= \frac{4\sigma^4\Gamma[(n+4)/2]}{\Gamma(n/2)} \\
&= \frac{4\sigma^4[(n+2)/2]n/2\Gamma(n/2)}{\Gamma(n/2)} \\
&= n(n+2)\sigma^4
\end{aligned}$$

$$\sigma_X^2 = EX^2 - \mu_X^2 = n(n+2)\sigma^4 - n\sigma^4 = 2n\sigma^4$$

A random variable Let $X_1, X_2 \; . \; . \; . \; X_n$ be independent zero-mean Gaussian variables each with variance σ_X^2. Then $Y = X_1^2 + X_2^2 + . \; . \; . + X_n^2$ has χ_n^2 distribution with mean $n\sigma^2$ and varaiance $2n\sigma^4$. This result gives an application of the chi-square random variable.

Relation between the chi-square distribution and the Gaussian distribution

A random variable Let $X_1, X_2 \ . \ . \ . X_n$ be independent zero-mean Gaussian variables each with variance σ_X^2. Then $Y = X_1^2 + X_2^2 + . \ . \ . + X_n^2$ has χ_n^2 distribution with mean

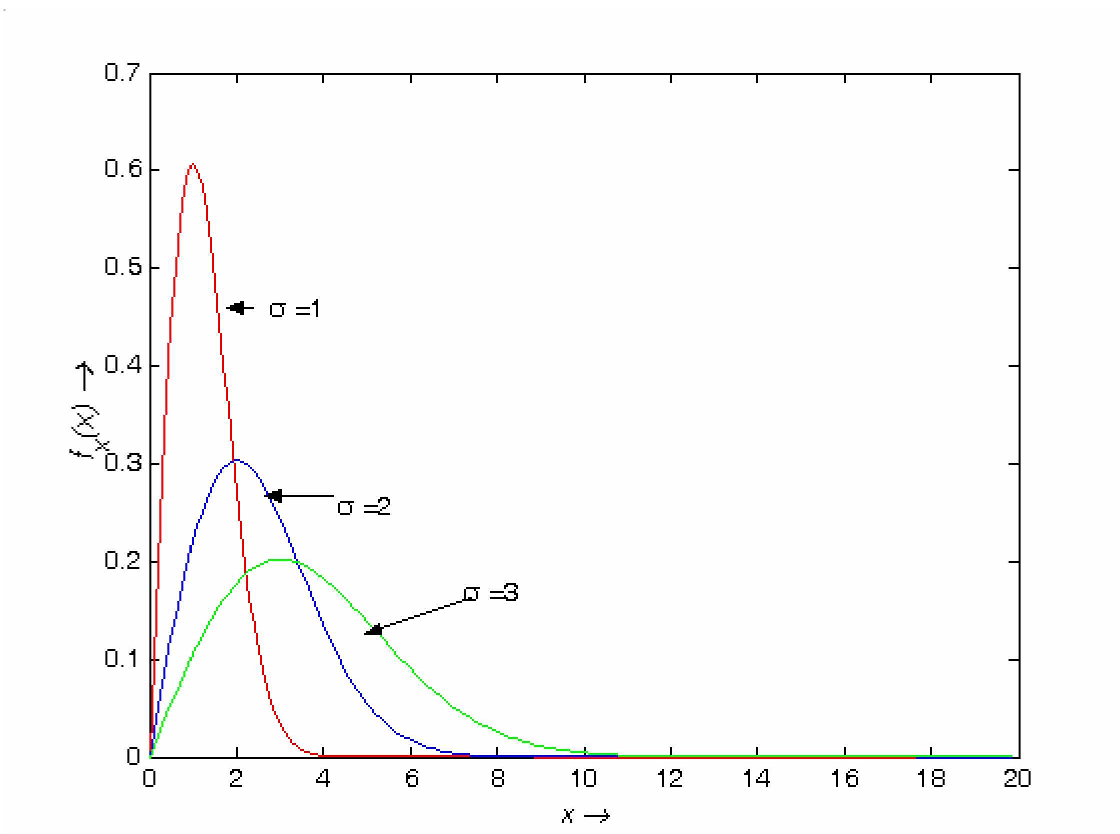
Rayleigh Random Variable

A Rayleigh random variable 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 Fig.



Mean and variance of the Rayleigh distribution

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

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

$$= \frac{\sqrt{2\pi}}{\sigma} \int_0^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx$$

$$= \frac{\sqrt{2\pi}}{\sigma} \frac{\sigma^2}{2}$$

$$= \sqrt{\frac{\pi}{2}}\sigma$$

Similarly

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

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

$$= 2\sigma^2 \int_0^{\infty} u e^{-u} du \quad \textbf{(Substituting } u = \frac{x^2}{2\sigma^2})$$

$$= 2\sigma^2 \qquad \textbf{(Noting that } \int_0^{\infty} u e^{-u} du \textbf{ is the mean of the exponential RV with } \lambda\textbf{=1)}$$

$$\therefore \sigma_X^2 = 2\sigma^2 - (\sqrt{\frac{\pi}{2}}\sigma)^2$$

$$= (2 - \frac{\pi}{2})\sigma^2$$

Relation between the Rayleigh distribution and the Gaussian distribution

A Rayleigh RV is related to Gaussian RVs as follow:

Let X_1 and X_2 be two independent zero-mean Gaussian RVs with a common variance σ^2. Then $X = \sqrt{X_1^2 + X_2^2}$ has the Rayleigh pdf with the parameter σ.

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

Application of the Rayleigh RV

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

$$s(t) = X_1 \cos wt + Y_1 \sin wt$$

If $X_1 \sim N(0, \sigma^2)$ and $X2 \sim N(0, \sigma^2)$ are independent, then the envelope $X = \sqrt{X_1^2 + X_2^2}$ has the Rayleigh distribution.

Simulation of Random Variables

- In many fields of science and engineering, computer simulation is used to study random phenomena in nature and the performance of an engineering system in a noisy environment. For example, we may study through computer simulation the performance of a communication receiver. Sometimes a probability model may not be analytically tractable and computer simulation is used to calculate probabilities.
- The heart of all these application is that it is possible to simulate a random variable with an *empirical* CDF or pdf that fits well with the theoretical CDF or pdf.

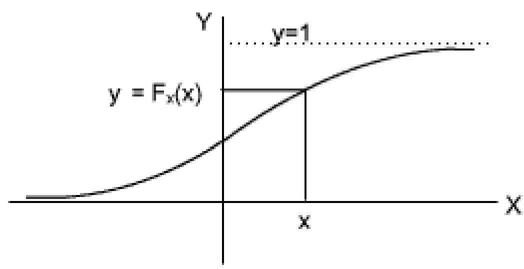
Generation of Random Numbers

Generation of random numbers means producing a sequence of independent random numbers with a specified CDF or pdf. All the random number generators rely on a routine to generate random numbers with the uniform pdf. Such routine is of vital importance because the quality of the generated random numbers with any other distribution depends on it. By the *quality of the generated random numbers* we mean how closely the empirical CDF or PDF fits the true one.

There are several algorithms to generate $U[0\ 1]$ random numbers. Note that these algorithms generate random number by a reproducible deterministic method. These numbers are *pseudo random numbers* because they are reproducible and the same sequence of numbers repeats after some period of count specific to the generating algorithm. This period is very high and a finite sample of data within the period appears to be uniformly distributed. We will not discuss about these algorithms. Software packages provide routines to generate such numbers.

Method of Inverse transform

Suppose we want to generate a random variable X with a prescribed distribution function $F_X(x)$. We have observed that the random variable Y, defined by $Y = F_X(X) \sim U[0, 1]$.. Thus given $U[0, 1]$. random number Y, the inverse transform $X = F_X^{-1}(Y)$ will have the CDF $F_X(x)$.



The algorithmic steps for the inverse transform method are as follows:

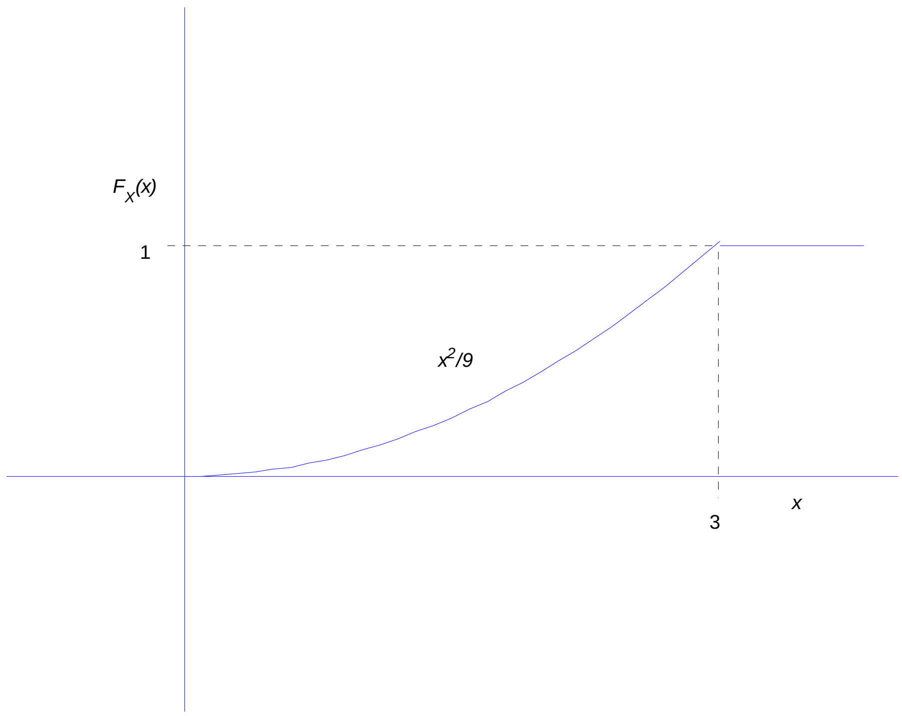
1. Generate a random number from $Y \sim U[0,\ 1]$. Call it y. .
2. Compute the value x such that F_X (x) =y.
3. Take x to be the random number generated.

Example Suppose, we want to generate a random variable with the pdf $f_X(x)$ ribution given by

$$f_X(x) = \begin{cases} \frac{2}{9}x & 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

The CDF of X is given by

$$F_X(x) = \begin{cases} 0 & \\ \frac{1}{9}x^2 & 0 \le x \le 3 \\ 1 & \text{x>0} \end{cases}$$



Therefore, we generate a random number y from the $U[0,\ 1]$ distribution and set $F_X(x) = y.$

We have

$$\frac{1}{9}x^2 = y$$

$$\Rightarrow \mathrm{x}=\sqrt{9\mathrm{y}}$$

Example Suppose, we want to generate a random variable with the exponential distribution given by $f_X(x) = \lambda e^{-\lambda x} \quad \lambda > 0, \ x \geq 0.$ Then

$F_X(x) = 1 - e^{-\lambda x}$

Therefore given y, we can get x by the mapping

$$1 - e^{-\lambda x} = y$$

$$x = -\frac{\log_e(1-y)}{\lambda}$$

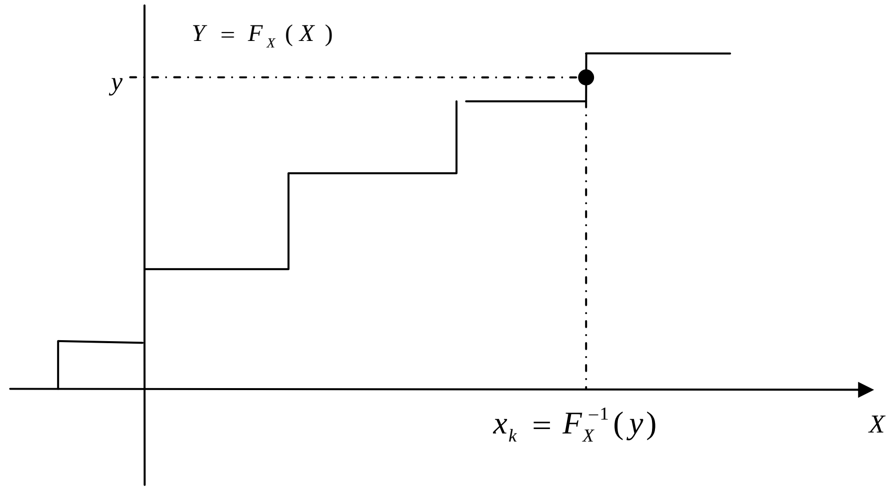
Since $1-y$ is also uniformly distributed over [0, 1], the above expression can be simplified as,

$$x = -\frac{\log y}{\lambda}$$

Generation of Gaussian random numbers

Generation of discrete random variables

We observed that the CDF of a discrete random variable is also $U[0,\ 1]$ distributed. Suppose X is a discrete random variable with the probability mass function $p_X(x_i),\ i=1,2,..n.$ Given $y = F_X(x),$ the inverse mapping is defined as shown in the Fig. below.



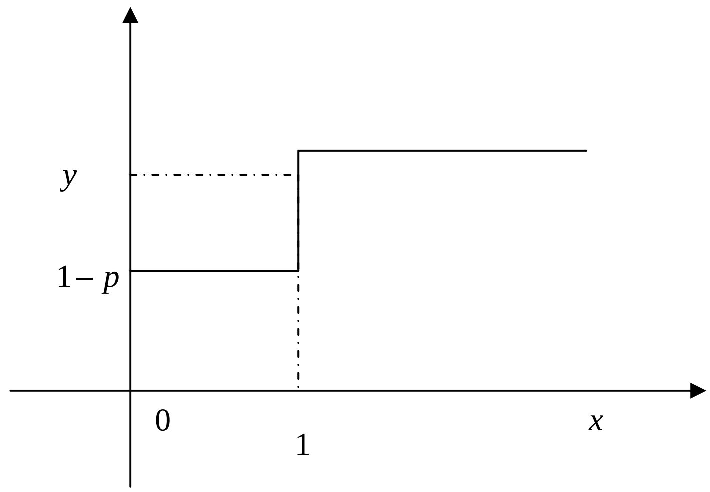
Thus if $F_X(x_{k-1}) \leq y < F_X(x_k),$ then
$F_X^{-1}(y) = x_k$

The algorithmic steps for the inverse transform method for simulating discrete random variables are as follows:

1. Generate a random number from $Y \sim U[0,\ 1].$ Call it $y.$
2. Compute the value x_k such that $F_X(x_{k-1}) \leq y < F_X(x_k),$
3. Take x_k to be the random number generated.

Example Generation of Bernoulli random numbers
Suppose we want to generate a random number from $X \sim Br(p).$ Generate y from the $U[0,\ 1]$ distribution. Set

$$x = \begin{cases} 0 & \text{for } y \leq 1-p \\ 1 & \text{otherwise} \end{cases}$$



y
1− p
0
1
x

Jointly distributed random variables

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

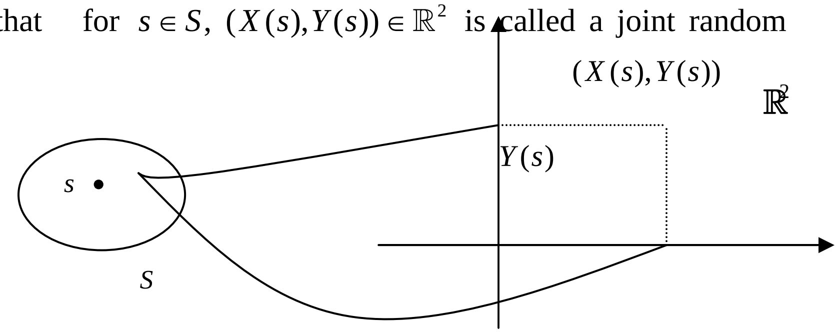


Figure Joint Random Variable

Remark

- The above figure illustrates the mapping corresponding to a joint random variable. The joint random variable in the above case id denoted by (X, Y).
- We may represent a joint random variable as a two-dimensional vector $\mathbf{X} = [X \;\; \mathrm{Y}]'$.
- We can extend the above definition to define joint random variables of any dimension.The mapping $S \to \mathbb{R}^n$ such that for $s \in S$, $(X_1(s), X_2(s), ..., X_n(s)) \in \mathbb{R}^n$ is called a n-dimesional random variable and denoted by the vector $\mathbf{X} = [X_1 \;\; X_2 \;\; ....X_n]'$.

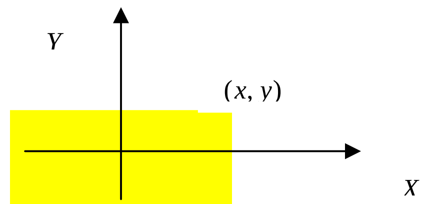
Example1: Suppose we are interested in studying the height and weight of the students in a class. We can define the joint RV (X, Y) where X represents height and Y represents the weight.

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

Joint Probability Distribution Function

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



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

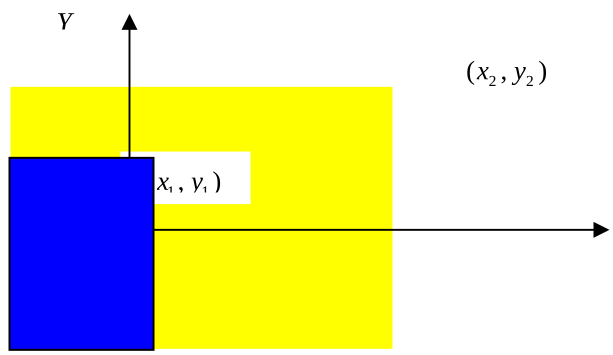
- $F_{X,Y}(x_1,y_1) \le F_{X,Y}(x_2,y_2)$ if $x_1 \le x_2$ and $y_1 \le y_2$

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

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

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

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



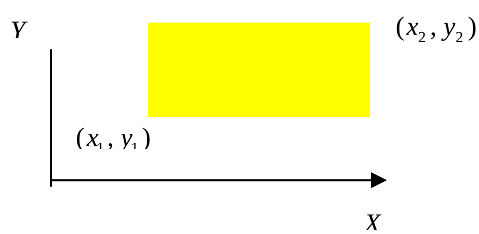
- $F_{X,Y}(-\infty,y) = F_{X,Y}(x,-\infty) = 0$

Note that

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

- $F_{X,Y}(\infty,\infty) = 1.$
- $F_{X,Y}(x,y)$ is right continuous in both the variables.
- I If $x_1 < x_2$ and $y_1 < y_2$,

$P\{x_1 < X \le x_2,\ y_1 < Y \le y_2\} = F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1) \ge 0.$



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

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

  To prove this

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

  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.

Example

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

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

(a) Find the marginal CDFs
(b) Find the probability $P\{1_1 < X \le 2,\ 1 < Y \le 2\}$

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

(b)
$$P\{1_1 < X \le 2,\ 1 < Y \le 2\} = F_{X,Y}(2,2) + F_{X,Y}(1,1) - F_{X,Y}(1,2) - F_{X,Y}(2,1)$$
$$= (1-e^{-4})(1-e^{-2}) + (1-e^{-2})(1-e^{-1}) - (1-e^{-2})(1-e^{-2}) - (1-e^{-4})(1-e^{-1})$$
$$= 0.0272$$

Jointly distributed discrete random variables

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

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

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

Remark

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

  This is because

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

- *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

  $$\begin{aligned} p_X(x) &= P\{X = x\} \cup R_Y \\ &= \sum_{y \in R_Y} p_{X,Y}(x, y) \end{aligned}$$

  and similarly

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

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

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

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

Joint Probability Density Function

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

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

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

Properties of Joint Probability Density Function

- $f_{X,Y}(x,y)$ is always a non-negaitive 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 B 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

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

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

Remark

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

- With the help of the two-dimensional Dirac Delta function, we can define the joint pdf of two discrete jointly random variables. Thus for discrete jointly random variables X and Y.
$$f_{X,Y}(x,y) = \sum_{(x_i,y_j)\in R_X \times R_Y} \sum p_{X,Y}(x,y)\delta(x-x_j, y-y_j)$$

Example The joint density function $f_{X,Y}(x,y)$ in the previous example are

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

Example: The joint pdf of two random variables X and Y are given by
$$\begin{aligned} f_{X,Y}(x,y) &= cxy \quad 0 \le x \le 2,\ 0 \le y \le 2 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

(i) Find c.

(ii) Find $F_{X,y}(x,y)$

(iii) Find $f_X(x)$ and $f_Y(y)$.

(iv) What is the probability $P(0 < X \le 1, 0 < Y \le 1)$?

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

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

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

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

Similarly

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

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

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

$$= \frac{1}{16} + 0 - 0 - 0$$

$$= \frac{1}{16}$$

Jointly distributed random variables

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

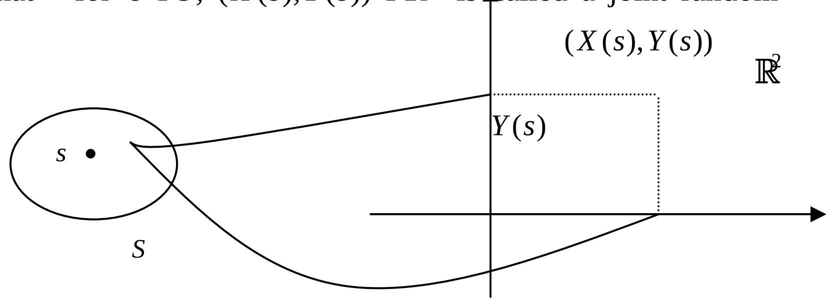


Figure Joint Random Variable

Remark

- The above figure illustrates the mapping corresponding to a joint random variable. The joint random variable in the above case id denoted by (X, Y).
- We may represent a joint random variable as a two-dimensional vector $\mathbf{X} = [X \ \ Y]'$.
- We can extend the above definition to define joint random variables of any dimension.The mapping $S \to \mathbb{R}^n$ such that for $s \in S$, $(X_1(s), X_2(s), ..., X_n(s)) \in \mathbb{R}^n$ is called a n-dimesional random variable and denoted by the vector $\mathbf{X} = [X_1 \ X_2 \ ....X_n]'$.

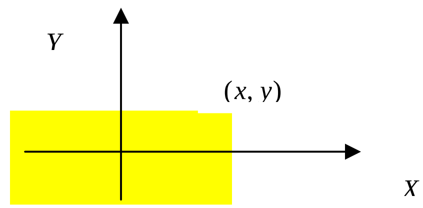
Example1: Suppose we are interested in studying the height and weight of the students in a class. We can define the joint RV (X, Y) where X represents height and Y represents the weight.

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

Joint Probability Distribution Function

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

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



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

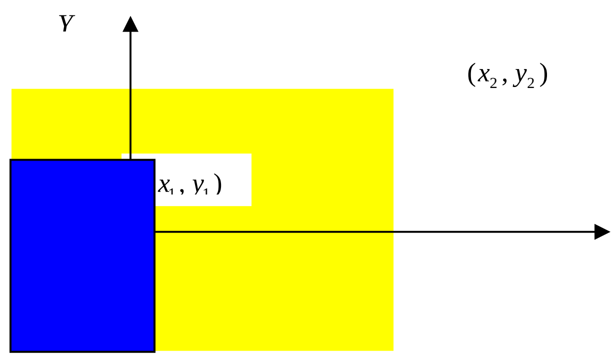
- $F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$ if $x_1 \le x_2$ and $y_1 \le y_2$

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

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

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

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



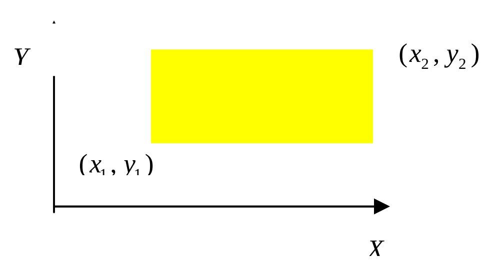
- $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$

Note that

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

- $F_{X,Y}(\infty, \infty) = 1.$
- $F_{X,Y}(x,y)$ is right continuous in both the variables.
- If $x_1 < x_2$ and $y_1 < y_2$,

$P\{x_1 < X \le x_2,\ y_1 < Y \le y_2\} = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1) \ge 0.$



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

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

  To prove this

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

  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.

Example

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

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

(a) Find the marginal CDFs
(b) Find the probability $P\{1_1 < X \le 2,\ \ 1 < Y \le 2\}$

(a)

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

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

(b)

$$P\{1_1 < X \le 2,\ \ 1 < Y \le 2\} = F_{X,Y}(2,2) + F_{X,Y}(1,1) - F_{X,Y}(1,2) - F_{X,Y}(2,1)$$
$$= (1-e^{-4})(1-e^{-2}) + (1-e^{-2})(1-e^{-1}) - (1-e^{-2})(1-e^{-2}) - (1-e^{-4})(1-e^{-1})$$
$$=0.0272$$

Jointly distributed discrete random variables

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

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

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

Remark

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

  This is because

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

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

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

  $$=P(S) = 1$$

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

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

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

  and similarly

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

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

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

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

Joint Probability Density Function

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

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

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

Properties of Joint Probability Density Function

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

  $f_{X,Y}(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$
- The probability of any Borel set B can be obtained by
  $P(B) = \iint_{(x,y)\in B} f_{X,Y}(x,y)dxdy$

Marginal density functions

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

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

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

Remark

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

- With the help of the two-dimensional Dirac Delta function, we can define the joint pdf of two discrete jointly random variables. Thus for discrete jointly random variables X and Y.

  $$f_{X,Y}(x,y) = \sum_{(x_i,y_j)\in R_X\times R_Y} \sum p_{X,Y}(x,y)\delta(x-x_j, y-y_j)$$

Example The joint density function $f_{X,Y}(x,y)$ in the previous example are

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

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

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

(i) Find $c.$

(ii) Find $F_{X,y}(x,y)$

(iii) Find $f_X(x)$ and $f_Y(y).$

(iv) What is the probability $P(0 < X \leq 1, 0 < Y \leq 1)$?

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

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

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

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

Similarly

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

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

Conditional probability mass functions

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

Similarly we can define the conditional probability mass function $p_{X/Y}(x/y)$

- From the definition of conditional probability mass functions, we can define two *independent random variables.* Two discrete random variables X and Y are said to be *independent* if and only if

  $p_{Y/X}(y/x) = p_Y(y)$

  so that

  $p_{X,y}(x, y) = p_X(x)p_Y(y)$

- Bayes rule:

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

Example Consider the random variables X and Y with the joint probability mass function as tabulated in Table .

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		

The marginal probabilities are as shown in the last column and the last row

$$p_{Y/X}(0/1) = \frac{p_{X,Y}(0,1)}{p_X(1)}$$
$$= \frac{0.14}{0.39}$$

Conditional Probability Density Function

$f_{Y/X}(y/X = x) = f_{Y/X}(y/x)$ is called conditional density of Y given X.
Let us define the conditional distribution function .

We cannot define the conditional distribution function for the continuous random variables X and Y by the relation

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

as both the numerator and the denominator are zero for the above expression.

The conditional distribution function is defined in the *limiting sense* as follows:

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

The conditional density is defined in the limiting sense as follows

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

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

The right hand side in equation (1) is

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

$$\boxed{\therefore f_{Y/X}(x/y)=f_{X,Y}(x,y)/f_X(x)} \qquad (2)$$

Similarly we have

$$\boxed{\therefore f_{X/Y}(x/y)=f_{X,Y}(x,y)/f_Y(y)} \qquad (3)$$

- 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

- $f_{X,Y}(x,y)=f_X(x)f_Y(y)$ (4)

Bayes rule for continuous random variables:

From (2) and (3) we get Baye's rule

$$\therefore f_{X/Y}(x/y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$
$$= \frac{f_X(x)f_{Y/X}(y/x)}{f_Y(y)}$$
$$= \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx}$$
$$= \frac{f_{Y/X}(y/x)f_X(x)}{\int_{-\infty}^{\infty} f_X(u)f_{Y/X}(y/x)du} \qquad (4)$$

Given the joint density function we can find out the conditional density function.

Example: For random variables X and Y, the joint probability density function is given by

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

Find the marginal density $f_X(x), f_Y(y)$ and $f_{Y/X}(y/x)$. Are X and Y independent?

$$f_X(x) = \int_{-1}^{1} \frac{1+xy}{4} dy$$
$$= \frac{1}{2}$$

Similarly

$$f_Y(y) = \frac{1}{2} \qquad -1 \le y \le 1$$

and

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

Baye's Rule for mixed random variables

Let X be a discrete random variable with probability mass function $p_X(x)$ and Y be a continuous random variable with the conditional probability density function $f_{Y/X}(y/x)$. In practical problem we may have to estimate X from observed Y. Then

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

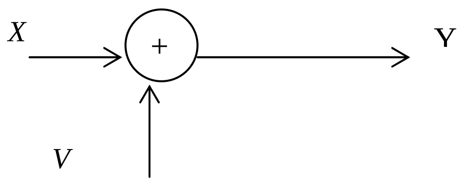
$$= \lim_{\Delta y \to 0} \frac{P(X = x, y < Y \leq y + \Delta y)}{P(y < Y \leq y + \Delta y)}$$

$$= \lim_{\Delta y \to 0} \frac{p_X(x) f_{Y/X}(y/x)\Delta y}{f_Y(y)\Delta y}$$

$$= \frac{p_X(x) f_{Y/X}(y/x)}{f_Y(y)}$$

$$= \frac{p_X(x) f_{Y/X}(y/x)}{\sum_x p_X(x) f_{Y/X}(y/x)}$$

Example



X is a binary random variable with

$$X = \begin{cases} 1 \text{ with probability } p \\ -1 \text{ with probability } 1\text{-}p \end{cases}$$

V is the Gaussian noise with mean 0 and variance $\sigma.^2$

Then

$$p_{X/Y}(x=1/y) = \frac{p_X(x) f_{Y/X}(y/x)}{\sum_x p_X(x) f_{Y/X}(y/x)}$$

$$= \frac{pe^{-(y-1)^2/2\sigma^2}}{pe^{-(y-1)^2/2\sigma^2} + (1-p)e^{-(y+1)^2/2\sigma^2}}$$

<u>Independent Random Variables</u>

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

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

and the corresponding joint density function $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$

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

$$
\begin{aligned}
F_{X,Y}(x,y) &= P\{X \le x, Y \le y\} \\
&= P\{X \le x\}P\{Y \le y\} \\
&= F_X(x)F_Y(y)
\end{aligned}
$$

$$
\begin{aligned}
\therefore\ f_{X,Y}(x,y) &= \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \\
&= f_X(x)\, f_Y(y)
\end{aligned}
$$

and equivalently

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

Remark:

Suppose X and Y are two discrete random variables with joint probability mass function $p_{X,Y}(x,y)$. Then X and Y are independent if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad \forall (x,y) \in R_X \times R_Y$$

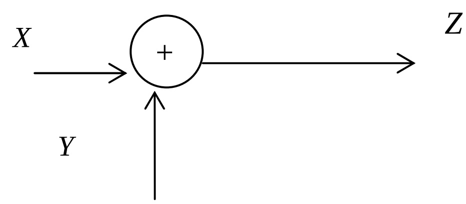
Transformation of two random variables:

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 $Z = XY$.

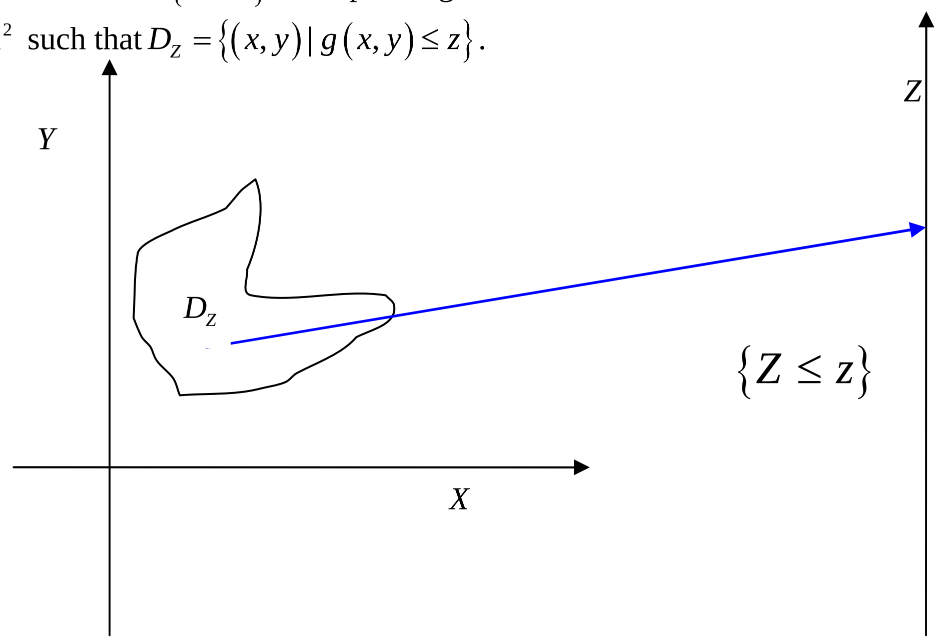
We have to know about the probability distribution of Z in any analysis of Z. More formally, given two random variables X and Y with joint probability density function $f_{X,Y}(x, y)$ and a function $Z = g(X,Y)$, we have to find $f_Z(z)$.

In this lecture, we will try to 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 $D_Z = \{(x, y) \mid g(x, y) \le z\}$.



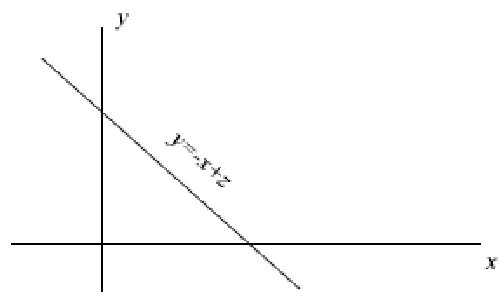
$$\therefore F_Z(z) = P(\{Z \le z\})$$
$$= P\{(x,y) \mid (x,y) \in D_z\}$$
$$= \iint\limits_{(x,y)\in D_z} f_{X,Y}(x,y)\,dydx$$

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

Example: Suppose $Z = X+Y$. Find the PDF $f_Z(z)$.

$Z \le z$
$\Rightarrow X + Y \le Z$

Therefore D_Z is the shaded region in the fig below.



$$\therefore F_Z(z) = \iint\limits_{(x,y)\in D_z} f_{X,Y}(x,y)\,dxdy$$
$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-x} f_{X,Y}(x,y)dy\right]dx$$
$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z} f_{X,Y}(x,u-x)\,du\right]dx \quad \text{substituting}$$
$$= \int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f_{X,Y}(x,u-x)\,dx\right]du \quad \text{interchanging the order of integration}$$

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

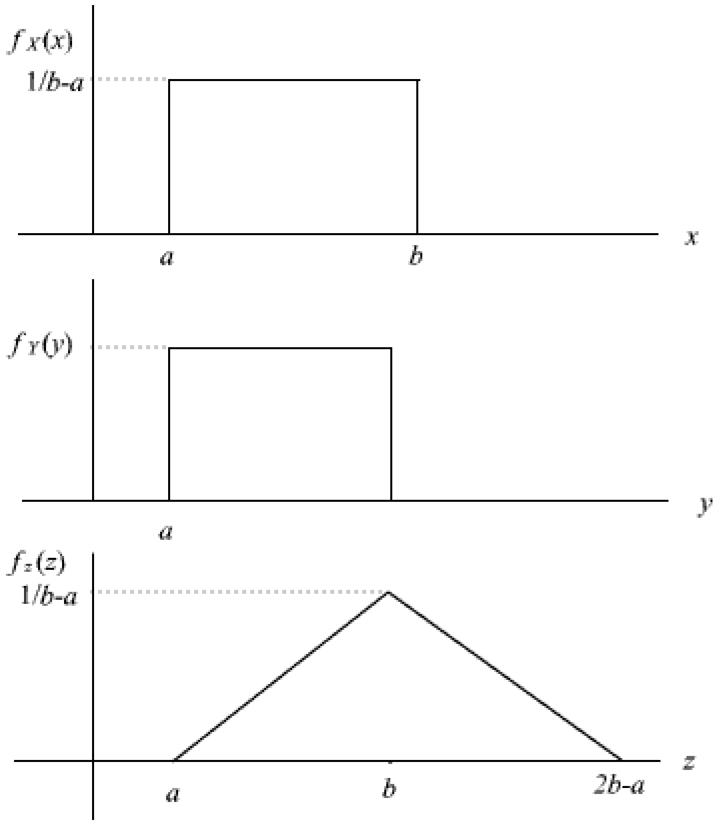
If X and Y are independent

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

Where * is the convolution operation.

Example:

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 is a triangular probability density function as shown in the figure.

Probability density function of $Z = XY$

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

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

$$= \int_{-\infty}^{\infty} \frac{1}{|x|} \int_{-\infty}^{\infty} f_{X,Y}\left(x, \frac{u}{x}\right) dudx$$

Substituting $u = xy$
$du = xdy$

$$\therefore f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{z}{x}\right) dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X,Y}\left(\frac{z}{y}, y\right) dy$$

Probability density function of $Z = \dfrac{Y}{X}$

$$Z \le z$$
$$\Rightarrow \frac{Y}{X} \le z$$
$$\Rightarrow Y \le xz$$

$$\therefore D_z = \{(x, y) \mid Z \le z\}$$
$$= \{-\infty < x \le \infty, y \le xz\}$$

$$\therefore F_Z(z) = \iint_{(x,y)\in D_z} f_{X,Y}(x, y)\, dydx$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{zx} f_{XY}(x, y)\, dydx$$
$$= \int_{-\infty}^{\infty} |x| \int_{-\infty}^{z} f_{X,Y}(x, ux)\, dudx$$

Suppose, X and Y are independent random variables. Then, $f_{X,Y}(x, zx) = f_X(x)\, f_Y(zx)$

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

$$\therefore f_Z(z) = \frac{d}{dz} F_Z(z)$$
$$= \int_{-\infty}^{\infty} |x|\, f_{X,Y}(x, z, k)\, dx$$
$$= \int_{-\infty}^{\infty} |y|\, f_{X,Y}(z, y, u)\, dy$$

Example:

Suppose X and Y are independent zero mean Gaussian random variable with unity standard derivationand $Z = \frac{Y}{X}$. Then

$$f_Z(z) = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{3^2 x^2}{2}} dx$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} |x| e^{-\frac{1}{2}x^2(1+z^2)} dx$$

$$= \frac{1}{\pi} \int_{0}^{\infty} x e^{-\frac{1}{2}x^2(1+z^2)} dx$$

$$= \frac{1}{\pi(1+z^2)}$$

which is the Cauchy probability density function.

Probability density function of $Z = \sqrt{X^2 + Y^2}$

$$\therefore D_z = \{(x, y) \mid Z \le z\}$$

$$= \left\{(x, y) \mid \sqrt{x^2 + y^2} \le z\right\}$$

$$= \{(r, \theta) \mid 0 \le r \le z, 0 \le \theta \le 2\pi\}$$

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

$$= \int_0^{2\pi} \int_0^{z} f_{XY}(r\cos\theta, r\sin\theta)\, rdrd\theta$$

$$\therefore f_Z(z) = \frac{d}{dz} F_Z(z) = \int_0^{2\pi} f_{XY}(z\cos\theta, z\sin\theta)\, zd\theta$$

Example Suppose X and Y are two independent Gaussian random variables each with mean 0 and variance σ^2 and $Z = \sqrt{X^2 + Y^2}$. Then

$$f_Z(z) = \int_0^{2\pi} f_{XY}(z\cos\theta, z\sin\theta)\, zd\theta$$

$$= z\int_0^{2\pi} f_X(z\cos\theta) f_Y(z\sin\theta)\, d\theta$$

$$= z\int_0^{2\pi} \frac{e^{-\frac{z^2}{2\sigma^2}\cos^2\theta} . e^{-\frac{z^2}{2\sigma^2}\sin^2\theta}}{2\pi\sigma^2} d\theta$$

$$= \frac{ze^{-\frac{z^2}{2\sigma^2}}}{\sigma^2} \qquad z \geq 0$$

The above is the Rayleigh density function we discussed earlier.

Rician Distribution:

*S*uppos*e* X and Y are independent Gaussian variables with non zero mean μ_X and μ_Y respectively and constant variance. We have to find the joint density function of the random variable $Z = \sqrt{X^2 + Y^2}$.

- Envelope of a sinusoidal + a narrow band Gaussian noise.
- Received noise in a multipath situation.

Here

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

and

$$Z = \sqrt{X^2 + Y^2}$$

We have shown that

$$f_Z(z) = \int_0^{2\pi} f_{XY}(z\cos\theta, z\sin\theta)\, zd\theta$$

Suppose $\mu_X = \mu\cos\phi$ and $\mu_Y = \mu\sin\phi$. Then

$$f_{X,Y}\left(z\cos\theta, z\sin\theta\right) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left(\left(z\cos\theta - \mu\cos\phi\right)^2 + \left(z\sin\theta - \mu\,\mathrm{c}\sin\phi\right)^2\right)}$$

$$= \frac{1}{2\pi\sigma^2} e^{-\frac{(z^2+\mu^2)}{2\sigma^2}} e^{z\mu\cos(\theta-\phi)}$$

$$\therefore\ f_Z\left(z\right) = \int_0^{2\pi} \frac{1}{2\pi\sigma^2} e^{-\frac{(z^2+\mu^2)}{2\sigma^2}} e^{\frac{z\mu\cos(\theta-\phi}{\sigma^2})} z\,d\theta$$

$$= \frac{ze^{-\frac{(z^2+\mu^2)}{2\sigma^2}}}{2\pi\sigma^2} \int_0^{2\pi} e^{\frac{z\mu\cos(\theta-\phi}{\sigma^2})} d\theta$$

Joint Probability Density Function of two functions of two random variables

We consider the transformation $(g_1, g_2): \mathbb{R}^2 \to \mathbb{R}^2$. We have to find out the joint probability density function $f_{Z_1,Z_2}(z_1, z_2)$ where $Z_1 = g_1(X,Y)$ and $Z_2 = g_2(X,Y)$. We hve to find out the joint probability density function $f_{Z_1,Z_2}(z_1, z_2)$ where $z_1 = g_1(x, y)$ and $z_2 = g_2(x, y)$. Suppose the inverse mapping relation is

$x = h_1(z_1, z_2)$ and $y = h_2(z_1, z_2)$

Consider a differential region of area $dz_1\ dz_2$ at point (z_1, z_2) in the $Z_1 - Z_2$ plane.

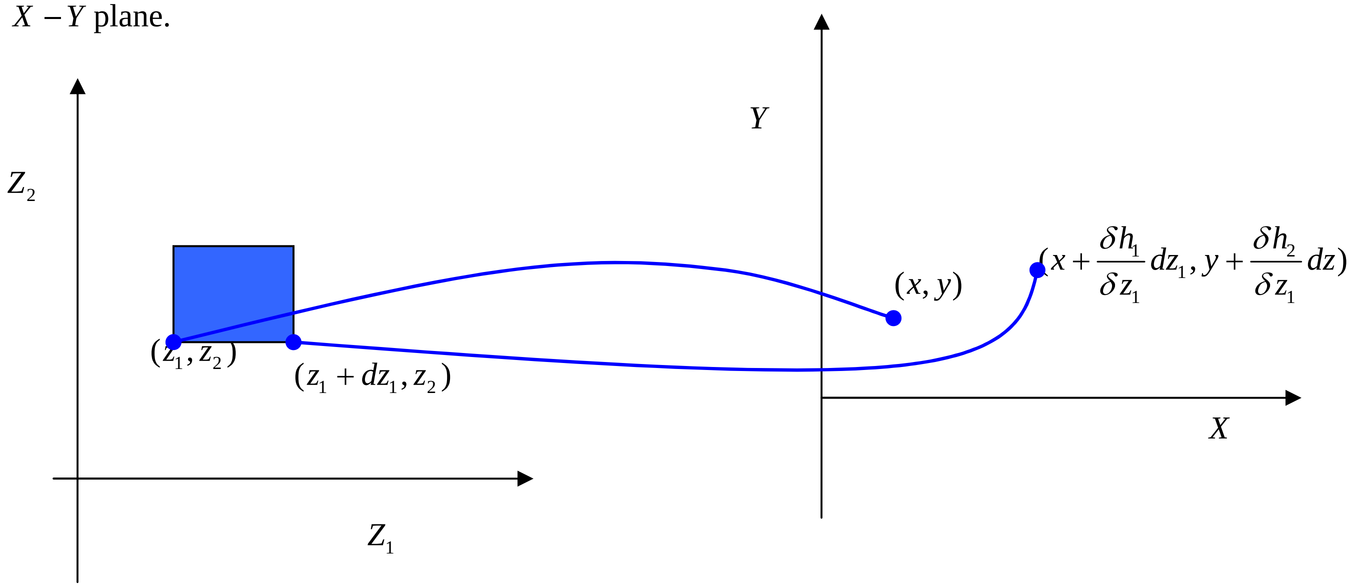
Let us see how the corners of the differential region are mapped to the $X - Y$ plane. Observe that

$$h_1(z_1 + dz_1, z_2) = h_1(z_1, z_2) + \frac{\delta h_1}{\delta z_1} dz_1 = x + \frac{\delta h_1}{\delta z_1} dz_1$$

$$h_2(z_1 + dz_1, z_2) = h_1(z_1, z_2) + \frac{\delta h_2}{\delta z_1} dz_1 = y + \frac{\delta h_2}{\delta z_1} dz_1$$

Therefore,

The point $(z_1 + dz_1, z_2)$ is mapped to the point $(x + \frac{\delta h_1}{\delta z_1} dz_1, y + \frac{\delta h_2}{\delta z_1} dz)$ in the $X - Y$ plane.



We can similarly find the points in the $X - Y$ plane corresponding to $(z_1, z_2 + dz_2)$ and $(z_1 + dz_1, z_2 + dz_2)$. The mapping is shown in Fig. We notice that each differential region in the $X - Y$ plane is a parallelogram. It can be shown the differential parallelogram at (x, y) has a area $|J(z_1, z_2)| dz_1 dz_2$ where $J(z_1, z_2)$ is the *Jacobian* of the transformation defined as the determinant

$$J(z_1, z_2) = \begin{vmatrix} \dfrac{\delta h_1}{\delta z_1} & \dfrac{\delta h_1}{\delta z_2} \\ \dfrac{\delta h_2}{\delta z_1} & \dfrac{\delta h_2}{\delta z_2} \end{vmatrix}$$

Further, it can be shown that the absolute values of the Jacobians of the forward and the inverse transform are inverse of each other so that

$$|J(z_1, z_2)| = \frac{1}{|J(x, y)|}$$

where

$$J(x, y) = \begin{vmatrix} \dfrac{\delta g_1}{\delta x} & \dfrac{\delta g_1}{\delta y} \\ \dfrac{\delta g_2}{\delta x} & \dfrac{\delta g_2}{\delta y} \end{vmatrix}$$

Therefore, the differential parallelogram in Fig. has an area of $\dfrac{dz_1 dz_2}{|J(x, y)|}$.

Suppose the transformation $z_1 = g_1(x, y)$ and $z_2 = g_2(x, y)$ has n roots and let $(x_i, y_i),\ i = 1, 2, ..n$ be the roots. The inverse mapping of the differential region in the $X - Y$ plane will be n differential regions corresponding to n roots. The inverse mapping is illustrated in the following figure for $n = 4$. As these parallelograms are non-overlapping,

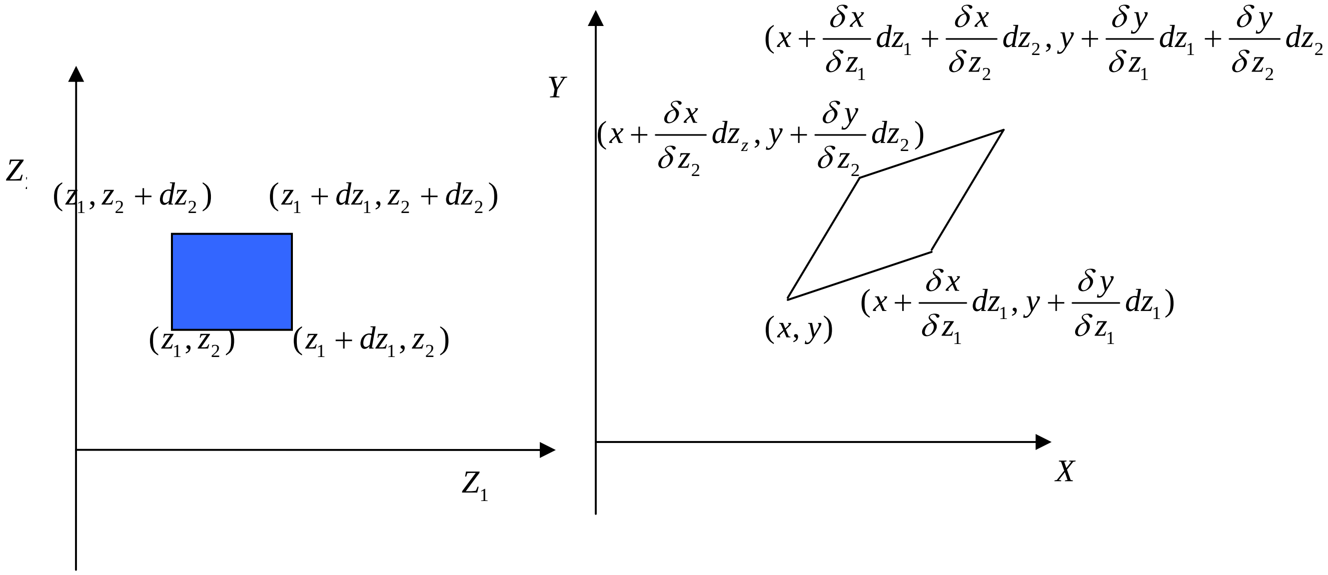
$$f_{Z_1, Z_2}(z_1, z_2)\, dz_1 dz_2 = \sum_{i=1}^{n} f_{X,Y}(x, y) \frac{dz_1 dz_2}{|J(x_i, y_i)|}$$

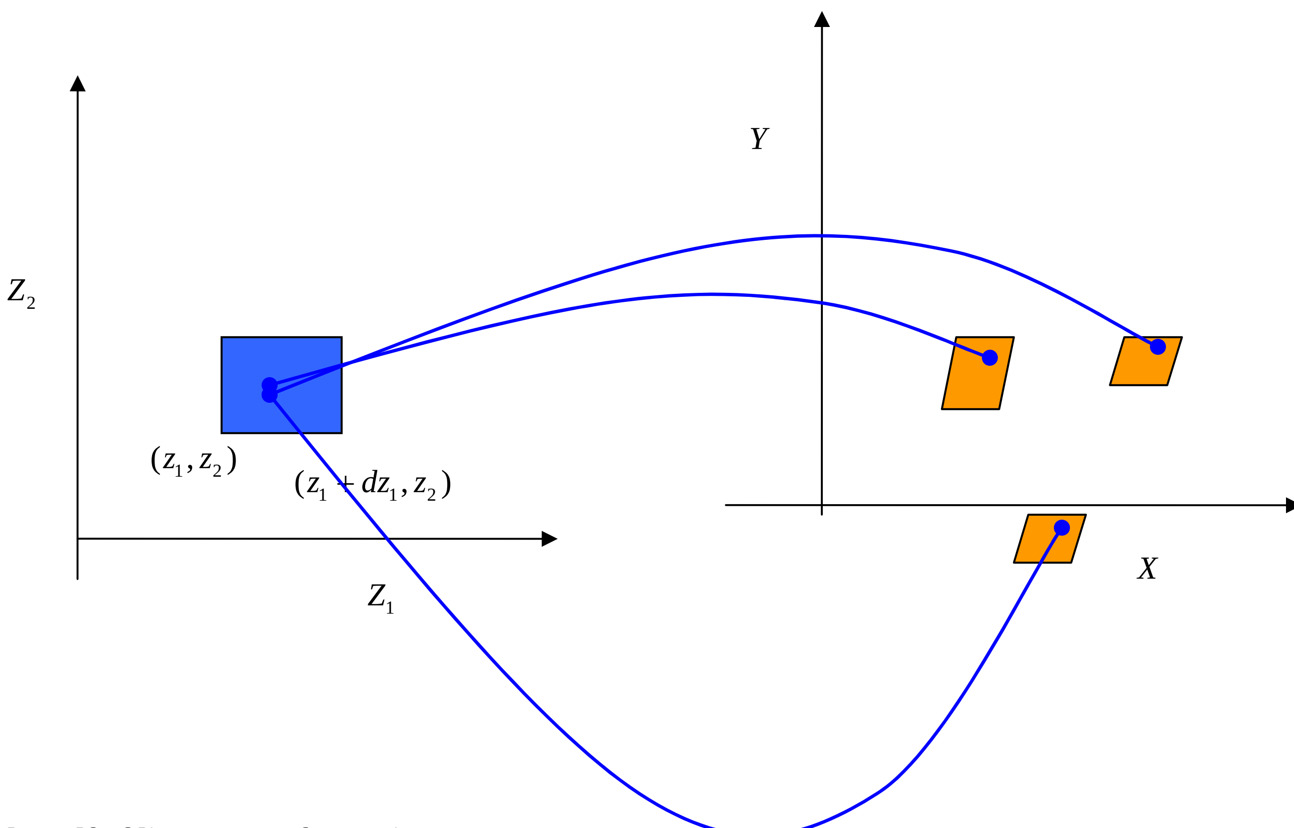
$$\therefore \qquad f_{Z_1, Z_2}(z_1, z_2) = \sum_{i=1}^{n} \frac{f_{X,Y}(x, y)}{|J(x_i, y_i)|}$$

Remark

- If $z_1 = g_1(x, y)$ and $z_2 = g_2(x, y)$ does not have a root in (x, y),

  then $\qquad f_{Z_1, Z_2}(z_1, z_2) = 0.$





Example: pdf of linear transformation

$Z_1 = aX + bY$

$Z_2 = cX + dY$

Then

$$x=\frac{dz_1-bz_2}{ad-bc},\ y=\frac{az_2-cz_1}{ad-bc}$$

$$J(x,y)=\begin{vmatrix} a & b \\ c & d \end{vmatrix}=ad-bc$$

Suppose X and Y are two independent Gaussian random variables each with mean 0 and variance σ^2. Given $R=\sqrt{X^2+Y^2}$ and $\theta=\tan^{-1}\frac{Y}{X}$, find $f_R(r)$ and $f_\theta(\theta)$.

Solution:

We have $x=r\cos\theta$ and $y=r\sin\theta$ so that $r^2=x^2+y^2$ ………… (1)

and $\tan\theta=\frac{y}{x}$ ……………… (2)

From (1)

$$\frac{\partial r}{\partial x}=\frac{x}{r}=\cos\theta$$

and $$\frac{\partial r}{\partial y}=\frac{y}{r}=\sin\theta$$

From (2)

$$\frac{\partial\theta}{\partial x}=\frac{-y}{x^2+y^2}=-\frac{\sin\theta}{r}$$

$$\frac{\partial\theta}{\partial y}=\frac{x}{x^2+y^2}=\frac{\cos\theta}{r}$$

$$\therefore |J|=\left|\det\begin{bmatrix}\cos\theta & \sin\theta \\ -\frac{\sin\theta}{r} & \frac{\cos\theta}{r}\end{bmatrix}\right|=\frac{1}{r}$$

$$\therefore f_{R,\theta}(r,\theta)=\left.\frac{f_X(x,y)}{|J|}\right]_{\substack{x=r\cos\theta\\ y=r\sin\theta}}$$

$$=\frac{r}{2\pi\sigma^2}e^{-\frac{r^2}{2\sigma^2}\cos^2\theta}.e^{-\frac{r^2}{2\sigma^2}\sin^2\theta}$$

$$=\frac{r}{2\pi\sigma^2}e^{-\frac{r^2}{2\sigma^2}}$$

$$\therefore f_R(r) = \int_0^{2\pi} f_{R,\theta}(r,\theta)d\theta$$

$$= \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \quad 0 \le r < \infty$$

$$f_\theta(\theta) = \int_0^\infty f_{R,\theta}(r,\theta)dr$$

$$= \frac{1}{2\pi\sigma^2} \int_0^\infty r e^{-\frac{r^2}{2\sigma^2}} dr$$

$$= \frac{1}{2\pi} \quad 0 \le \theta \le 2\pi$$

Rician Distribution:

- X and Y are independent Gaussian variables with non zero mean μ_X and μ_Y respectively and constant variance.

- We have to find the joint density function of the random variable $Z = \sqrt{X^2 + Y^2}$.
- Envelope of a sinusoidal + a narrow band Gaussian noise.
- Received noise in a multipath situation.

$$Z = \sqrt{X^2 + Y^2}$$
$$\phi = \tan^{-1} \frac{Y}{X}$$

We have to find $J(x, y)$ corresponding to z and ϕ .

$$J(x,y) = \det \begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial y} \end{bmatrix}$$

From $\quad z = \sqrt{x^2 + y^2} \quad$ and $\tan\phi = \frac{y}{x}$

We have, $\quad z^2 = x^2 + y^2 \quad$ and $\tan\phi = \frac{y}{x}$

Therefore,

$$\frac{\partial z}{\partial x} = \frac{x}{z} = \cos\phi$$

and $$\frac{\partial z}{\partial y} = \frac{y}{z} = \sin\phi$$

also $$\frac{\partial \phi}{\partial x} = -\frac{y}{x^2 \sec^2\phi} = -\frac{y}{x^2}\cos^2\phi$$

and $$\frac{\partial \phi}{\partial y} = \frac{y}{x \sec^2\phi} = \frac{1}{x}\cos^2\phi$$

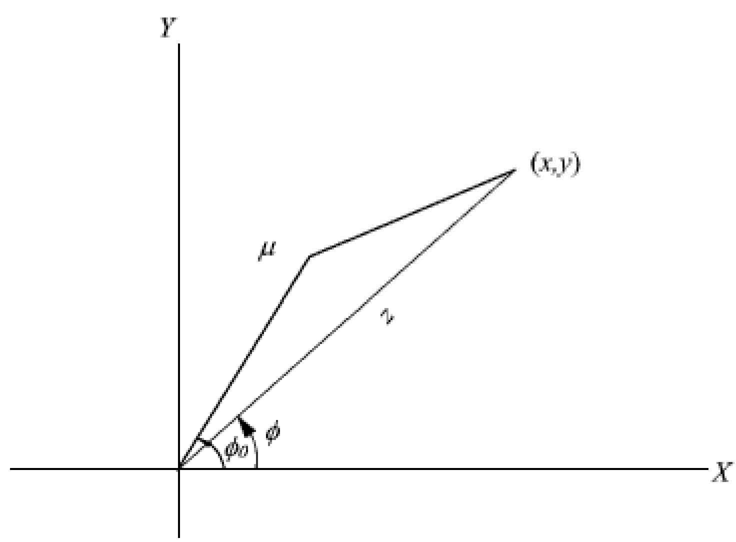
$$\therefore J(x,y) = \det\begin{bmatrix} \cos\phi & \sin\phi \\ -\dfrac{y^2\cos^2\phi}{x^2} & \dfrac{1}{x}\cos^2 x \end{bmatrix}$$

$$= \frac{\cos^3\phi}{x} + \frac{y\sin\phi\cos^2\phi}{x^2}$$

$$= \cos^2\phi\left(\frac{x\cos\phi + y\sin\phi}{x^2}\right)$$

$$= \frac{z\cos^2\phi}{x^2} = \frac{1}{z}$$

Consider the transformation as shown in the diagram below:



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

We have to find the density at (x, y) corresponding to z and ϕ.

From the above figure,

$$X-\mu_X = Z\cos\phi-\mu\cos\phi_0$$

and

$$Y-\mu_Y = Z\sin\phi-\mu\sin\phi_0$$

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{1}{2\sigma^2}\left((z\cos\phi-\mu\cos\phi_0)^2+(z\sin\phi-\mu\sin\phi)^2\right)}$$

$$=\frac{1}{2\pi\sigma^2}e^{-\frac{1}{2\sigma^2}\left(z^2-2z\mu\cos(\phi-\phi_0)+\mu^2\right)}$$

$$=\frac{1}{2\pi\sigma^2}e^{-\frac{z^2+\mu^2}{2\sigma^2}}e^{\frac{z\mu}{\sigma^2}\cos(\phi-\phi_0)}$$

Expected Values of Functions of Random Variables

Recall that

- If $Y = g(X)$ is a function of a continuous random variable X, then
  $$EY = Eg(X) = \int_{-\infty}^{\infty} g(x) f_X(x)dx$$
- 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

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

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

$$\{z < Z \le z + \Delta z\} = \bigcup_{i=1}^{n} \{(x_i, y_i) \in \Delta D_i\}$$

where

ΔD_i is the differential region containing (x_i, y_i). The mapping is illustrated in Fig. for $n = 3$.

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

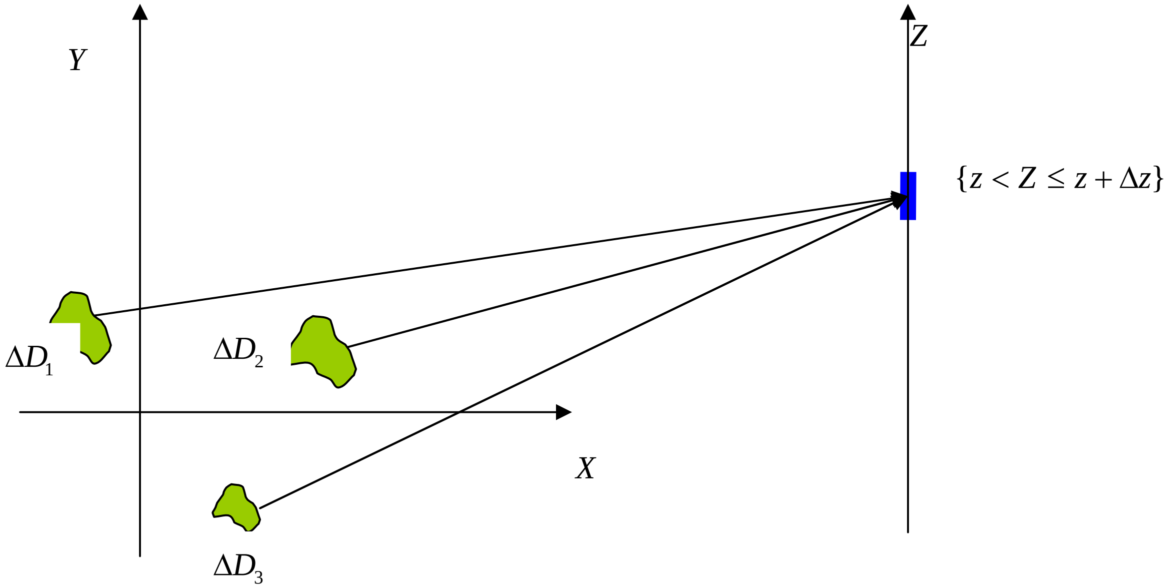
$$\therefore zf_Z(z)\Delta z = \sum_{(x_i, y_i) \in \Delta D_i} zf_{X,Y}(x_i, y_i)\Delta x_i \Delta y_i$$
$$= \sum_{(x_i, y_i) \in \Delta D_i} g(x_i, y_i) f_{X,Y}(x_i, y_i)\Delta x_i \Delta y$$

As z is varied over the entire the entire Z axis, the corresponding (nonoverlapping) differential regions in $X - Y$ plane covers the entire plane.

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

Thus,

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



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

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

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

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

$$=0 \ \text{ otherwise}$$

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

$$Eg(X,Y)=EX^2Y$$

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

$$=\int_0^2\int_0^2 x^2y\frac{1}{4}xydxdy$$

$$=\frac{1}{4}\int_0^2 x^3dx\int_0^2 y^2dy$$

$$=\frac{1}{4}\times\frac{2^4}{4}\times\frac{2^3}{3}$$

$$=\frac{8}{3}$$

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

$$EZ = aEX + bEY$$

Proof:

$$EZ = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)f_{X,Y}(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} axf_{X,Y}(x,y)dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} byf_{X,Y}(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} ax \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx + \int_{-\infty}^{\infty} by \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy$$

$$= a\int_{-\infty}^{\infty} xf_X(x)dx + b\int_{-\infty}^{\infty} yf_Y(y)dy$$

$$= aEX + bEY$$

Thus, expectation is a linear operator.

Example:

Consider the discrete random variables X and Y discussed in Example .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		

Clearly $EXY = \sum_{x,y \in \mathbb{R}_X \times \mathbb{R}Y} \sum g(x,y)p_{X,Y}(x,y)$

$$= 1\times1\times0.35 + 1\times2\times0.01$$

$$= 0.37$$

Remark

(1) We have earlier shown that expectation is a linear operator. We can generally write

$$E[a_1g_1(X,Y) + a_2g_2(X,Y)] = a_1Eg_1(X,Y) + a_2Eg_2(X,Y)$$

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

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

Joint Moments of Random Variables

Just like the moments of a random variable provides 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

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

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

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. We will give an interpretation of $Cov(X,Y)$ and $\rho(X,Y)$ later on.

We will show that $|\rho(X,Y)| \leq 1.$ To establish the relation, we prove the following result:

For two random variables X and Y, $E^2(XY) \leq EX^2 EY^2$

<u>**Proof:**</u>

Consider the random variable $Z = aX + Y$

$$\begin{aligned} &E(aX + Y)^2 \geq 0 \\ \Rightarrow\ &a^2 EX^2 + EY^2 + 2aEXY \geq 0 \end{aligned}.$$

Non-negatively of the left-hand side => 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 $\dfrac{E^2XY}{EX^2} + EY^2 - 2\dfrac{E^2XY}{EX^2}$

Minimum is nonnegative =>

$$\begin{aligned} &EY^2 - \frac{E^2XY}{EX^2} \geq 0 \\ \Rightarrow\ &E^2XY \leq EX^2\, EY^2 \\ \Rightarrow\ &\boxed{EXY \leq \sqrt{EX^2}\,\sqrt{EY^2}} \end{aligned}$$

Now

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

$$\therefore\left|\rho(X,Y)\right|=\frac{\left|E(X-\mu_X)(Y-\mu_Y)\right|}{\sqrt{E(X-\mu_X)^2E(Y-\mu_Y)^2}}$$

$$\leq\frac{\sqrt{E(X-\mu_X)^2}\sqrt{E(Y-\mu_Y)^2}}{\sqrt{E(X-\mu_X)^2E(Y-\mu_Y)^2}}$$

$$=1$$

$$\left|\rho(X,Y)\right|\leq 1$$

Uncorrelated random variables

Two random variables X and Y are called uncorrelated if

$$Cov(X,Y)=0$$

which also means

$$E(XY)=\mu_X\mu_Y$$

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

$$EXY=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_{X,Y}(x,y)dxdy \qquad \text{assuming } X \text{ and } Y \text{ are continuous}$$

Then
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_X(x)f_Y(y)dxdy$$

$$=\int_{-\infty}^{\infty}xf_X(x)dx\int_{-\infty}^{\infty}yf_Y(y)dy$$

$$=EXEY$$

Thus two independent random variables are always uncorrelated.

The converse is not always true.

(3) Two random variables may be dependent, but still they may be uncorrelated. If there exists correlation between two random variables, one may be represented as a linear regression of the others. We will discuss this point in the next section.

Linear prediction of Y from X

$\hat{Y}=aX+b$ Regression

Prediction error $Y-\hat{Y}$

Mean square prediction error

$$E(Y-\hat{Y})^2 = E(Y-aX-b)^2$$

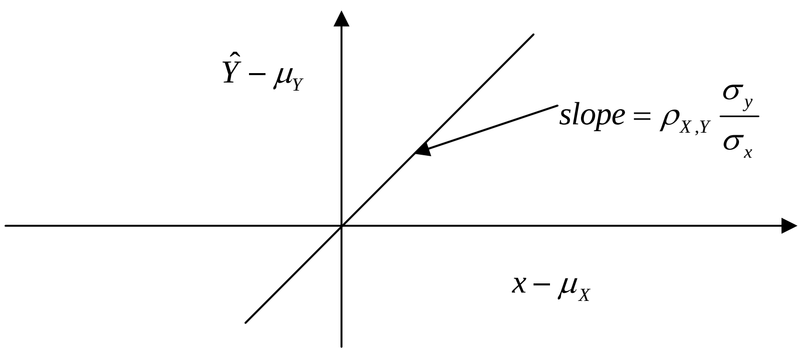
For minimising the error will give optimal values of a and b. Corresponding to the optimal solutions for a and b, we have

$$\frac{\partial}{\partial a}E(Y-aX-b)^2 = 0$$
$$\frac{\partial}{\partial b}E(Y-aX-b)^2 = 0$$

Solving for a and b, $\quad \hat{Y}-\mu_Y = \frac{1}{\sigma_X^2}\sigma_{X,Y}(x-\mu_X)$

so that $\hat{Y}-\mu_Y = \rho_{X,Y}\frac{\sigma_y}{\sigma_x}(x-\mu_X)$

where $\boxed{\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}}$ is the **correlation coefficient**.



Remark

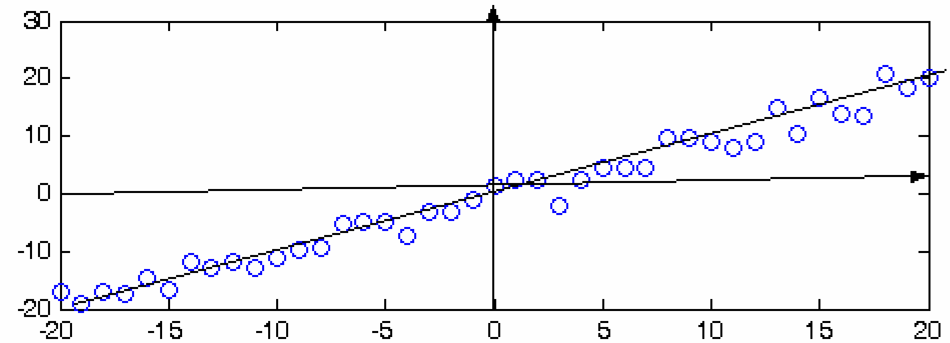
If $\rho_{X,Y} > 0$, then X and Y are called *positively correlated*.

If $\rho_{X,Y} < 0$, then X and Y are called *negatively correlated*

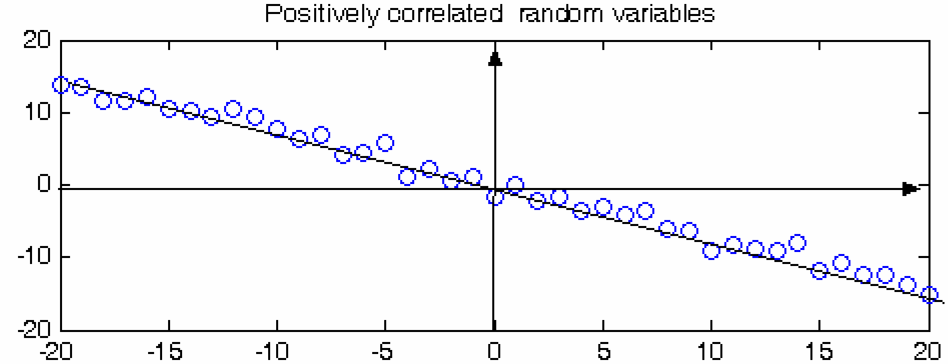
If $\rho_{X,Y} = 0$ then X and Y are uncorrelated.

$$\Rightarrow \hat{Y}-\mu_Y = 0$$
$$\Rightarrow \hat{Y} = \mu_Y \text{ is the best prediction.}$$

(To be labeled and animated)

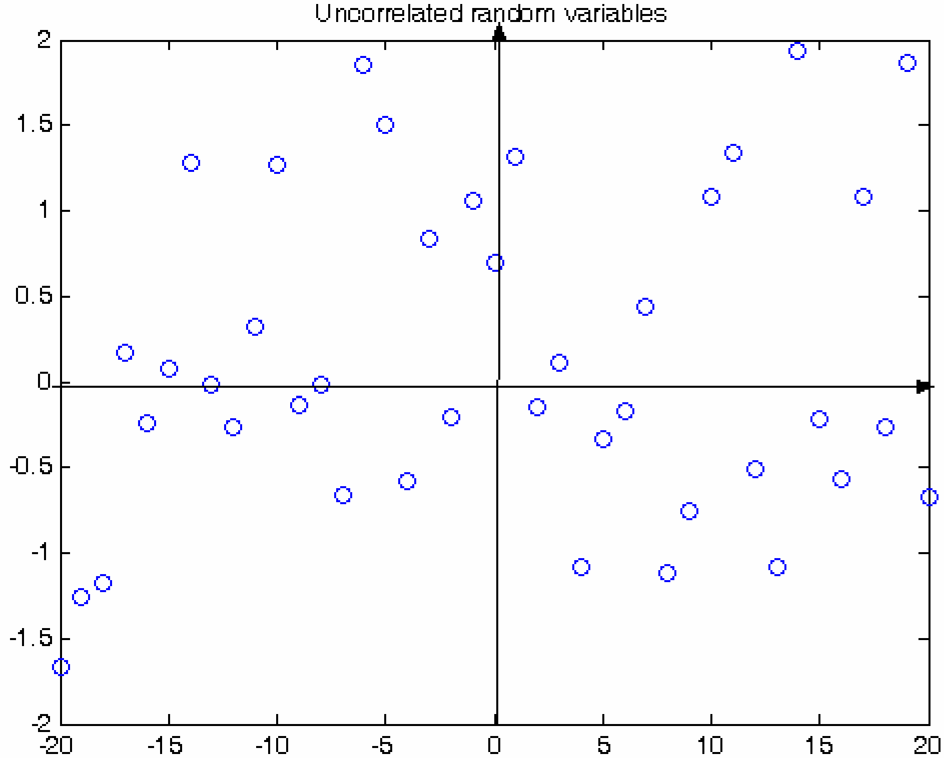


Positively correlated random variables



Negatively Correlated Random variables

Uncorrelated random variables



If $\rho_{X,Y} = 0$ then X and Y are uncorrelated.

$$=> \hat{Y} - \mu_Y = 0$$
$$=> \hat{Y} = \mu_Y \text{ is the best prediction.}$$

Note that independence => Uncorrelatedness. But uncorrelated generally does not imply independence (except for jointly Gaussian random variables).

Example :

$Y = X^2$ and $f_X(x)$ is uniformly distributed between (1,-1).

X and Y are dependent, but they are uncorrelated.

Because
$$\begin{aligned} Cov(X,Y) = \sigma_X &= E(X-\mu_X)(Y-\mu_Y) \\ &= EXY = EX^3 = 0 \\ &= EXEY \ (\because \text{EX} = 0) \end{aligned}$$

In fact for any zero- mean symmetric distribution of X, X and X^2 are uncorrelated.

(4) is a linear estimator $\hat{X} = aY$

Jointly Gaussian Random variables

Many practically occurred random variables are modeled as jointly Gaussian random variables. For example, noise occurring in the communication systems is modeled as jointly Gaussian random variables.

Two random variables X and Y are called jointly Gaussian if their joint density function is

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

$$-\infty < x < \infty, \ -\infty < y < \infty$$

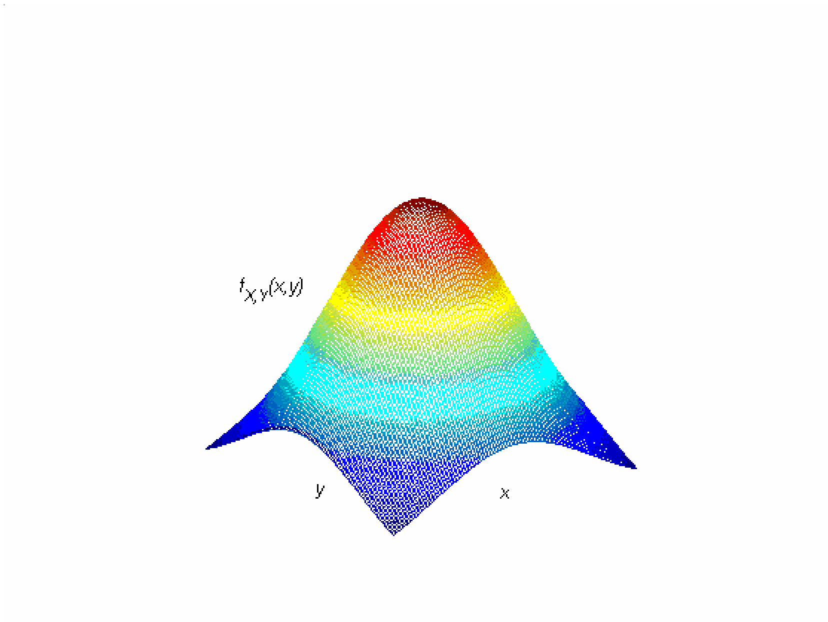
The joint pdf is determined by 5 parameters

- 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 pdf has a bell shape centred at (μ_X, μ_Y) as shown in the Fig. 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.



Properties of jointly Gaussian random variables

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

We have

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

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

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

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

Area under a Gaussian

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

Similarly

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}$$

(2) The converse of the above result is not true. If each of X and Y is Gaussian, X and Y are not necessarily jointly Gaussian. Suppose

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

$f_{X,Y}(x,y)$ in this example is non-Gaussian and qualifies to be a joint pdf. Because,

$f_{X,Y}(x,y)\geq 0$ and

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}(1+\sin x\sin y)dydx\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}dydx+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}\sin x\sin ydydx\\
&=1+\frac{1}{2\pi\sigma_X\sigma_Y}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}}\sin xdx\underbrace{\int_{-\infty}^{\infty}e^{-\frac{1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}}\sin ydy}_{\text{Odd function in } y}\\
&=1+0\\
&=1
\end{aligned}$$

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

$$\begin{aligned}
f_X(x)&=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}(1+\sin x\sin y)dy\\
&=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}dy+\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_X\sigma_Y}\underbrace{e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}\sin x\sin y}_{\text{Odd function in } y}dy\\
&=\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}+0\\
&=\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}
\end{aligned}$$

Similarly, $f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}$

Thus X and Y are both Gaussian, but not jointly Gaussian.

(3) If X and Y are jointly Gaussian, then for any constants a and b, then the random variable

Z, given by $Z=aX+bY$ is Gaussian with mean $\mu_Z=a\mu_X+b\mu_Y$ and variance

$$\sigma_Z{}^2 = a^2\sigma_X{}^2 + b^2\sigma_Y{}^2 + 2ab\sigma_X\sigma_Y\rho_{X,Y}$$

(4) Two jointly Gaussian RVs X and Y are independent if and only if X and Y are uncorrelated $(\rho_{X,Y} = 0)$. Observe that if X and Y are uncorrelated, then

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} \\ &= \frac{1}{\sqrt{2\pi}\sigma_X} e^{\frac{(x-\mu_X)^2}{\sigma_X^2}} \frac{1}{\sqrt{2\pi}\sigma_Y} e^{\frac{(y-\mu_Y)^2}{\sigma_Y^2}} . \\ &= f_X(x) f_Y(y) \end{aligned}$$

Joint Characteristic Functions of Two Random Variables

The *joint characteristic function* of two random variables X and Y is defined by

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

If X and Y are jointly continuous random variables, then

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

Note that $\phi_{X,Y}(\omega_1, \omega_2)$ is same as the two-dimensional Fourier transform with the basis function $e^{j\omega_1 x + j\omega_2 y}$ instead of $e^{-(j\omega_1 x + j\omega_2 y)}$.

$f_{X,y}(x, y)$ is related to the joint characteristic function by the Fourier inversion formula

$$f_{X,y}(x, y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_{X,Y}(\omega_1, \omega_2)e^{-j\omega_1 x - j\omega_2 y} d\omega_1 d\omega_2$$

If X and Y are discrete random variables, we can define the joint characteristic function in terms of the joint probability mass function as follows:

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

Properties of Joint Characteristic Functions of Two Random Variables

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,

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

Hence,

$$1 = \phi_{X,Y}(0,0)$$

$$EX = \frac{1}{j}\frac{\partial}{\partial\omega_1}\phi_{X,Y}(\omega_1,\omega_2)\Big|_{\omega_1=0}$$

$$EY = \frac{1}{j}\frac{\partial}{\partial\omega_2}\phi_{X,Y}(\omega_1,\omega_2)\Big|_{\omega_2=0}$$

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

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

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

Example 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

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

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

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

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

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

Area under a Gaussian

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

$$= e^{\mu_X j\omega-\sigma_X^{4}{}^2\omega^2/2}$$

If X and Y are 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 below:

Example 2 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

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

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

which is the characteristic function of a Gaussian random variable with

mean $\mu_Z = \mu_X + \mu_Y$ and

variance $\sigma_Z^2 = \sigma_X^2 + 2\rho_{X,Y}\sigma_X\sigma_Y + \sigma_Y^2$

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

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

Conditional Expectation

Recall that

- If X and Y are continuous random variables, then the conditional density function of Y given $X = x$ is given by

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

- If X and Y are discrete random variables, then the probability mass function Y given $X = x$ is given by

  $p_{Y/X}(y/x) = \mathrm{p}_{X,Y}(x,y)/p_X(x)$

The conditional expectation of Y given $X = x$ is defined by

$$\mu_{Y/X=x} = E(Y/X=x) = \begin{cases} \int_{-\infty}^{\infty} y f_{Y/X}(x/y) & \text{if } X \text{ and } Y \text{ are continuous} \\ \sum_{y \in R_Y} y p_{Y/X}(x/y) & \text{if } X \text{ and } Y \text{ are discrete} \end{cases}$$

Remark

- The conditional expectation of Y given $X = x$ is also called the conditional mean of Y given $X = x$.

- We can similarly define the conditional expectation of X given $X = x,$ denoted by $E(X/Y = y)$

- Higher-order conditional moments can be defined in a similar manner.

- Particulaly, the conditional variance of Y given $X = x$ is given by

$$\sigma^2_{Y/X=x} = E[(Y - \mu_{Y/X=x})^2 / X = x]$$

Example:

Consider the discrete random variables X and Y discussed in example .The joint probability mass function of the random variables are tabulated in Table . Find the joint expectation of $\mathrm{E}\,(Y/X = 2)$

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	

The conditional probability mass function is given by

$$p_{Y/X}(y/2) = p_{X,Y}(2,y)/p_X(2)$$

$$\therefore p_{Y/X}(0/2) = p_{X,Y}(2,0)/p_X(2)$$

$$= \frac{0.15}{0.16} = 15/16$$

and

$$p_{Y/X}(1/2) = p_{X,Y}(2,1)/p_X(2)$$

$$= \frac{0.01}{0.16} = 1/16$$

$$E(Y/x=2) = 0\times p_{Y/X}(0/2) + 1\times p_{Y/X}(1/2)$$

$$= 1/16$$

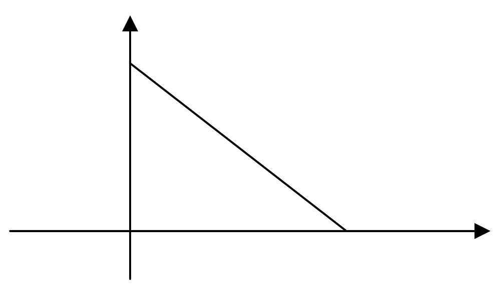
Example

Suppose X and Y are jointlyuniform random variables with the joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & x\geq 0, y\geq 0, x+y\leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(Y/X=x)$

From the figure, $f_{X,Y}(x,y) = \frac{1}{2}$ in the shaded area.



We have

$$\therefore\ f_X(x) = \int_0^{2-x} f_{X,Y}(x,y)dy$$
$$= \int_0^{2-x} \frac{1}{2}dy$$
$$= \frac{1}{2}(2-x) \quad 0 \le \mathrm{x} \le 2$$

$$\therefore\ f_{Y/X}(y/x) \quad = f_{X,Y}(x,y)/f_X(x)$$
$$= \frac{1}{2-x}$$
$$\therefore E(Y/X=x) = \int_{-\infty}^{\infty} y f_{Y/X}(y/x)dy$$
$$= \int_0^{2-x} y\frac{1}{2-x}dy$$
$$= \frac{2-x}{2}$$

Example

Suppose X and Y are jointly Gaussian random variables with the joint probability density function given by

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

We have to find $E(Y/X=x)$.

We have $f_{Y/X}(y/x) = f_{X,Y}(x,y)/f_X(x)$

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

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

Therefore,

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

$$= \mu_Y + \frac{\sigma_Y\rho_{X,Y}}{\sigma_X}(x-\mu_x)$$

Conditional Expectation as a random variable

Note that $E(Y/X=x)$ is a function of x.

Using this function, we may define a random variable $\phi(X)=E(Y/X)$ Thus we may consider EY/X as a function of the random variable X and $E(Y/X=x)$ as the value of $E(Y/X)$ at $X=x$.

$\boxed{E(Y/X) \text{ is a random variable, } E(Y/X=x) \text{ is the value of } E(Y/X) \text{ at } X=x}$

$EE(Y/X)=EY$

$EE(Y/X)=EY$

Proof:

$$
\begin{aligned}
EE(Y/X) &= \int_{-\infty}^{\infty} E(Y/X=x) f_X(x)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{Y/X}(x/y)dy\, f_X(x)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_X(x) f_{Y/X}(x/y)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dydx \\
&= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} y f_Y(y)dy \\
&= EY
\end{aligned}
$$

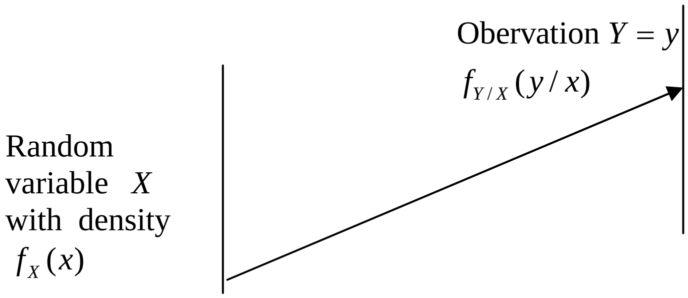
Thus $EE(Y/X)=EY$ and similarly

$EE(X/Y)=EX$

Baysean Estimation theory and conditional expectation

Consider two random variables X and Y with joint pdf $f_{X,Y}(x,y)$. Suppose Y is observable and $f_X(x)$ is known. We have to estimate X for a given value in some optimal sense. I in a sense that some values of θ are more likely (*a priori* information). We can represent this prior information in the form of a prior density function.

<u>In the folowing we omit the suffix in density functions just for notational simplicity.</u>



The conditional density function $f_{Y/X}(y/x)$ is called *likelihood function* in estimation terminology.

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

Also we have the Bayes rule

$$f_{X/Y}(x/y) = \frac{f_X(x) f_{Y/X}(y/x)}{f_Y(y)}$$

where $f_{\Theta/\mathbf{X}}(\theta)$ is the *a posteriori* density function

Suppose the optimam estimator $\hat{X}(Y)$ is a function of the random variable Y such that it minimizes the mean-square error

The Estimation error $E(\hat{X}(Y) - X)^2$. Such an estimator is known as the *minimum mean-square error estimator.*

Estimation problem is

Minimize $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\hat{X}(y) - x)^2 \; f_{X,Y}(x, y) dx\, dy$$

with respect to $\hat{\mathrm{X}}(\mathrm{y})$.

This is equivalent to minimizing

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\hat{X}(y)-x)^2 \; f_Y(y) f_{X/Y}(x/y)dx\,dy$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\hat{X}(y)-x)^2 \; f_{X/Y}(x/y)dx \; f_Y(y)dy$$

Since $f_Y(y)$ is always +ve, the above integral will be minimum if the inner integral is minimum. This results in the problem:

Minimize $\int_{-\infty}^{\infty}(\hat{X}(y)-x)^2 \; f_{X/Y}(x/y)dx$

with respect to $\hat{X}(y)$. The minimum is given by

$$\frac{\partial}{\partial\hat{X}(y)}\int_{-\infty}^{\infty}(\hat{X}(y)-x)^2 \; f_{X/Y}(x/y)dx \; = 0$$

$$\Rightarrow -2\int_{-\infty}^{\infty}(\hat{X}(y)-x) \; f_{X/Y}(x/y)dx = 0$$

$$\Rightarrow \int_{-\infty}^{\infty}\hat{X}(y) \; f_{X/Y}(x/y)dx = \int_{-\infty}^{\infty} x \; f_{X/Y}(x/y)dx = E(X/Y=y)$$

Multiple Random Variables

In many applications we have to deal with many 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 n random variables defined on the same sample space.

Joint CDF of n random variables

Consider n random variables $X_1, X_2,.., X_n$ defined on the same probability space $(S, \mathbb{F}, P).$ We define the random vector $\boldsymbol{X}$ as,

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_n \end{bmatrix} \qquad \text{or} \qquad \boldsymbol{X}' = \begin{bmatrix} X_1 & X_2 & . & . & X_n \end{bmatrix}$$

where $'$ indicates the transpose operation.

Thus an $n-$dimensional random vector $\mathbf{X}.$ is defined by the mapping $\mathbf{X}: S \to \mathbb{R}^n.$ A particular value of the random vector $\mathbf{X}.$ is denoted by

$\mathbf{x}$=[x_1 x_2.. x_n]'.

The CDF of the random vector $\mathbf{X}.$ is defined as the joint CDF of $X_1, X_2,.., X_n.$ Thus

$$F_{X_1,X_2,..,X_n}(x_1, x_2,..x_n) = F_{\mathbf{X}}(\mathbf{x})$$
$$= P(\{X_1 \le x_1, X_2 \le x_2,..X_n \le x_n\})$$

Some of the most important properties of the joint CDF are listed below. These properties are mere extensions of the properties of two joint random variables.

Properties of the joint CDF of n random variables

(a) $F_{X_1,X_2,..X_n}(x_1, x_2,..x_n)$ is a non-decreasing function of each of its arguments.

(b) $F_{X_1,X_2,..X_n}(-\infty, x_2,..x_n) = F_{X_1,X_2,..X_n}(x_1, -\infty,..x_n) = ... = F_{X_1,X_2,..X_n}(x_1, x_2,.. -\infty) = 0$

(c) $F_{X_1,X_2,..X_n}(\infty, \infty,..,\infty) = 1$

(d) $F_{X_1,X_2,..X_n}(x_1, x_2,..x_n)$ is right-continuous in each of its arguments.

(e) The marginal CDF of a random variable X_i is obtained from $F_{X_1,X_2,..X_n}(x_1,x_2,..x_n)$ by letting all random variables except X_i tend to ∞. Thus

$$F_{X_1}(x_1) = F_{X_1,X_2,..X_n}(x_1,\infty,..,\infty),$$

$$F_{X_2}(x_2) = F_{X_1,X_2,..X_n}(\infty,x_2,..,\infty)$$

and so on.

Joint pmf of n discrete random variables

Suppose $\mathbf{X}$ is a discrete random vector defined on the same probability space (S,F,P). Then $\mathbf{X}$ is completely specified by the joint probability mass function

$$P_{X_1,X_2,...,X_n}(x_1,x_2,..x_n) = P_{\mathbf{X}}(\mathbf{x})$$
$$= P(\{X_1 = x_1, X_2 = x_2,..X_n = x_n\})$$

Given $P_{X_1,X_2,...,X_n}(x_1,x_2,..x_n)$ we can find the *marginal probability mass function*

$$p_{X_1}(x_1) = \sum_{X_2}\sum_{X_3}\cdots\sum_{X_n} p_{X_2,X_3,...,X_n}(x_1,x_2,...,x_n)$$

n -1 summations

Joint PDF of n random variables

If $\mathbf{X}$ is a continuous random vector, that is, $F_{X_1,X_2,..X_n}(x_1,x_2,..x_n)$ is continuous in each of its argument, then $\mathbf{X}$ can be specified by the joint probability density function

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,X_2,..X_n}(x_1,x_2,..x_n)$$
$$= \frac{\partial^n}{\partial x_1 \partial x_2 ... \partial x_n} F_{X_1,X_2,..X_n}(x_1,x_2,..x_n)$$

Properties of joint PDF of n random variables

The joint pdf of n random variables satisfies the following important properties

(1) $f_{X_1,X_2,..X_n}(x_1,x_2,..x_n)$ is always a non-negaitive quantity. That is,

$$f_{X_1,X_2,..X_n}(x_1,x_2,..x_n) \geq 0 \quad \forall (x_1,x_2,..x_n) \in \mathbb{R}^n$$

(2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} f_{X_1,X_2,..X_n}(x_1,x_2,..x_n)dx_1dx_2...dx_n = 1$

(3) Given $f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,X_2,..X_n}(x_1,x_2,..x_n)$ for all $(x_1,x_2,..x_n) \in \mathbb{R}^n$, we can find the probability of a Borel set (region) $B \subseteq \mathbb{R}^n$,

$$P(\{(x_1,x_2,..x_n) \in B\}) = \int\int_{B}..\int f_{X_1,X_2,..X_n}(x_1,x_2,..x_n)dx_1dx_2...dx_n$$

(4) The marginal CDF of a random variable X_i is related to $f_{X_1,X_2,..X_n}(x_1,x_2,..x_n)$ by the $(n-1)-$ *fold* integral

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} f_{X_1,X_2,..X_n}(x_1, x_2, ..x_i..x_n)dx_1dx_2...dx_n$$

where the integral is performed over all the arguments except x_i.

Similarly,

$$f_{X_i,X_j}(x_i, x_j) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} f_{X_1,X_2,..X_n}(x_1, x_2, ..x_i..x_n)dx_1dx_2...dx_n \quad (n-2)-fold \text{ integral}$$

and so on.

The *conditional density functions* are defined in a similar manner. Thus

$$f_{X_{m+1},X_{m+2},......,X_n/X_1,X_2,.......,X_m}\left(x_{m+1}, x_{m+2}, ...x_n / x_1, x_2, ...,x_m\right) = \frac{f_{X_1,X_2,...,X_n}\left(x_1, x_2, ...,x_n\right)}{f_{X_1,X_2,...,X_m}\left(x_1, x_2, ...,x_m\right)}$$

Independent random variables:

The random variables $X_1, X_2, ..., X_n$ are called (mutually) independent if and only if

$$f_{X_1,X_2,..X_n}(x_1, x_2, ..x_n) = \prod_{i=1}^{n} f_{X_i}\left(x_i\right)$$

For example, if $X_1, X_2, ..., X_n$ are independent Gaussian random variables, then

$$f_{X_1,X_2,..X_n}(x_1, x_2, ..x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\frac{\left(x_i-\mu_i\right)^2}{\sigma_i^2}}$$

Remark, $X_1, X_2, ..., X_n$ may be pair wise independent, but may not be mutually independent.

Identically distributed random variables:

The random variables $X_1, X_2, ..., X_n$ are called identically distributed if each random variable has the same marginal distribution function, that is,

$$F_{X_1}\left(x\right) = F_{X_2}\left(x\right) = ... = F_{X_n}\left(x\right) \qquad \forall x$$

An important subclass of independent random variables is the independent and identically distributed (iid) random variables. The random variables $X_1, X_2, ..., X_n$ are called iid if $X_1, X_2, ..., X_n$ are mutually independent and each of $X_1, X_2, ..., X_n$ has the same marginal distribution function.

Example: If $X_1, X_2, ..., X_n$ may be iid random variables generated by n independent throwing of a fair coin and each taking values 0 and 1, then $X_1, X_2, ..., X_n$ are iid and

$$p_X\left(1,1,......,1\right) = \left(\frac{1}{2}\right)^n$$

Moments of Multiple random variables

Consider n jointly random variables represented by the random vector $\mathbf{X} = [X_1, X_2, ..., X_n]'$. The expected value of any scalar-valued function $g(\mathbf{X})$ is defined using the $n-$ *fold* integral as

$$E(g(\mathbf{X}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} g(x_1, x_2, ..x_n) f_{X_1, X_2, ..X_n}(x_1, x_2, ..x_n) dx_1 dx_2 ...dx_n$$

The mean vector of $\mathbf{X}$, denoted by $\boldsymbol{\mu}_{\mathbf{X}}$, is defined as

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}} &= E(\mathbf{X}) \\ &= \left[E(X_1)\ E(X_2)......E(X_n)\right]'. \\ &= \left[\mu_{X_1}\ \mu_{X_2} ...\ \mu_{X_n}\right]'.\end{aligned}$$

Similarly for each (i, j) $i = 1, 2, .., n,$ $j = 1, 2, .., n$ we can define the covariance

$$Cov(X_i, X_j) = E(X_i - \mu_{X_i})(X_j - \mu_{X_j})$$

All the possible covariances can be represented in terms of a matrix called the covariance matrix $\mathbf{C}_{\mathbf{X}}$ defined by

$$\begin{aligned}\mathbf{C}_{\mathbf{X}} &= E(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})' \\ &= \begin{bmatrix} \operatorname{var}(X_i) & \operatorname{cov}(X_1, X_2)\cdots & \operatorname{cov}(X_1, X_n) \\ \operatorname{cov}(X_2, X_1) & \operatorname{var}(X_2)\cdots. & \operatorname{cov}(X_2, X_n) \\ \vdots & \vdots & \vdots \\ \operatorname{cov}(X_n, X_1) & \operatorname{cov}(X_n, X_2)\cdots & \operatorname{var}(X_n) \end{bmatrix}\end{aligned}$$

Properties of the Covariance Matrix

- $\mathbf{C}_{\mathbf{X}}$ is a symmetric matrix because $Cov(X_i, X_j) = Cov(X_j, X_i)$

- $\mathbf{C}_{\mathbf{X}}$ is a *non-negative definite* matrix in the sense that for any real vector $\mathbf{z} \neq 0,$ the quadratic form $\mathbf{z}'\mathbf{C}_{\mathbf{X}}\mathbf{z} \geq 0.$ The result can be proved as follows:

$$\begin{aligned}\mathbf{z}'\mathbf{C}_{\mathbf{X}}\mathbf{z} &= \mathbf{z}'E((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})')\mathbf{z} \\ &= E(\mathbf{z}'(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})'\mathbf{z}) \\ &= E(\mathbf{z}'(\mathbf{X} - \mu_{\mathbf{X}}))^2 \\ &\geq 0\end{aligned}$$

The covariance matrix represents second-order relationship between each pair of the random variables and plays an important role in applications of random variables.

- The n random variables $X_1, X_2, .., X_n$ are called *uncorrelated* if for each (i, j) $i = 1, 2, .., n,$ $j = 1, 2, .., n$

$$Cov(X_i, X_j) = 0$$

If $X_1, X_2, .., X_n$ are uncorrelated, $\mathbf{C}_{\mathbf{X}}$ will be a diagonal matrix.

Multiple Jointly Gaussian Random variables

For any positive integer n, $X_1, X_2, ..., X_n$ represent n jointly random variables. These n random variables define a random vector $\mathbf{X} = [X_1, X_2, ..., X_n]'$. These random variables are called jointly Gaussian if the random variables $X_1, X_2, ..., X_n$ have joint probability density function given by

$$f_{X_1, X_2, ..., X_n}(x_1, x_2, ... x_n) = \frac{e^{-\frac{1}{2}\mathbf{X}'\mathbf{C}_X^{-1}\mathbf{X}}}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(\mathbf{C}_\mathbf{X})}}$$

where $\mathbf{C}_\mathbf{X} = E(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})'$ is the covariance matrix and $\boldsymbol{\mu}_\mathbf{X} = E(\mathbf{X}) = \left[E(X_1), E(X_2)......E(X_n)\right]'$ is the vector formed by the means of the random variables.

Vector space Interpretation of Random Variables

Consider a set $\mathbf{V}$ with elements called *vectors* and the field of real numbers $\mathbb{R}$.

$\mathbf{V}$ is called a vector space if and only if

1. An operation *vector addition* '+' is defined in $\mathbf{V}$ such that $(\mathbf{V},+)$ is a *commutative group*. Thus $(\mathbf{V},+)$ satisfies the following properties.

(i) For any pair of elements $\mathbf{v},\mathbf{w}\in\mathbf{V}$, there exists a unique element $(\mathbf{v}+\mathbf{w})\in\mathbf{V}$.

(ii) Vector addition is associative: $\mathbf{v}+(\mathbf{w}+\mathbf{z})=(\mathbf{v}+\mathbf{w})+\mathbf{z}$ for any three vectors $\mathbf{v},\mathbf{w},\mathbf{z}\in\mathbf{V}$

(iii) There is a vector $\mathbf{0}\in\mathbf{V}$ such that $\mathbf{v}+\mathbf{0}=\mathbf{0}+\mathbf{v}$ for any $\mathbf{v}\in\mathbf{V}$.

(iv) For any $\mathbf{v}\in\mathbf{V}$ there is a vector $-\mathbf{v}\in\mathbf{V}$ such that $\mathbf{v}+(\mathbf{-v})=\mathbf{0}=(\mathbf{-v})+\mathbf{v}$.

(v) for any $\mathbf{v},\mathbf{w}\in\mathbf{V}$, $\mathbf{v}+\mathbf{w}=\mathbf{w}+\mathbf{v}$.

2. For any element $\mathbf{v}\in\mathbf{V}$ and any $r\in\mathbb{R}$. the scalar product $r\mathbf{v}\in\mathbf{V}$
   This scalar product has the following properties for any $r,s\in\mathbb{R}$. and any $\mathbf{v},\mathbf{w}\in\mathbf{V}$,
3. $r(s\mathbf{v}) = (rs)\mathbf{v}$ for $r,s\in\mathbb{R}$ and $\mathbf{v}\in\mathbf{V}$
4. $\mathbf{r(v+w)=rv+rw}$
5. $\mathbf{(r+s)v=rv+sv}$
6. $1\mathbf{v}=\mathbf{v}$

It is easy to verify that the set of all random variables defined on a probability space $(S,\mathbb{F},P)$ forms a vector space with respect to addition and scalar multiplication. Similarly the set of all $n-\dim\textit{ensional}$ random vectors forms a vector space. Interpretation of random variables as elements of a vector space help in understanding many operations involving random variables.

Linear Independence

Consider n random vectors $\mathbf{v_1},\mathbf{v_2},....,\mathbf{v_N}$.

If $c_1\mathbf{v_1}+c_2\mathbf{v_2}+....+c_N\mathbf{v_N}=\mathbf{0}$ implies that

$c_1=c_2=....=c_N=0$, then $\mathbf{v_1},\mathbf{v_2},....,\mathbf{v_N}$ are *linearly independent*.

For N random vectors $\mathbf{X_1, X_2, ..., X_N}$, if $c_1\mathbf{X_1} + c_2\mathbf{X_2} + ... + c_N\mathbf{X_N} = \mathbf{0}$ implies that $c_1 = c_2 = ... = c_N = 0$, then $\mathbf{X_1, X_2, ..., X_N}$ are linearly independent.

Inner Product

If $\mathbf{v}$ and $\mathbf{w}$ are real vectors in a vector space $\mathbf{V}$ defined over the field $\mathbb{R}$, the inner product $<\mathbf{v}, \mathbf{w}>$ is a scalar such that

$\forall \mathbf{v}, \mathbf{w}, \mathbf{z} \in V$ and $r \in \mathbb{R}$

$$1. <\mathbf{v}, \mathbf{w}> = <\mathbf{w}, \mathbf{v}>$$
$$2. <\mathbf{v}, \mathbf{v}> = \|\mathbf{v}\|^2 \geq 0, \text{ where } \|\mathbf{v}\| \text{ is a } \textit{norm} \text{ inducedby the inner product}$$
$$3. <\mathbf{v}+\mathbf{w}, \mathbf{z}> = <\mathbf{v}, \mathbf{z}> + <\mathbf{w}, \mathbf{z}>$$
$$4. <r\mathbf{v}, \mathbf{w}> = r<\mathbf{v}, \mathbf{w}>$$

In the case of two random variables X and Y, the joint expectation EXY defines an inner product between X and Y. Thus

$$<X, Y> = EXY$$

We can easily verify that EXY satisfies the axioms of inner product.
The norm of a a random variable X is given by

$$\|X\|^2 = EX^2$$

For two $n-$dimensional random vectors $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, the inner product is

$$<\mathbf{X}, \mathbf{Y}> = E\mathbf{X}'\mathbf{Y} = \sum_i^n EX_iY_i$$

The norm of a a random vector $\mathbf{X}$ is given by

$$\|\mathbf{X}\|^2 = <\mathbf{X}, \mathbf{X}> = E\mathbf{X}'\mathbf{X} = \sum_i^n EX_i^2$$

- The set of RVs along with the inner product defined through the joint expectation operation and the corresponding norm defines a *Hilbert Space*.

Schwarz Inequality

For any two vectors $\mathbf{v}$ and $\mathbf{w}$ belonging to a Hilbert space $\mathbf{V}$

$$|<\mathbf{v},\mathbf{w}>|\leq \|\mathbf{v}\|\|\mathbf{w}\|$$

This means that for any two random variables X and Y

$$\left|E(XY)\right|\leq\sqrt{EX^2}\sqrt{EY^2}$$

Similarly for any two random vectors $\mathbf{X}$ and $\mathbf{Y}$

$$\left|E(\mathbf{X}'\mathbf{Y})\right|\leq\sqrt{E\mathbf{X}'\mathbf{X}}\sqrt{E\mathbf{Y}'\mathbf{Y}}$$

Orthogonal Random Variables and Orthogonal Random Vectors

Two vectors $\mathbf{v}$ and $\mathbf{w}$ are called orthogonal if $<\mathbf{v},\mathbf{w}>=0$

Two random variables X and Y are called orthogonal if $EXY=0$.

Similarly two random vectors $\mathbf{X}$ and $\mathbf{Y}$ are called orthogonal if

$$E\mathbf{X}'\mathbf{Y}=\sum_{i}^{n}EX_iY_i=0$$

Just like the of independent random variables and the uncorrelated random variables, the orthogonal random variables form an important class of random variables.

Remark

If X and Y are uncorrelated, then

$$E(X-\mu_X)(Y-\mu_Y)=0$$

$\therefore (X-\mu_X)$ is orthogonal to $(Y-\mu_Y)$

If each of X and Y is zero-mean

$$Cov(X,Y)=EXY$$

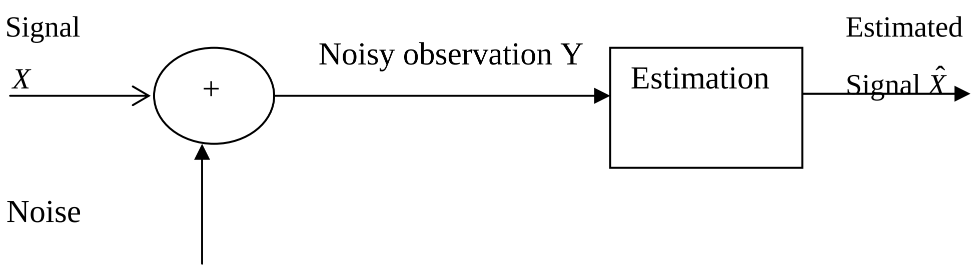
In this case, $EXY=0\Leftrightarrow Cov(XY)=0.$

Minimum Mean-square-error Estimation

Suppose X is a random variable which is not *observable* and Y is another observable random variable which is statistically dependent on X through the joint probability density function $f_{X,Y}(x, y)$. We pose the followingproblem.

Given a value of Y what is the best guess for X ?

This problem is known as the *estimation problem* and has many practical applications. One application is the signal estimation from noisy observations as illustrated in the Fig. below:



Let $\hat{X}(Y)$ be the estimate of the random variable X based on the random variable Y. Clearly $\hat{X}(Y)$ is a function of Y. We have to find best estimate be $\hat{X}(Y)$ in some meaningful sense. Observe that

- X is the unknown random variable
- $\hat{X}(Y)$ is the estimate of X.
- $X - \hat{X}(Y)$ is the estimation error.
- $E(X - \hat{X}(Y))^2$ is the mean of the square error.

One meaningful criterion is to minimize $E(X - \hat{X}(Y))^2$ with respect to $\hat{X}(Y)$ and the corresponding estimation principle is called *minimum mean square error principle*. Such a function which we want to minimize is called a *cost function* in o*ptimization theory*. For finding $\hat{X}(Y)$, we have to minimize the cost function

$$E(X-\hat{X}(Y))^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\hat{X}(y))^2 f_{X,Y}(x,y)dydx$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\hat{X}(y))^2 f_Y(y) f_{X/Y}(x)dydx$$
$$= \int_{-\infty}^{\infty} f_Y(y)(\int_{-\infty}^{\infty}(x-\hat{X}(y))^2 f_{X/Y}(x)dx)dy$$

Since $f_Y(y)$ is always positive, therefore the minimization of $E(X-\hat{X}(Y))^2$ with respect to $\hat{X}(Y)$ is equivalent to minimizing the inside integral $\int_{-\infty}^{\infty}(x-\hat{X}(y))^2 f_{X/Y}(x)dx$ with respect to $\hat{X}(Y)$. The condition for the minimum is

$$\frac{\partial}{\partial \hat{X}}\int_{-\infty}^{\infty}(x-\hat{X}(y))^2 f_{X/Y}(x/y)dx = 0$$

Or $2\int_{-\infty}^{\infty}(x-\hat{X}(y)) f_{X/Y}(x/y)dx = 0$

$$\Rightarrow \int_{-\infty}^{\infty}\hat{X}(y) f_{X/Y}(x/y)dx = \int_{-\infty}^{\infty} x f_{X/Y}(x/y)dx$$
$$\Rightarrow \hat{X}(y) = E(X/Y=y)$$

Thus, the minimum *mean-square error* estimation involves conditional expectation $E(X/Y=y)$. To find $E(X/Y=y)$, we have to determine the *a posteriori* probability density $f_{X/Y}(x/y)$ and perform $\int_{-\infty}^{\infty} x f_{X/Y}(x/y)dx$. These operations are computationally exhaustive when we have to perform these operations numerically.

Example Consider two zero-mean jointly Gaussian random variables X and Y with the joint pdf

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

The marginal density $f_Y(y)$ is a Gaussian random variable and given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{y^2}{2\sigma_Y^2}}$$

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

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

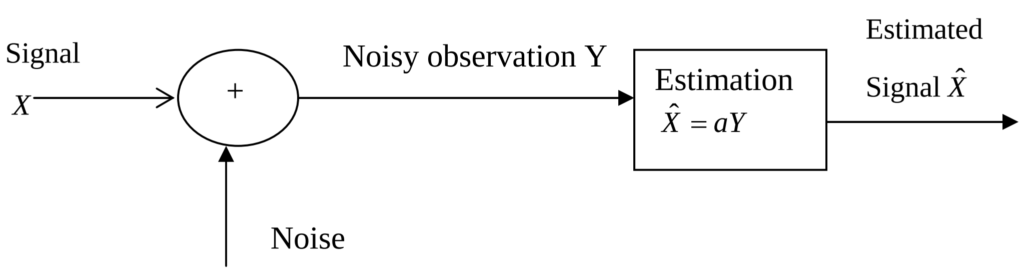
which is Gaussian with mean $\frac{\rho_{XY}\sigma_X}{\sigma_Y}y.$ Therefore, the MMSE estimator of X given $Y = y$ is given by

$$\hat{X}(y) = E(X/Y = y)$$
$$= \frac{\rho_{XY}\sigma_X}{\sigma_Y}y$$

This example illustrates that in the case of jointly Gaussian random variables X and Y, the mean-square estimator of X given $Y = y$, is linearly related with y. This important result gives us a clue to have simpler version of the mean-square error estimation problem discussed below.

Linear Minimum Mean-square-error Estimation and the Orthogonality Principle

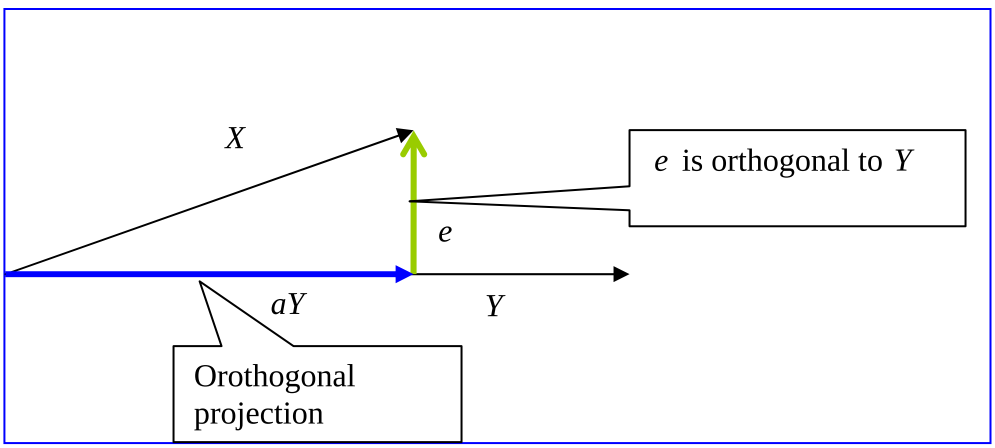
We assume that X and Y are both zero-mean and $\hat{X}(y) = ay.$ The estimation problem is now to find the optimal value for $a.$ Thus we have the *linear minimum mean-square error* criterion which minimizes $E(X-aY)^2$ with respect to $a.$



$$
\begin{aligned}
&\tfrac{d}{da}E(X-aY)^2=0\\
&\Rightarrow E\tfrac{d}{da}(X-aY)^2=0\\
&\Rightarrow E(X-aY)Y=0\\
&\Rightarrow EeY=0
\end{aligned}
$$

where e is the estimation error.

Thus the optimum value of a is such that *the estimation error* $(X-aY)$ *is orthogonal to the observed random variable* Y and the optimal estimator aY is the *orthogonal projection* of X on Y. This *orthogonality principle* forms the heart of a class of estimation problem called *Wiener filtering.* The orthogonality principle is illustrated geometrically in the following figure (Fig.).



The optimum value of a is given by

$$
\begin{aligned}
&E(X-aY)Y=0\\
\Rightarrow\ &EXY-aEY^2=0\\
\Rightarrow\ &a=\frac{EXY}{EY^2}
\end{aligned}
$$

The corresponding minimum linear mean-square error (LMMSE) is

$$
\begin{aligned}
LMMSE &= E(X-aY)^2 \\
&= E(X-aY)X - aE(X-aY)Y \\
&= E(X-aY)X - 0 \\
&\qquad (\,E(X-aY)Y = 0, \text{ using the } \textit{orthogonality principle}) \\
&= EX^2 - aEXY
\end{aligned}
$$

The orthogonality principle can be applied to optimal estimation of a random variable from more than one observation. We illustrate this in the following example.

Example Suppose X is a zero-mean random variable which is to be estimated from two zero-mean random variables Y_1 and Y_2. Let the LMMSE estimator be $\hat{X} = a_1Y_1 + a_2Y_2$. Then the optimal values of a_1 and a_2 are given by

$$\frac{\partial E(X - a_1Y_1 - a_2Y_2)^2}{\partial a_i} = 0 \qquad i = 1, 2.$$

This results in the *orthogonality* conditions

$$E(X - a_1Y_1 - a_2Y_2)Y_1 = 0$$

and

$$E(X - a_1Y_1 - a_2Y_2)Y_2 = 0$$

Rewriting the above equations we get

$$a_1EY_1^2 + a_2EY_2Y_1 = EXY_1$$

and

$$a_1EY_2Y_1 + a_2EY_2^2 + = EXY_2$$

Solving these equations we can find a_1 and a_2.

Further the corresponding minimum linear mean-square error (LMMSE) is

$$
\begin{aligned}
LMMSE &= E(X - a_1Y_1 - a_2Y_2)^2 \\
&= E(X - a_1Y_1 - a_2Y_2)X - a_1E(X - a_1Y_1 - a_2Y_2)Y_1 - a_2E(X - a_1Y_1 - a_2Y_2)Y_2 \\
&= E(X - a_1Y_1 - a_2Y_2)X - 0 \\
&\qquad (\text{ using the } \textit{orthogonality principle}) \\
&= EX^2 - a_1EXY_1 - a_2EXY_2
\end{aligned}
$$

Convergence of a sequence of random variables

Let $X_1, X_2, ..., X_n$ be a sequence n independent and identically distributed random variables. Suppose we want to estimate the mean of the random variable on the basis of the observed data by means of the relation

$$\mu_n = \frac{1}{n}\sum_{i=1}^{N} X_i$$

How closely does μ_n *represent the true mean* μ_X *as* n *is increased? How do we measure the closeness between* μ_n *and* μ_X *?*

Notice that μ_n is a random variable. *What do we mean by the statement* μ_n *converges to* μ_X ?

- Consider a deterministic sequence of real numbers $x_1, x_2, ...x_n....$ The sequence converges to a *limit* x if correspond to every $\varepsilon > 0$ we can find a positive integer N such that $|x - x_n| < \varepsilon$ for $n > N.$ For example, the sequence $1, \frac{1}{2}, ..., \frac{1}{n}, ...$ converges to the number 0.
- The **Cauchy criterion** gives the condition for convergence of a sequence without actually finding the limit. The sequence $x_1, x_2, ...x_n....$ converges if and only if , for every $\varepsilon > 0$ there exists a positive integer N such that $|x_{n+m} - x_n| < \varepsilon$ for all $n > N$ and all $m > 0.$

Convergence of a random sequence $X_1, X_2, ...X_n....$ cannot be defined as above. Note that for each $s \in S,$ $X_1(s), X_2(s), ...X_n(s)....$ represent a sequence of numbers . Thus $X_1, X_2, ...X_n....$ represents a family of sequences of numbers. Convergence of a random sequence is to be defined using different criteria. Five of these criteria are explained below.

Convergence Everywhere

A sequence of random variables is said to converge everywhere to X if

$|X(s) - X_n(s)| \to 0$ for $n > N$ and $\forall s \in S.$

Note here that the sequence of numbers for each sample point is convergent.

Almost sure (a.s.) convergence or convergence with probability 1

A random sequence $X_1, X_2, ...X_n, ...$ may not converge for every $s \in S.$

Consider the event $\{s \mid X_n(s) \to X\}$

The sequence $X_1, X_2, ...X_n, ...$ is said to converge to X *almost sure* or *with probability 1* if

$P\{s \mid X_n(s) \to X(s)\} = 1 \quad as \quad n \to \infty,$

or equivalently for every ε>0 there exists N such that

$P\{s \left| X_n(s) - X(s) \right| < \varepsilon$ for all for $n \geq N \ \} = 1$

We write $X_n \xrightarrow{a.s.} X$ in this case

One important application is the ***Strong Law of Large Numbers(SLLN)***:

If $X_1, X_2, ...X_n ....$ are independent and identically distributed random variables with a finite mean μ_X, then $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu_X$ with probability 1as $n \to \infty$.

Remark:

- $\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the sample mean.
- The strong law of large number states that the sample mean converges to the true mean as the sample size increases.
- The SLLN is one of the fundamental theorems of probability. There is a weaker versions of the law that we will discuss letter

Convergence in mean square sense

A random sequence $X_1, X_2, ...X_n ....$ is said to converge in the mean-square sense (m.s) to a random variable X if

$E(X_n - X)^2 \to 0 \quad as \quad n \to \infty$

X is called the mean-square limit of the sequence and we write

$l.i.m.\, X_n = X$

where $l.i.m.$ means limit in mean-square. We also write

$X_n \xrightarrow{m.s.} X$

- The following **Cauchy criterion** gives the condition for m.s. convergence of a random sequence without actually finding the limit. The sequence $X_1, X_2, ...X_n ....$ converges in m.s. if and only if , for every $\varepsilon > 0$ there exists a positive integer N such that

  $E\left[\left|x_{n+m} - x_n\right|\right]^2 \to 0$ as $n \to \infty$ for all $m > 0$.

Example :

If $X_1, X_2, ...X_n ....$ are iid random variables, then

$\frac{1}{n}\sum_{i=1}^{N} X_i \to \mu_X$ in the mean square 1as $n \to \infty$.

We have to show that $\lim_{n\to\infty} E(\frac{1}{n}\sum_{i=1}^{N} X_i - \mu_X)^2 = 0$

Now,

$$E(\frac{1}{n}\sum_{i=1}^{N} X_i - \mu_X)^2 = E(\frac{1}{n}(\sum_{i=1}^{N}(X_i - \mu_X))^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{N} E(X_i - \mu_X)^2 + \frac{1}{n^2}\sum_{\text{i=1}}^{\text{n}}\sum_{\text{j=1,j}\neq\text{i}}^{\text{n}} E(X_i - \mu_X)(X_j - \mu_X)$$

$$= \frac{n\sigma_X^2}{n^2} + 0 \text{ (Because of independence)}$$

$$= \frac{\sigma_X^2}{n}$$

$$\therefore \lim_{n\to\infty} E(\frac{1}{n}\sum_{i=1}^{N} X_i - \mu_X)^2 = 0$$

Convergence in probability

Associated with the sequence of random variables $X_1, X_2, ...X_n, ...,$ we can define a sequence of probabilities $P\{|X_n - X| > \varepsilon\}, n = 1, 2, ...$ for every $\varepsilon > 0$.

The sequence $X_1, X_2, ...X_n ....$ is said to convergent to X in probability if this sequence of probability is convergent that is

$$P\{|X_n - X| > \varepsilon\} \to 0 \quad as \quad n \to \infty$$

for every $\varepsilon > 0$. We write $X_n \xrightarrow{P} X$ to denote convergence in probability of the sequence of random variables $X_1, X_2, ...X_n ....$ to the random variable X.

If a sequence is convergent in mean, then it is convergent in probability also, because

$P\{|X_n - X|^2 > \varepsilon^2\} \le E(X_n - X)^2 / \varepsilon^2$ (Markov Inequality)

We have

$$P\{|X_n - X| > \varepsilon\} \le E(X_n - X)^2 / \varepsilon^2$$

If $E(X_n - X)^2 \to 0 \quad as \quad n \to \infty,$ (mean square convergent) then

$$P\{|X_n - X| > \varepsilon\} \to 0 \quad as \quad n \to \infty.$$

<u>Example:</u>

Suppose $\{X_n\}$ be a sequence of random variables with

$P\{X_n = 1\} = 1 - \frac{1}{n}$

and

$P\{X_n = -1\} = \frac{1}{n}$

Clearly

$$P\{|X_n - 1| > \varepsilon\} = P\{X_n = -1\} = \frac{1}{n} \to 0$$

$$as\ n \to \infty.$$

Therefore $\{X_n\} \xrightarrow{P} \{X = 0\}$

Thus the above sequence converges to a constant in probability.

Remark:

Convergence in probability is also called stochastic convergence.

Weak Laws of Large numbers

If $X_1, X_2, \ldots. X_n \ldots.$ are independent and identically distributed random variables, with *sample mean* $\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\mu_n \xrightarrow{P} \mu_X$ as $n \to \infty$.

We have

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\therefore E\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \mu_X$$

and

$$E(\mu_n - \mu_X)^2 = \frac{\sigma_X^2}{n} \text{ (as shown above)}$$

$$P\{|\mu_n - \mu_X| \geq \varepsilon\} \leq E(\mu_n - \mu_X)^2 / \varepsilon^2$$

$$= \frac{\sigma_X^2}{n\varepsilon^2}$$

$$\therefore P\{|\mu_n - \mu_X| \geq \varepsilon\} \to 0 \text{ as } n \to \infty.$$

Convergence in distribution

Consider the random sequence $X_1, X_2, \ldots. X_n \ldots.$ and a random variable X. Suppose $F_{X_n}(x)$ and $F_X(x)$ are the distribution functions of X_n and X respectively. The sequence is said to converge to X in distribution if

$$F_{X_n}(x) \to F_X(x) \qquad as \quad n \to \infty.$$

for all x at which $F_X(x)$ is continuous. Here the two distribution functions eventually coincide. We write $X_n \xrightarrow{d} X$ to denote convergence in distribution of the random sequence $X_1, X_2, ...X_n ....$ to the random variable X.

Example: Suppose $X_1, X_2, ...X_n ....$ is a sequence of RVs with each random variable X_i having the uniform density

$$f_{X_i}(x) = \begin{cases} \frac{1}{b} & x \le b \\ 0 \text{ other wise} \end{cases}$$

Define $Z_n = \max(X_1, X_2, ...X_n)$

We can show that

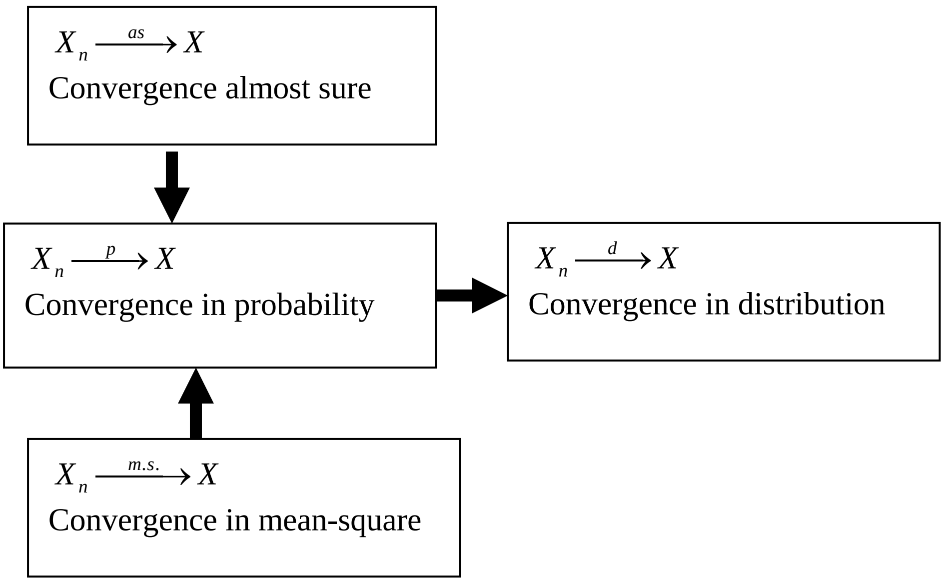
$$F_{Z_n}(z) = \begin{cases} = 0, & z < 0 \\ \frac{z^n}{a^n}, & 0 \le z < a \\ 1 & \text{otherwise} \end{cases}$$

Clearly,

$$\lim_{n\to\infty} F_{Z_n}(z) = F_Z(z) = \begin{cases} 0, & z < a \\ 1 & z \ge a \end{cases}$$

$\therefore \{z_n\}$ Converges to Z in *distribution.*

Relation between Types of Convergence



Central Limit Theorem

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

Form a random variable

$$Y_n = X_1 + X_2 + ...X_n$$

The mean and variances of Y_n are given by

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

and

$$\begin{aligned}\text{var}(Y_n) = \sigma^2_{Y_n} &= E\{\sum_{i=1}^{n}(X_i - \mu_i)\}^2 \\ &= \sum_{i=1}^{n} E(X_i - \mu_i)^2 + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} E(X_i - \mu_i)(X_j - \mu_j) \\ &= \sigma^2_{X_1} + \sigma^2_{X_2} + ... + \sigma^2_{X_n} \\ &\quad \because X_i \text{ **and** } X_j \text{ **are independent for** } (i \neq j)\end{aligned}$$

Thus we can determine the mean and variances 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 $\{Y_n = \sum_{i=1}^{n} X_i\}$ converges *in distribution* to $Y \sim N(\mu_Y, \sigma^2_Y)$ as $n \to \infty$. The conditions are:

1. The random variables $X_1, X_2, ..., X_n$ are independent with same mean and variance, but not identically distributed.
2. The random variables $X_1, X_2, ..., X_n$ are independent with different means and same variance and not identically distributed.
3. The random variables $X_1, X_2, ..., X_n$ 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:

Suppose $X_1, X_2 .... X_n$ is a sequence of independent and identically distributed random variables each with mean μ_X and variance σ_X^2 and $Y_n = \sum_{i=1}^{n} \frac{(X_i - \mu_X)}{\sqrt{n}}$. Then, the sequence $\{Y_n\}$ *converges in distribution* to a Gaussian random variable Y with mean 0 and variance σ_X^2. That is,

$$\lim_{n\to\infty} F_{Y_n}(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma_X} e^{-u^2/2\sigma_X^2} du$$

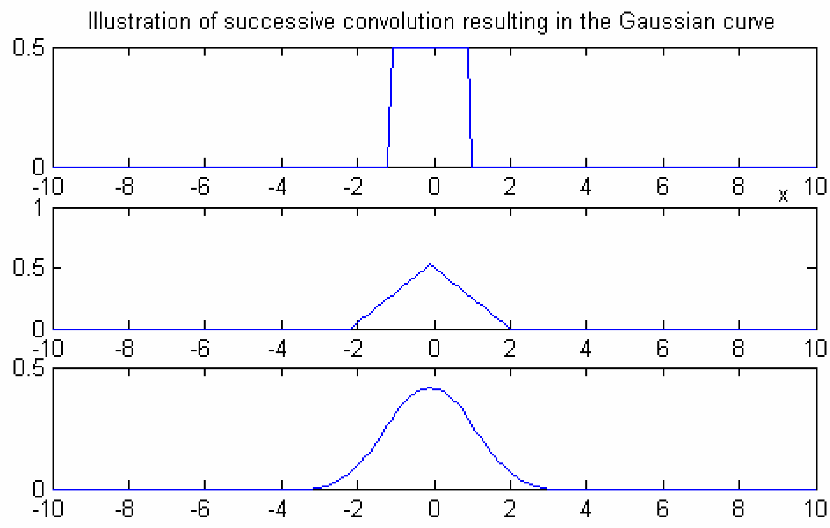
Remarks

- The central limit theorem is really a property of convolution, Consider the sum of two statistically independent random variables, say, $Y = X_1 + X_2$. Then the pdf $f_Y(y)$ the convolution of $f_{X_1}(x)$ and $f_{X_2}(x)$. This can be shown with the help of the characteristic functions as follows:

$$\begin{aligned}
\phi_Y(\omega) &= E\left[e^{j\omega(x_1+x_2)}\right] \\
&= E\left(e^{j\omega x_1}\right)E\left(e^{j\omega x_2}\right) = \phi_{x_1}(\omega)\phi_{x_2}(\omega) \\
\therefore f_Y(x) &= f_{X_1}(y) * f_{X_2}(y) \\
&= \int_{-\infty}^{\infty} f_{X_1}(\tau) f_{X_2}(y-\tau) \mathrm{d}\tau
\end{aligned}$$

where * is the convolution operation.

We can illustrate this by convolving two uniform distributions repeatedly. The convolution of two uniform distributions gives a triangular distribution. Further convolution gives a parabolic distribution and so on.



(To be animated)

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 $X_1, X_2 .... X_n$ to have zero mean. Thus,

$$Y_n = (X_1 + X_2 + ... X_n)/\sqrt{n}.$$

Clearly,

$$\mu_{Y_n} = 0,$$
$$\sigma^2_{Y_n} = \sigma^2_X,$$
$$E(Y_n^3) = E(X^3)/\sqrt{n} \text{ and so on.}$$

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^{jwY_n} in power series

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

Assume all the moments of Y_n to be finite. Then

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

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

Therefore,

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

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

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

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

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

$$\therefore \lim_{n\to\infty} \text{In } \phi_{Y_n}(\omega) = e^{-\frac{\omega^2\sigma_X^2}{2}}$$

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

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

Remark:

(1) Under the conditions of the CLT, the sample mean

$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{N} X_i$ converges in distribution to $N(\mu_X, \frac{\sigma_X^2}{\sqrt{n}})$. In other words, if samples are taken from any distribution with mean μ_X and variance σ_X^2, as the sample size n increases, the distribution function of the sample mean approaches to the distribution function of a Gaussian random variable.

(2) The CLT states that the distribution function $F_{X_n}(x)$ converges to a Gaussian distribution function. The theorem does not say that the pdf $f_{Y_n}(y)$ is a Gaussian pdf in the limit. For example, suppose each X_i has a Bernoulli distribution. Then the pdf of Y consists of impulses and can never approach the Gaussian pdf.

(3) The the Cauchy distribution does not meet the conditions for the central limit theorem to hold. As we have noted earlier, this distribution does not have a finite mean or a variance.

Suppose a random variable X_i has the Cauchy distribution

$$f_{X_i}(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty.$$

The characteristic function of X_i is given by

$$\phi_{X_i}(w) = e^{-|w|}$$

The sample mean $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{N} X_i$ will have the chacteristic function

$$\phi_{Y_n}(w) = e^{-|w|}$$

Thus the sum of large number of Cauchy random variables will not follow a Gaussian distribution.

(4) The central-limit theorem one of the most widely used results of probability. If a random variable is result of several independent causes, then the random variable can be considered to be Gaussian. For example,

-the thermal noise in a resistor is result of the independent motion of billions electrons and is modelled as a Gaussian.

-the observation error/ measurement error of any process is modeled as a Gaussian.

(5) The CLT can be used to simulate a Gaussian distribution given a routine to simulate a particular random variable

Normal approximation of the Binomial distribution

One of the application of the CLT is in approximation of the Binomial coefficients.

Suppose $X_1, X_2, X_3, \ldots\ldots\ldots\ldots X_n, \ldots\ldots$ is a sequence of Bernoulli(p) random variables with P{ X_i =1}= p and P{ x_i =0}= 1- p.

Then $y_n = \sum_{i=1}^{n} X_i$ is a Binomial distribution with $\mu_{y_n} = np$ and $\sigma^2_{y_n} = n\,p\,(1-p)$.

Thus, $\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0, 1)$

or $Y_n \xrightarrow{d} N(np,\ np(1-p))$

$$\therefore \quad P\left(k-1 < Y_n \le k\right) = \int_{k-1}^{k-1} \frac{1}{\sqrt{2\pi np(1-p)}} \; e^{-\frac{1}{2}\cdot\frac{(y-np)^2}{np(1-p)}} \, .dy$$

$$\therefore \quad P\left(Y_n = k\right) = \frac{1}{\sqrt{2\pi np(1-p)}} \, e^{-\frac{1}{2}\cdot\frac{(y-np)^2}{np(1-p)}} \qquad \text{(assume the integrand interval = 1)}$$

This is normal approximation to the Binomial coefficients and known as the De-Moirre-Laplace approximation.

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, 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..

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 $\{X(s,t),\ \mathrm{s} \in \mathrm{S}, t \in \Gamma\}$ 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 ξ and index variable t and may be written as $X(t,\xi)$.

Remark

- For a fixed $t(=t_0)$, $X(t_0,\xi)$ is a random variable.
- For a fixed $\xi(=\xi_0)$, $X(t,\xi_0)$ is a single realization of the random process and is a deterministic function.
- For a fixed $\xi(=\xi_0)$ and a fixed $t(=t_0)$, $X(t,\xi_0)$ is a single number.
- When both t and ξ are varying we have the random process $X(t,\xi)$.

The random process $\{X(s,t), s \in S, t \in T\}$ is normally denoted by $\{X(t)\}$. Following figure illustrates a random procee.

A random process is illustrated below.

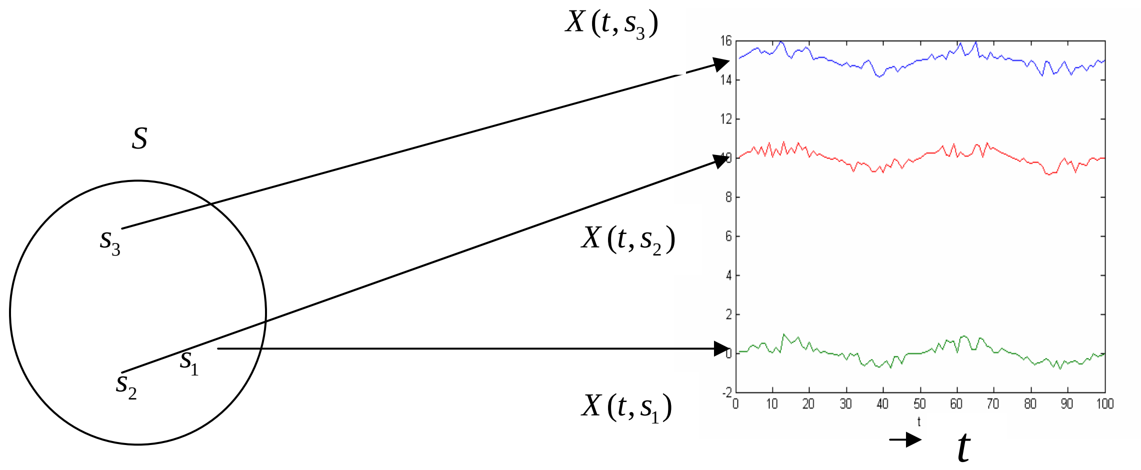


Figure   Random Process

(To Be animated)

Example 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 $p_A(-1) = 1 - p$.

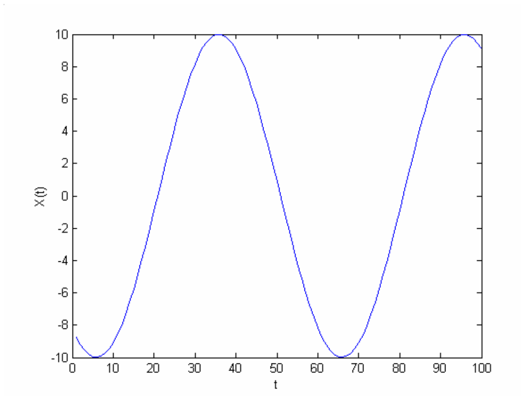
Clearly, $\{X(t),\ t \in \Gamma\}$ is a random process with two possible realizations $X_1(t) = \cos\omega t$ and $X_2(t) = -\cos\omega t$. At a particular time t_0 $X(t_0)$ is a random variable with two values $\cos\omega t_0$ and $-\cos\omega t_0$.

Continuous-time vs. discrete-time process

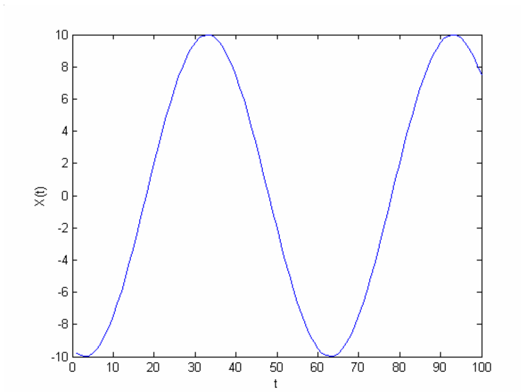
If the index set Γ is continuous, $\{X(t), t \in \Gamma\}$ is called a continuous-time process.

Example Suppose $X(t) = A\cos(w_0 t + \phi)$ where A and w_0 are constants and ϕ is uniformly distributed between 0 and 2π. $X(t)$ is an example of a continuous-time process.

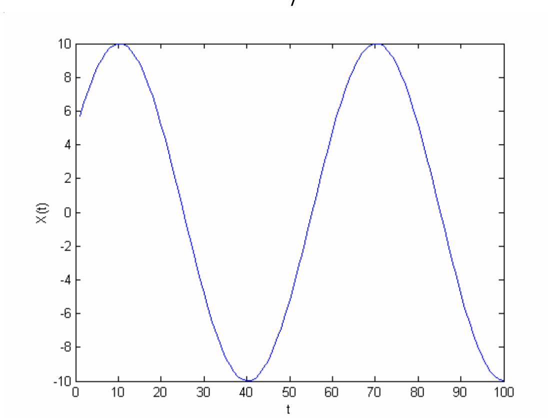
4 realizations of the process is illustrated below.
(TO BE ANIMATED)



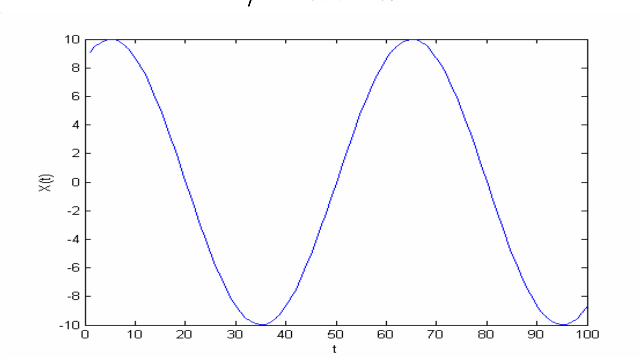
$\phi = 0.8373\pi$



$\phi = 0.9320\pi$



$\phi = 1.6924\pi$



$\phi = 1.8636\pi$

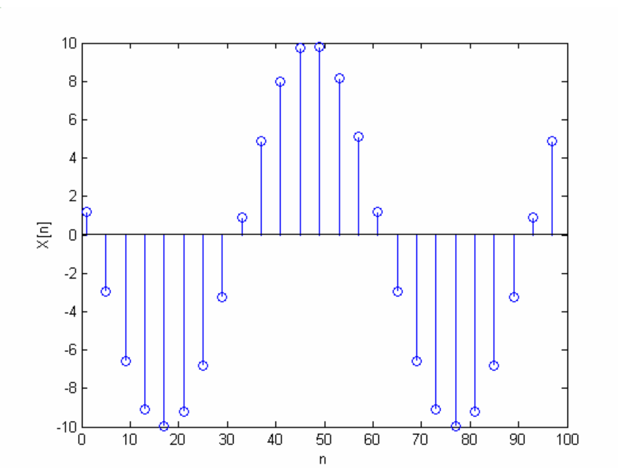
If the index set Γ is a countable set, $\{X(t), t \in \Gamma\}$ is called a discrete-time process. Such a random process can be represented as $\{X[n], n \in Z\}$ and called a *random sequence*.

Sometimes the notation $\{X_n, n \geq 0\}$ is used to describe a random sequence indexed by the set of positive integers.

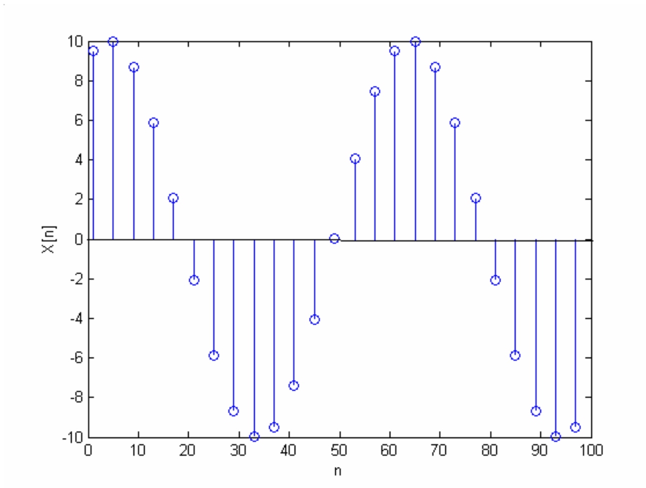
We can define a discrete-time random process on discrete points of time. Particularly, we can get a discrete-time random process $\{X[n], n \in Z\}$ by sampling a continuous-time process $\{X(t), t \in \Gamma\}$ at a uniform interval T such that $X[n] = X(nT)$.

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.

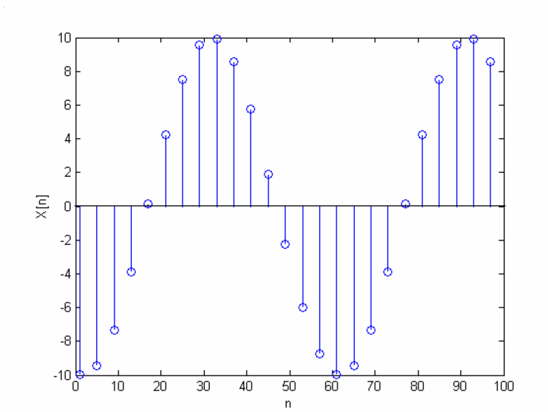
Example Suppose $X_n = \sqrt{2}\cos(\omega_0 n + Y)$ where ω_0 is a constant and Y is a random variable uniformly distributed between π and $-\pi$.
X_n is an example of a discrete-time process.



$\phi = 0.4623\pi$



$$\phi = 1.9003\pi$$



$$\phi = 0.9720\pi$$

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 countable. Other-wise the random process is called continuous state.

Example Consider the random sequence $\{X_n, n \geq 0\}$ generated by repeated tossing of a fair coin where we assign 1 to Head and 0 to Tail.

Clearly X_n can take only two values- 0 and 1. Hence $\{X_n, n \geq 0\}$ is a discrete-time *two-state* process.

How to describe a random process?

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* $f_{X(t)}(x) = \dfrac{dF_{X(t)}(x)}{dx}.$

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-th order joint distribution function

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

or th the n-th order joint density function

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

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

$$p_{X(t_1),X(t_2).....X(t_n)}(x_1, x_2.....x_n) = P(X(t_1) = x_1, X(t_2) = x_2.....X(t_n) = x_n), \ \forall\ n \geq 1 \text{ and } \forall t_n \in \Gamma$$

If the random process is continuous-state, it can be specified by

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

$R_X(t_1,t_2) = R_X(t_2,t_1,)$ and
$R_X(t,t) = EX^2(t) =$ *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

$$\begin{aligned} C_X(t_1,t_2) &= E(X(t_1)-\mu_X(t_1))(X(t_2)-\mu_X(t_2)) \\ &= R_X(t_1,t_2)-\mu_X(t_1)\mu_X(t_2) \end{aligned}$$

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

These moments give partial information about the process.

The ratio $\rho_X(t_1,t_2) = \dfrac{C_X(t_1,t_2)}{\sqrt{C_X(t_1,t_1)\,C_X(t_2,t_2)}}$ 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

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

The above definitions are easily extended to a random sequence $\{X_n, n \geq 0\}$.

Example

(a) Gaussian Random Process

For any positive integer n, $X(t_1), X(t_2),.....X(t_n)$ represent n jointly random variables. These n random variables define a random vector $\mathbf{X} = [X(t_1), X(t_2),.....X(t_n)]'$. The process $X(t)$ is called Gaussian if the random vector $[X(t_1), X(t_2),.....X(t_n)]'$ is jointly Gaussian with the joint density function given by

$$f_{X(t_1),X(t_2)...X(t_n)}(x_1,x_2,...x_n) = \frac{e^{-\frac{1}{2}\mathbf{X}'\mathbf{C}_X^{-1}\mathbf{X}}}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(\mathbf{C}_\mathbf{X})}}$$

where $\mathbf{C}_\mathbf{X} = E(\mathbf{X}-\boldsymbol{\mu}_\mathbf{X})(\mathbf{X}-\boldsymbol{\mu}_\mathbf{X})'$

and $\boldsymbol{\mu}_\mathbf{X} = E(\mathbf{X}) = \left[E(X_1), E(X_2)......E(X_n)\right]'$.

The Gaussian Random Process is completely specified by the autocovariance matrix and hence by the mean vector and the autocorrelation matrix $\mathbf{R_X} = E\mathbf{XX}'$.

(b) Bernoulli Random Process

A **Bernoulli process** is a discrete-time random process consisting of a sequence of independent and identically distributed **Bernoulli random variables**. Thus the discrete – time random process $\{X_n, n \geq 0\}$ is **Bernoulli process if**

$P\{X_n = 1\} = p$ and

$P\{X_n = 0\} = 1 - p$

Example

Consider the random sequence $\{X_n, n \geq 0\}$ generated by repeated tossing of a fair coin where we assign 1 to Head and 0 to Tail. Here $\{X_n, n \geq 0\}$ is a Bernoulli process where each random variable X_n is a **Bernoulli random variable** with

$p_X(1) = P\{X_n = 1\} = \frac{1}{2}$ and

$p_X(0) = P\{X_n = 0\} = \frac{1}{2}$

(c) A sinusoid with a random phase

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

$f_\Phi(\phi) = \frac{1}{2\pi}$

$X(t)$ at a particular t is a random variable and it can be shown that

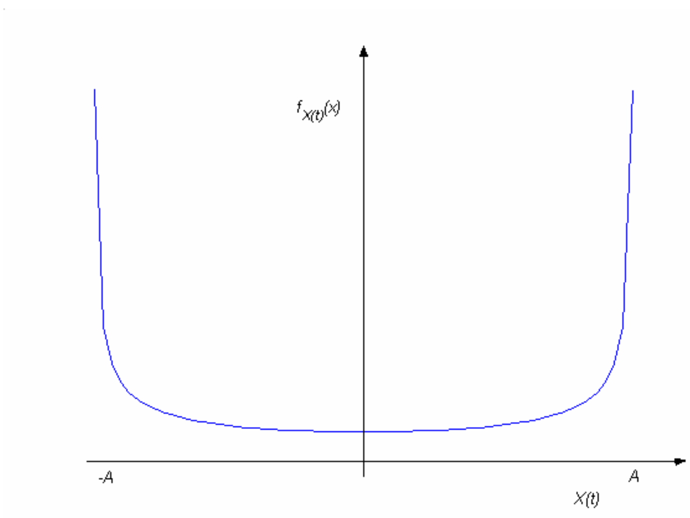
$$f_{X(t)}(x) = \begin{cases} \dfrac{1}{\pi\sqrt{A^2 - x^2}} & |x| < A \\ 0 & \text{otherwise} \end{cases}$$

The pdf is sketched in the Fig. below:

The mean and autocorrelation of $X(t)$:

$$\mu_{X(t)} = EX(t)$$
$$= EA\cos(w_0 t + \phi)$$
$$= \int_{-\infty}^{\infty} A\cos(w_0 t + \phi)\frac{1}{2\pi}d\phi$$
$$= 0$$

$$R_X(t_1, t_2) = EA\cos(w_0 t_1 + \phi)A\cos(w_0 t_2 + \phi)$$
$$= A^2 E\cos(w_0 t_1 + \phi)\cos(w_0 t_2 + \phi)$$
$$= \frac{A^2}{2}E(\cos(w_0(t_1 - t_2)) + \cos(w_0(t_1 + t_2 + 2\phi)))$$
$$= \frac{A^2}{2}\cos(w_0(t_1 - t_2)) + \frac{A^2}{2}\int_{-\pi}^{\pi}\cos(w_0(t_1 + t_2 + 2\phi))\frac{1}{2\pi}d\phi$$
$$= \frac{A^2}{2}\cos(w_0(t_1 - t_2))$$



Two or More Random Processes

In practical situations we deal with two or more random processes. We often deal with the input and output processes of a system. To describe two or more random processes we have to use the *joint distribution functions* and the *joint moments.*

Consider two random processes $\{X(t), t \in \Gamma\}$ and $\{Y(t), t \in \Gamma\}$. For any positive integer n, $X(t_1), X(t_2), \ldots\ldots X(t_n), Y(t_1^{/}), Y(t_2^{/}), \ldots\ldots Y(t_n^{/})$ represent $2n$ jointly distributed random variables. Thus these two random processes can be described by the $(n+m)$th order joint distribution function

$$F_{X(t_1),X(t_2).....X(t_n),Y(t_1'),Y(t_2'),.....Y(t_m')}(x_1,x_2.....x_n,y_1,y_2.....y_m)$$

$$=P(X(t_1)\le x_1,X(t_2)\le x_2.....X(t_n)\le x_n,Y(t_1')\le y_1,Y(t_2')\le y_2.....Y(t_m')\le y_m)$$

or the corresponding $(n+m)$th order joint density function

$$f_{X(t_1),X(t_2).....X(t_n),Y(t_1'),Y(t_2'),.....Y(t_n')}(x_1,x_2.....x_n,y_1,y_2.....y_n)$$

$$=\frac{\partial^{2n}}{\partial x_1\partial x_2...\partial x_n\partial y_1\partial y_2...\partial y_n}F_{X(t_1),X(t_2).....X(t_n),Y(t_1'),Y(t_2'),.....Y(t_n')}(x_1,x_2.....x_n,y_1,y_2.....y_n)$$

Two random processes can be partially described by the joint moments:

- $\textit{Cross}-\textit{correlation function}$ of the processes at times t_1,t_2
  $$R_{XY}(t_1,t_2)=E(X(t_1)Y(t_2))=E(X(t_1)Y(t_2))$$

- $\textit{Cross}-\text{cov}\,\textit{ariance function}$ of the processes at times t_1,t_2
  $$C_{XY}(t_1,t_2)==E(X(t_1)-\mu_X(t_1))(X(t_2)-\mu_X(t_2))$$
  $$=R_{XY}(t_1,t_2)-\mu_X(t_1)\mu_Y(t_2)$$
  .

- 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

- Independent processes: Two *random processes* $\{X(t),t\in\Gamma\}$ *and* $\{Y(t),t\in\Gamma\}$.

are called independent if

- Uncorrelated processes: Two *random processes* $\{X(t),t\in\Gamma\}$ *and* $\{Y(t),t\in\Gamma\}$.

are called uncorrelated if

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

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

$$\mathrm{R}_{XY}(t_1, t_2) = 0 \quad \forall\ t_1, t_2 \in \Gamma$$

Example Suppose $X(t) = A\cos(w_0 t + \phi)$ and $Y(t) = A\sin(w_0 t + \phi)$ where A and w_0 are constants and ϕ is uniformly distributed between 0 and 2π.

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, 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..

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 $\{X(s,t),\ \mathrm{s} \in \mathrm{S}, t \in \Gamma\}$ 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 ξ and index variable t and may be written as $X(t,\xi)$.

Remark

- For a fixed $t(=t_0)$, $X(t_0,\xi)$ is a random variable.
- For a fixed $\xi(=\xi_0)$, $X(t,\xi_0)$ is a single realization of the random process and is a deterministic function.
- For a fixed $\xi(=\xi_0)$ and a fixed $t(=t_0)$, $X(t,\xi_0)$ is a single number.
- When both t and ξ are varying we have the random process $X(t,\xi)$.

The random process $\{X(s,t), s \in S, t \in T\}$ is normally denoted by $\{X(t)\}$. Following figure illustrates a random procee.

A random process is illustrated below.

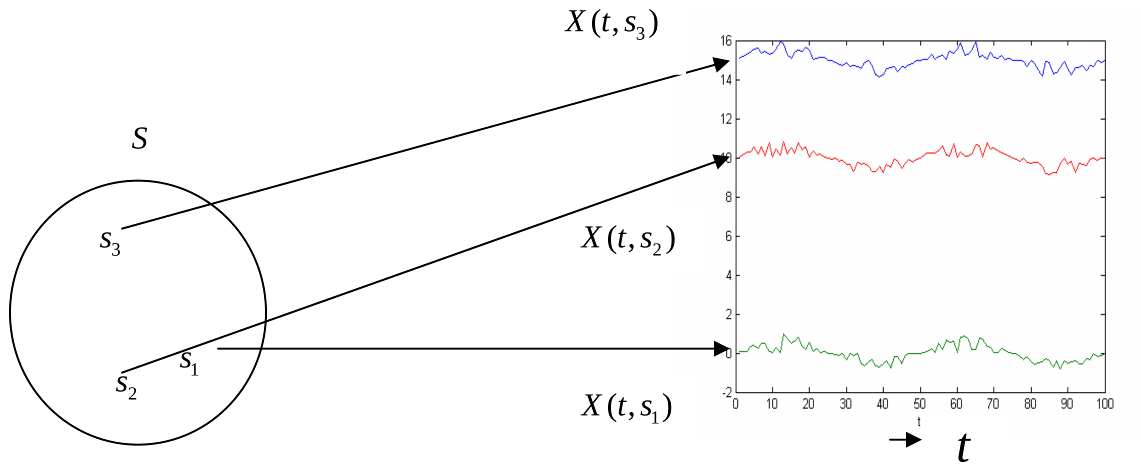


Figure   Random Process

(To Be animated)

Example 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 $p_A(-1) = 1 - p$.

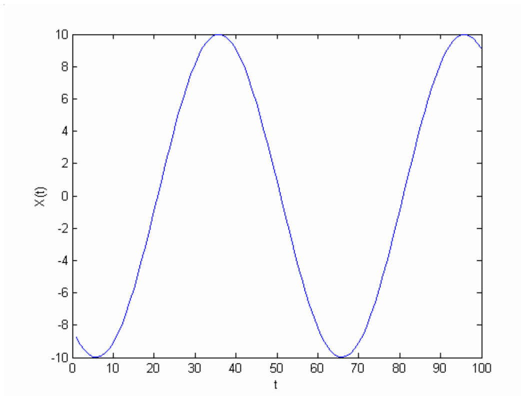
Clearly, $\{X(t),\ t \in \Gamma\}$ is a random process with two possible realizations $X_1(t) = \cos \omega t$ and $X_2(t) = -\cos \omega t$. At a particular time t_0 $X(t_0)$ is a random variable with two values $\cos \omega t_0$ and $-\cos \omega t_0$.

Continuous-time vs. discrete-time process

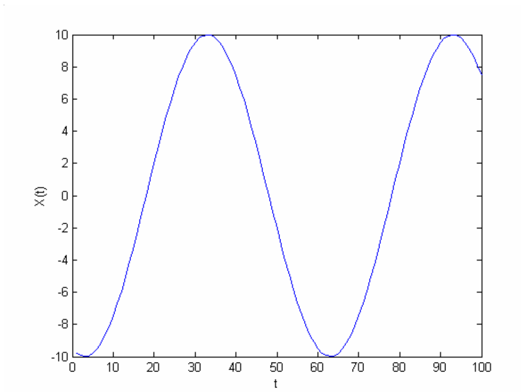
If the index set Γ is continuous, $\{X(t), t \in \Gamma\}$ is called a continuous-time process.

Example Suppose $X(t) = A\cos(w_0 t + \phi)$ where A and w_0 are constants and ϕ is uniformly distributed between 0 and 2π. $X(t)$ is an example of a continuous-time process.

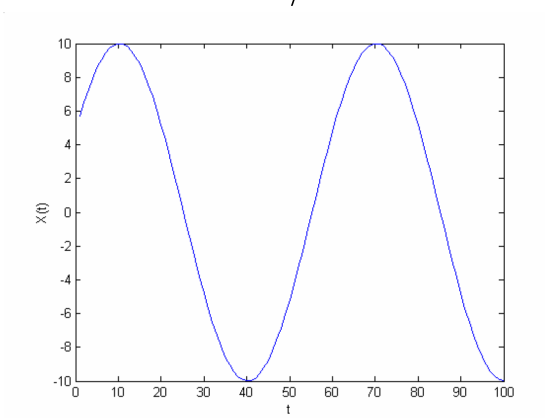
4 realizations of the process is illustrated below.
(TO BE ANIMATED)



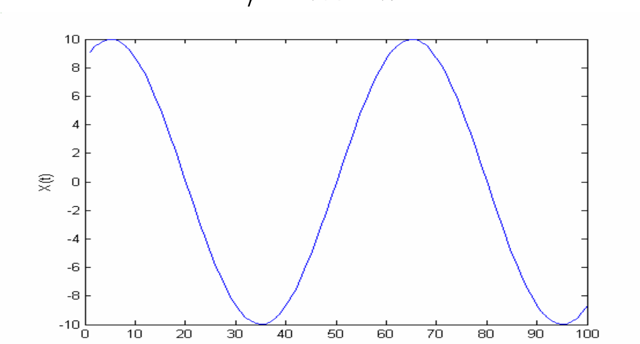
$\phi = 0.8373\pi$



$\phi = 0.9320\pi$



$\phi = 1.6924\pi$



$\phi = 1.8636\pi$

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*.

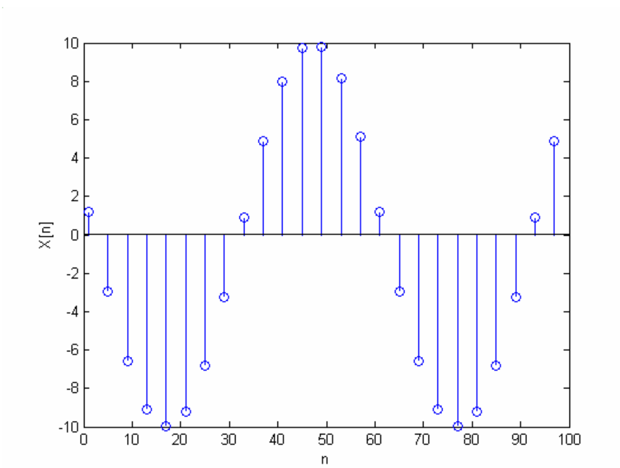
Sometimes the notation $\{X_n, n \geq 0\}$ is used to describe a random sequence indexed by the set of positive integers.

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 $X[n] = X(nT)$.

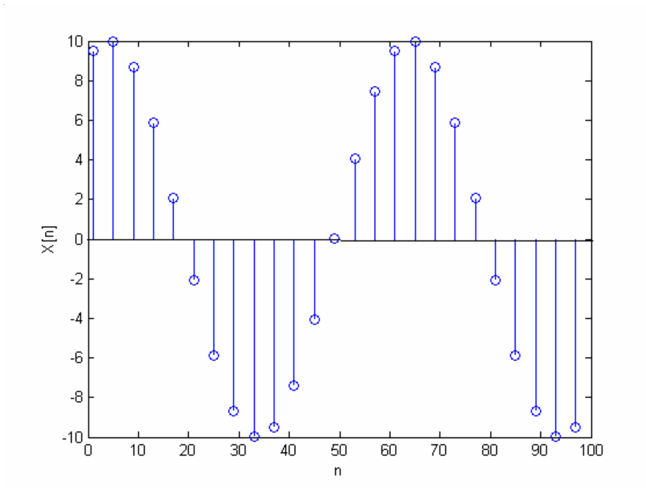
The discrete-time random process is more important in practical implementations. Advanced statistical signal processing techniques have been developed to process this type of signals.

Example Suppose $X_n = \sqrt{2}\cos(\omega_0 n + Y)$ where ω_0 is a constant and Y is a random variable uniformly distributed between π and $-\pi$.

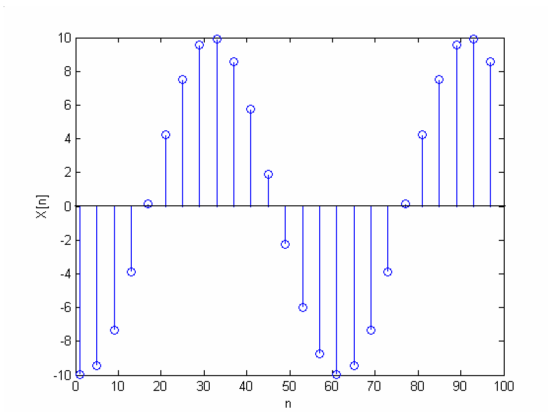
X_n is an example of a discrete-time process.



$\phi = 0.4623\pi$



$\phi = 1.9003\pi$



$\phi = 0.9720\pi$

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 countable. Other-wise the random process is called continuous state.

Example Consider the random sequence $\{X_n, n \geq 0\}$ generated by repeated tossing of a fair coin where we assign 1 to Head and 0 to Tail.

Clearly X_n can take only two values- 0 and 1. Hence $\{X_n, n \geq 0\}$ is a discrete-time *two-state* process.

How to describe a random process?

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* $f_{X(t)}(x) = \dfrac{dF_{X(t)}(x)}{dx}.$

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, $X(t_1), X(t_2), ...X(t_n)$ represents n jointly distributed random variables. Thus a random process $\{X(t), t \in \Gamma\}$ can thus be described by specifying the n-th order joint distribution function

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

or th the n-th order joint density function

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

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

$$p_{X(t_1), X(t_2).....X(t_n)}(x_1, x_2.....x_n) = P(X(t_1) = x_1, X(t_2) = x_2.....X(t_n) = x_n), \ \ \forall \ n \geq 1 \text{ and } \ \forall t_n \in \Gamma$$

If the random process is continuous-state, it can be specified by

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

$R_X(t_1,t_2) = R_X(t_2,t_1,)$ and

$R_X(t,t) = EX^2(t) =$ *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

$$\begin{aligned} C_X(t_1,t_2) &= E(X(t_1)-\mu_X(t_1))(X(t_2)-\mu_X(t_2)) \\ &= R_X(t_1,t_2)-\mu_X(t_1)\mu_X(t_2) \end{aligned}$$

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

These moments give partial information about the process.

The ratio $\rho_X(t_1,t_2) = \dfrac{C_X(t_1,t_2)}{\sqrt{C_X(t_1,t_1)\,C_X(t_2,t_2)}}$ 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

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

The above definitions are easily extended to a random sequence $\{X_n, n \geq 0\}$.

Example

(a) Gaussian Random Process

For any positive integer n, $X(t_1), X(t_2), ...X(t_n)$ represent n jointly random variables. These n random variables define a random vector $\mathbf{X} = [X(t_1), X(t_2), ...X(t_n)]'$. The process $X(t)$ is called Gaussian if the random vector $[X(t_1), X(t_2), ...X(t_n)]'$ is jointly Gaussian with the joint density function given by

$$f_{X(t_1),X(t_2)...X(t_n)}(x_1,x_2,...x_n) = \frac{e^{-\frac{1}{2}\mathbf{X}'\mathbf{C}_X^{-1}\mathbf{X}}}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(\mathbf{C_X})}}$$

where $\mathbf{C_X} = E(\mathbf{X}-\boldsymbol{\mu}_\mathbf{X})(\mathbf{X}-\boldsymbol{\mu}_\mathbf{X})'$

and $\boldsymbol{\mu}_\mathbf{X} = E(\mathbf{X}) = \left[E(X_1), E(X_2)......E(X_n)\right]'$.

The Gaussian Random Process is completely specified by the autocovariance matrix and hence by the mean vector and the autocorrelation matrix $\mathbf{R_X} = E\mathbf{XX}'$.

(b) Bernoulli Random Process

A **Bernoulli process** is a discrete-time random process consisting of a sequence of independent and identically distributed **Bernoulli random variables**. Thus the discrete – time random process $\{X_n, n \geq 0\}$ is **Bernoulli process if**

$P\{X_n = 1\} = p$ and

$P\{X_n = 0\} = 1 - p$

Example

Consider the random sequence $\{X_n, n \geq 0\}$ generated by repeated tossing of a fair coin where we assign 1 to Head and 0 to Tail. Here $\{X_n, n \geq 0\}$ is a Bernoulli process where each random variable X_n is a **Bernoulli random variable** with

$p_X(1) = P\{X_n = 1\} = \frac{1}{2}$ and

$p_X(0) = P\{X_n = 0\} = \frac{1}{2}$

(c) A sinusoid with a random phase

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

$f_\Phi(\phi) = \frac{1}{2\pi}$

$X(t)$ at a particular t is a random variable and it can be shown that

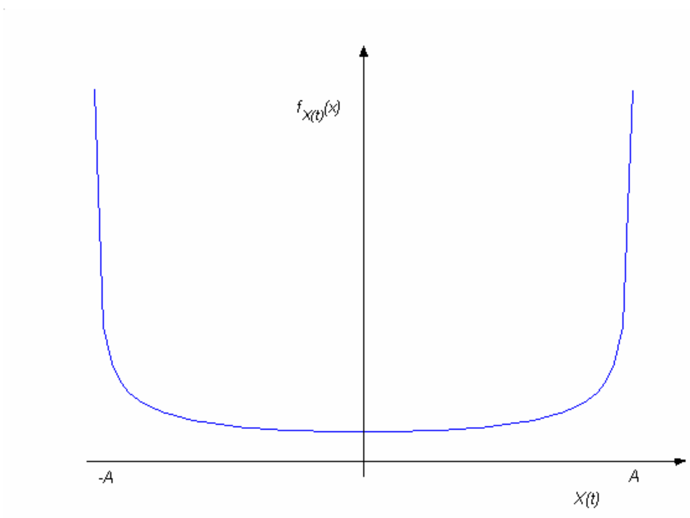
$$f_{X(t)}(x) = \begin{cases} \dfrac{1}{\pi\sqrt{A^2 - x^2}} & |x| < A \\ 0 & \text{otherwise} \end{cases}$$

The pdf is sketched in the Fig. below:

The mean and autocorrelation of $X(t)$:

$$
\begin{aligned}
\mu_{X(t)} &= EX(t) \\
&= EA\cos(w_0 t + \phi) \\
&= \int_{-\infty}^{\infty} A\cos(w_0 t + \phi)\frac{1}{2\pi}d\phi \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
R_X(t_1, t_2) &= EA\cos(w_0 t_1 + \phi)A\cos(w_0 t_2 + \phi) \\
&= A^2 E\cos(w_0 t_1 + \phi)\cos(w_0 t_2 + \phi) \\
&= \frac{A^2}{2} E(\cos(w_0(t_1 - t_2)) + \cos(w_0(t_1 + t_2 + 2\phi))) \\
&= \frac{A^2}{2}\cos(w_0(t_1 - t_2)) + \frac{A^2}{2}\int_{-\pi}^{\pi}\cos(w_0(t_1 + t_2 + 2\phi))\frac{1}{2\pi}d\phi \\
&= \frac{A^2}{2}\cos(w_0(t_1 - t_2))
\end{aligned}
$$



Two or More Random Processes

In practical situations we deal with two or more random processes. We often deal with the input and output processes of a system. To describe two or more random processes we have to use the *joint distribution functions* and the *joint moments.*
Consider two random processes $\{X(t), t \in \Gamma\}$ and $\{Y(t), t \in \Gamma\}$. For any positive integer n, $X(t_1), X(t_2), ...X(t_n), Y(t_1'), Y(t_2'), ...Y(t_n')$ represent $2n$ jointly distributed random variables. Thus these two random processes can be described by the $(n+m)$th order joint distribution function

$$F_{X(t_1),X(t_2).....X(t_n),Y(t_1'),Y(t_2'),.....Y(t_m')}(x_1,x_2.....x_n,y_1,y_2.....y_m)$$
$$=P(X(t_1)\le x_1,X(t_2)\le x_2.....X(t_n)\le x_n,Y(t_1')\le y_1,Y(t_2')\le y_2.....Y(t_m')\le y_m)$$

or the corresponding $(n+m)$th order joint density function

$$f_{X(t_1),X(t_2).....X(t_n),Y(t_1'),Y(t_2'),.....Y(t_n')}(x_1,x_2.....x_n,y_1,y_2.....y_n)$$
$$=\frac{\partial^{2n}}{\partial x_1\partial x_2...\partial x_n\partial y_1\partial y_2...\partial y_n}F_{X(t_1),X(t_2).....X(t_n),Y(t_1'),Y(t_2'),.....Y(t_n')}(x_1,x_2.....x_{n,}y_1,y_2.....y_n)$$

Two random processes can be partially described by the joint moments:

- $Cross-correlation\ function$ of the processes at times t_1,t_2
  $R_{XY}(t_1,t_2)\ =E(X(t_1)Y(t_2))\ =E(X(t_1)Y(t_2))$

- $Cross-\mathrm{cov}ariance\ function$ of the processes at times t_1,t_2
  $$C_{XY}(t_1,t_2)\ ==E(X(t_1)-\mu_X(t_1))(X(t_2)-\mu_X(t_2))$$
  $$=R_{XY}(t_1,t_2)-\mu_X(t_1)\mu_Y(t_2)$$
  .

- 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

- Independent processes: Two *random processes* $\{X(t),t\in\Gamma\}$ *and* $\{Y(t),t\in\Gamma\}$.

are called independent if

- Uncorrelated processes: Two *random processes* $\{X(t),t\in\Gamma\}$ *and* $\{Y(t),t\in\Gamma\}$.

are called uncorrelated if

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

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

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

Example Suppose $X(t) = A\cos(w_0 t + \phi)$ and $Y(t) = A\sin(w_0 t + \phi)$ where A and w_0 are constants and ϕ is uniformly distributed between 0 and 2π.

Important Classes of Random Processes

Having characterized the random process by the joint distribution (density) functions and joint moments we define the following two important classes of random processes.

(a) Independent and Identically Distributed Process

Consider a discrete-time random process $\{X_n\}$. For *any finite choice* of time instants $n_1, n_2, ..., n_N$, if the random variables $X_{n_1}, X_{n_2}, ..., X_{n_N}$ are jointly independent with a common distribution, then $\{X_n\}$ is called an independent and identically distributed (iid) random process. Thus for an iid random process $\{X_n\}$,

$$F_{X_{n_1}, X_{n_2}, ..., X_{n_N}}(x_1, x_2.....x_n) = F_X(x_1)F_X(x_2).....F_X(x_n)$$

and equivalently

$$p_{X_{n_1}, X_{n_2}, ..., X_{n_N}}(x_1, x_2.....x_n) = p_X(x_1)p_X(x_2).....p_X(x_n)$$

Moments of the IID process:

It is easy to verify that for an iid process $\{X_n\}$

- *Mean* $EX_n = \mu_X$ = constant
- *Variance* $E(X_n - \mu_X)^2 = \sigma_X^2$ =constant
- *Autocovariance* $C_X(n,m) = E(X_n - \mu_X)(X_m - \mu_X)$

$$= E(X_n - \mu_X)E(X_m - \mu_X)$$

$$= \begin{cases} 0 & \text{for } n \neq m \\ \sigma_X^2 & \text{otherwise} \end{cases}$$

$$= \sigma_X^2 \delta[n,m]$$

where $\delta[n,m] = 1$ for $n = m$ and 0 otherwise.

- *Autocorrelation* $R_X(n,m) = C_X(n,m) + \mu_X^2 = \sigma_X^2 \delta(n,m) + \mu_X^2$

Example Bernoulli process: Consider the Bernoulli process $\{X_n\}$ with

$p_X(1) = p$ and

$p_X(0) = 1 - p$

This process is an iid process.

Using the iid property, we can obtain the joint probability mass functions of any order in terms of p. For example,

$$p_{X_1, X_2}(1,0) = p(1-p)$$

$$p_{X_1, X_2, X_3}(0,0,1) = (1-p)^2 p$$

and so on.

Similarly, the mean, the variance and the autocorrelation function are given by

$$\mu_{X_n} = EX_n = p$$
$$\operatorname{var}(X_n) = p(1-p)$$
$$\begin{aligned} R_X(n_1, n_2) &= EX_{n_1} X_{n_2} \\ &= EX_{n_1} EX_{n_2} \\ &= p^2 \end{aligned}$$

(b) Independent Increment Process

A random process $\{X(t)\}$ is called an in*dependent increment* process if for any $n > 1$ and $t_1 < t_2 < \ldots < t_n \in \Gamma,$ the set of n random variables $X(t_1), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are jointly independent random variables.

If the probability distribution of $X(t+r) - X(t'+r)$ is same as that of $X(t) - X(t'),$ for any choice of $t, t'$and $r,$ the $\{X(t)\}$ is called s*tationary increment process.*

- The above definitions of the in*dependent increment* process and the s*tationary increment process* can be easily extended to discrete-time random processes.
- The independent increment property simplifies the calculation of joint probability distribution, density and mass functions from the corresponding first-order quantities. As an example,for $t_1 < t_2,\ x_1 < x_2,$

$$\begin{aligned} F_{X(t_1),X(t_2)}(x_1, x_2) &= P(\{X(t_1) \le x_1, X(t_2) \le x_2\} \\ &= P(\{X(t_1) \le x_1\}) P(\{X(t_2) \le x_2\} / \{X(t_1) \le x_1\}) \\ &= P(\{X(t_1) \le x_1) P(\{X(t_2) - X(t_1) \le x_2 - x_1\}) \\ &= F_{X(t_1)}(x_1) F_{X(t_2)-X(t_1)}(x_2 - x_1) \end{aligned}$$

- The independent increment property simplifies the computation of the autocovariance function.

For $t_1 < t_2,$ the autocorrelation function of $X(t)$ is given by

$$\begin{aligned} R_X(t_1, t_2) &= EX(t_1) X(t_2) \\ &= EX(t_1)(X(t_1) + X(t_2) - X(t_1)) \\ &= EX^2(t_1) + EX(t_1) E(X(t_2) - X(t_1)) \\ &= EX^2(t_1) + EX(t_1) EX(t_2) - (EX(t_1))^2 \\ &= \operatorname{var}(X(t_1)) + EX(t_1) EX(t_2) \end{aligned}$$
$$\therefore\ C_X(t_1, t_2) = EX(t_1) X(t_2) - EX(t_1) EX(t_2) = \operatorname{var}(X(t_1))$$

Similarly, for $t_1 > t_2,$

$$C_X(t_1, t_2) = \operatorname{var}(X(t_2))$$

Therefore

$$\boxed{C_X(t_1, t_2) = \operatorname{var}(X(\min(t_1, t_2)))}$$

Example: Two continuous-time independent increment processes are widely studied. They are:

(a) Wiener process with the increments following Gaussian distribution and

(b) Poisson process with the increments following Poisson distribution. We shall discuss these processes shortly.

Random Walk process

Consider an iid process $\{Z_n\}$ having two states $Z_n = 1$ $Z_n = -1$ with the probability mass functions

$p_Z(1) = p$ and $p_Z(-1) = q = 1 - p.$

Then the *sum process* $\{X_n\}$ given by

$$X_n = \sum_{i=1}^{n} Z_i = X_{n-1} + Z_n$$

with $X_0 = 0$ is called a Random Walk process.

- This process is one of the widely studied random processes.
- It is an independent increment process. This follows from the fact that $X_n - X_{n-1} = Z_n$ and $\{Z_n\}$ is an iid process.
- If we call $Z_n = 1$ as *success* and $Z_n = -1$ as *failure***,** then $X_n = \sum_{i=1}^{n} Z_i$ represents the total number of successes in n independent trials.
- If $p = \frac{1}{2}$, $\{X_n\}$ is called a *symmetrical random walk* process.

Probability mass function of the Random Walk Process

At an instant n, X_n can take integer values from $-n$ to n

Suppose $X_n = k.$

Clearly $k = n_1 - n_{-1}$

where $n_1 =$ number of *successes* and $n_{-1} =$ number of *failures* in n trials of Z_n such that $n_1 + n_{-1} = n.$

$\therefore\ n_1 = \dfrac{n+k}{2}$ and $n_{-1} = \dfrac{n-k}{2}$

Also n_1 and n_{-1} are necessarily non-negative integers.

$$\therefore\ p_{X_n}(k) = \begin{cases} {}^nC_{\frac{n+k}{2}}\, p^{\frac{n+k}{2}} (1-p)^{\frac{n-k}{2}} & \text{if } \dfrac{n+k}{2} \text{ and } \dfrac{n-k}{2} \text{ are non-negative integers} \\ 0 & \text{otherwise} \end{cases}$$

Mean, Variance and Covariance of a Random Walk process

Note that

$$EZ_n = 1\times p - 1\times(1-p) = 2p-1$$

$$EZ_n^2 = 1\times p + 1\times(1-p) = 1$$

and

$$\begin{aligned}\operatorname{var}(Z_n) &= EZ_n^2 - (EZ_n)^2\\ &= 1 - 4p^2 + 4p - 1\\ &= 4pq\end{aligned}$$

$$\therefore EX_n = \sum_{i=1}^{n} EZ_i = n(2p-1)$$

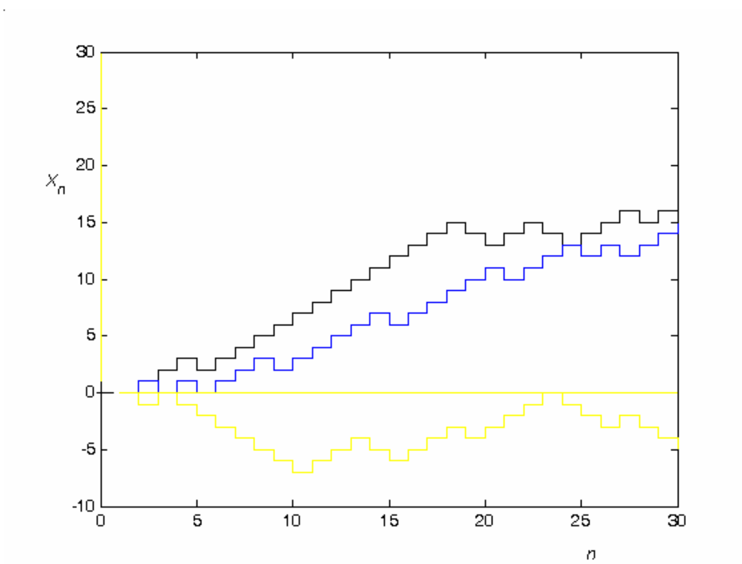
and

$$\begin{aligned}\operatorname{var}(X_n) &= \sum_{i=1}^{n}\operatorname{var}(Z_i) \qquad \because Z_i s \text{ are independent random variables}\\ &= 4npq\end{aligned}$$

Since the random walk process $\{X_n\}$ is an independent increment process, the autocovariance function is given by

$C_X(n_1, n_2) = 4pq\min(n_1, n_2)$

Three realizations of a random walk process is as shown in the Fig. below:



Remark If the increment Z_n of the random walk process takes the values of s and $-s$, then

$$\therefore EX_n = \sum_{i=1}^{n} EZ_i = n(2p-1)s$$

and

$$\begin{aligned}\operatorname{var}(X_n) &= \sum_{i=1}^{n} \operatorname{var}(Z_i)\\ &= 4npqs^2\end{aligned}$$

(c) Markov process

A process $\{X(t)\}$ is called a Markov process if for any sequence of time $t_1 < t_2 < \ldots\ldots < t_n$,

$$P(\{X(t_n) \leq x \mid X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_{n-1}) = x_{n-1}\}) = P(\{X(t_n) \leq x \mid X(t_{n-1}) = x_{n-1}\})$$

- Thus for a Markov process *"the future of the process, given present, is independent of the past."*
- A discrete-state Markov process is called *a Markov Chain*. If $\{X_n\}$ is a discrete-time discrete-state random process, the process is Markov if
$$P(\{X_n = x_n \mid X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}\}) = P(\{X_n = x_n \mid X_{n-1} = x_{n-1}\})$$
- An iid random process is a Markov process.
- Many practical signals with strong correlation between neighbouring samples are modelled as Markov processes

Example Show that the random walk process $\{X_n\}$ is Markov.

Here,

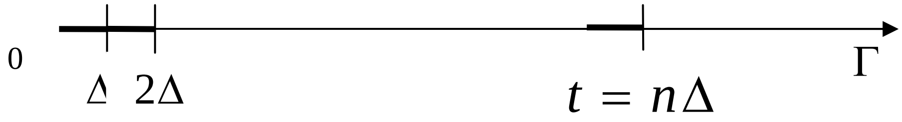
$$\begin{aligned}P(\{X_n = x_n \mid X_0 = 0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}\})\\ &= P(\{X_{n-1} + Z_n = x_n \mid X_0 = 0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}\})\\ &= P(\{Z_n = x_n - x_{n-1}\})\\ &= P(\{X_n = x_n \mid X_{n-1} = x_{n-1}\})\end{aligned}$$

Wiener Process

Consider a symmetrical random walk process $\{X_n\}$ given by

$X_n = X(n\Delta)$

where the discrete instants in the time axis are separated by Δ as shown in the Fig. below. Assume Δ to be infinitesimally small.



Clearly,

$EX_n = 0$

$\operatorname{var}(X_n) = 4pqns^2 = 4 \times \frac{1}{2} \times \frac{1}{2} ns^2 = ns^2$

For large $n,$ the distribution of X_n approaches the normal with men 0 and variance

$ns^2 = \frac{t}{\Delta} s^2 = \alpha t$

As $\Delta \to 0$ and $n \to \infty,$ X_n becomes the continuous-time process $X(t)$ with the pdf

$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-\frac{1}{2}\frac{x^2}{\alpha t}}$. This process $\{X(t)\}$ is called the Wienerprocess.

A random process $\{X(t)\}$ is called a Wiener process or the Brownian motion process if it satisfies the following conditions:

(1) $X(0) = 0$

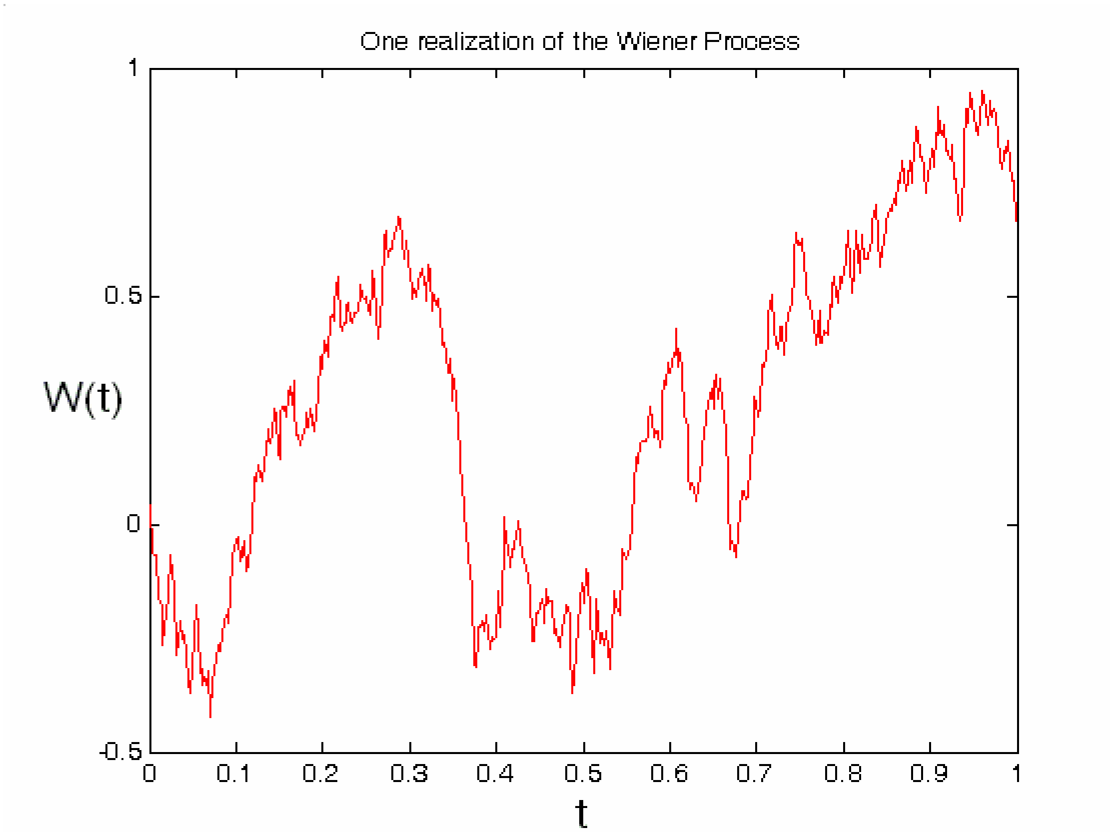
(2) $X(t)$ is an independent increment process.

(3) For each $s \geq 0, t \geq 0$ $X(s+t) - X(s)$ has the normal distribution with mean 0 and variance αt.

$$f_{X(s+t)-X(s)}(x) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-\frac{1}{2}\frac{x^2}{\alpha t}}$$

- Wiener process was used to model the Brownian motion – microscopic particles suspended in a fluid are subject to continuous molecular impacts resulting in the zigzag motion of the particle named Brownian motion after the British Botanist Brown.
- Wiener Process is the integration of the white noise process.

A realization of the Wiener process is shown in the figure below:



$$
\begin{aligned}
R_X(t_1,t_2) &= EX(t_1)X(t_2) \\
&= EX(t_1)\{X(t_2)-X(t_1)+X(t_1)\} \qquad \text{Assuming } t_2 > t_1 \\
&= EX(t_1)E\{X(t_2)-X(t_1)\}+EX^2(t_1) \\
&= EX^2(t_1) \\
&= \alpha t_1
\end{aligned}
$$

Similarly if $t_1 > t_2$

$$
\begin{aligned}
&R_X(t_1,t_2) = \alpha t_2 \\
&\therefore R_X(t_1,t_2) = \alpha \min(t_1,t_2)
\end{aligned}
$$

$$
f_{X(t)}(x) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-\frac{1}{2\alpha}\frac{x^2}{t}}
$$

Remark

$C_X(t_1,t_2) = \alpha \min(t_1,t_2)$

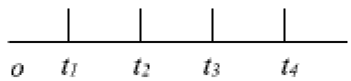
$X(t)$ is a Gaussian process.

Poisson Process

Consider a random process representing the number of occurrences of an event up to time *t* (over time interval $(0,\ t]$. Such a process is called a *counting process* and we shall denote it by $\{N(t), t \geq 0\}$. Clearly $\{N(t), t \geq 0\}$ is a continuous-time discrete state process and any of its realizations is non-decreasing function of time.

The counting process $\{N(t), t \geq 0\}$ is called Poisson's process with the rate parameter λ if

(i) $N(0)=0$
(ii) $N(t)$ is an independent increment process.



Thus the increments $N(t_2) - N(t_1)$ and $N(t_4) - N(t_3), et.$ are independent.

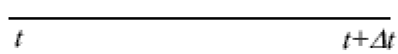
(iii) $P\left(\{N(\Delta t) = 1\}\right) = \lambda \Delta t + 0(\Delta t)$

where $0(\Delta t)$ implies any function such that $\lim\limits_{\Delta t \to 0} \dfrac{0(\Delta t)}{\Delta t} = 0.$

(iv) $P\left(\{N(\Delta t) \geq 2\}\right) = 0\left(\Delta t\right)$

The assumptions are valid for many applications. Some typical examples are

- Number of alpha particles emitted by a radio active substance.
- Number of binary packets received at switching node of a communication network.
- Number of cars arriving at a petrol pump during a particular interval of time.



$P\left(\{N(t+\Delta t) = n\}\right)$ = Probability of occurrence of n events up to time $t + \Delta t$

$$= P\left(\{N(t) = n, N(\Delta t) = 0\}\right) + P\left(\{N(t) = n-1, n(\Delta t) = 1\}\right) + P\left(\{N(t) < n-1, n(\Delta t) \geq 2\}\right)$$

$$= P\left(\{N(t) = n\}\right)(1 - \lambda \Delta t - 0(\Delta t)) + P\left(\{N(t) = n-1\}\right)(\lambda \Delta t + 0(\Delta t)) + P\left(\{N(t) = (n-1)\}\right)(0$$

$$P\left(\{N(t) = n, N(\Delta t) = 0\}\right) + P\left(\{N(t) = n-1, n(\Delta t) = 1\}\right) + P\left(\{N(t) < n-1, n(\Delta t) \geq 2\}\right)$$

$$= P\left(\{N(t) = n\}\right)(1 - \lambda \Delta t - 0(\Delta t)) + P\left(\{N(t) = n-1\}\right)(\lambda \Delta t + 0(\Delta t)) + P\left(\{N(t) = (n-1)\}\right)(0(\Delta t))$$

$$\lim_{\Delta t\to 0}\frac{P\left(\left\{N(t+\Delta t)=n\right\}\right)-P\left(\left\{N(t)=n\right\}\right)}{\Delta t}=-\lambda\left[P\left(\{N(t)=n\}\right)-P\left(\{N(t)=n-1\}\right)\right]$$

$$\therefore \frac{d}{dt}P\left(\left\{N(t)=n\right\}\right)=\lambda\left[P\left(\left\{N(t)=n\right\}\right)-P\left(\left\{N(t)=n-1\right\}\right)\right] \qquad (1)$$

The above is a first-order linear differential equation with initial condition $P(\{N(0)=n\})=0.$. This differential equation can be solved recursively.

First consider the problem to find $P\left(\left\{N(t)=0\right\}\right)$

From (1)

$$\frac{d}{dt}P\left(\left\{N(t)=0\right\}\right)=\lambda P\left(\left\{N(t)=0\right\}\right)$$

$$\Rightarrow P\left\{N(t)=0\right\}=e^{-\lambda t}$$

Next to find $P(N(t)=1)$

$$\frac{d}{dt}\left(\left\{P(N(t)=1\right\}\right)=-\lambda P\left(\left\{N(t)=1\right\}\right)-\lambda P\left(\left\{N(t)=0\right\}\right)$$

$$=-\lambda P\left\{N(t)=1\right\}-\lambda e^{-\lambda t}$$

with initial condition $P\left(\left\{N(0)=1\right\}\right)=0.$

Solving the above first-order linear differential equation we get

$$P\left\{\left\{N(t)=1\right\}\right\}=\lambda t e^{-\lambda t}$$

Now by mathematical indication we can show that

$$P\left(\left\{N(t)=n\right\}\right)=\frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

Remark

(1) The parameter λ is called the rate or intensity of the Poisson process.

It can be shown that

$$P\left(\left\{N(t_2)-N(t_1)=n\right\}\right)=\frac{(\lambda(t_2-t_1))^n e^{-\lambda(t_2-t_1)}}{n!}$$

Thus the probability of the increments depends on the length of the interval t_2-t_1 and not on the absolute times t_2 and t_1. Thus the Poisson process is a process with *stationary increments.*

(2) The independent and stationary increment properties help us to compute the joint probability mass function of $N(t)$. For example,

$$P\left(\left\{N(t_1)=n_1, N(t_2)=n_2\right\}\right)=P\left(\left\{N(t_1)=n_1\right\}\right)P(\{N(t_2)=n_2\right\}/\{N(t_1)=n_1\}\right)$$
$$=P\left(\left\{N(t_1)=n_1\right\}\right)P(\{N(t_2)-N(t_1)=n_2-n_1\right\}\right)$$
$$=\frac{(\lambda t_1)^{n_1}e^{-\lambda t_1}}{n_1!}\frac{(\lambda(t_2-t_1))^{n_2-n_1}e^{-\lambda(t_2-t_1)}}{(n_2-n_1)!}$$

Mean, Variance and Covariance of the Poisson process

We observe that at any time $t>0$, $N(t)$ is a Poisson random variable with the parameter λt. Therefore,

$$EN(t)=\lambda t$$

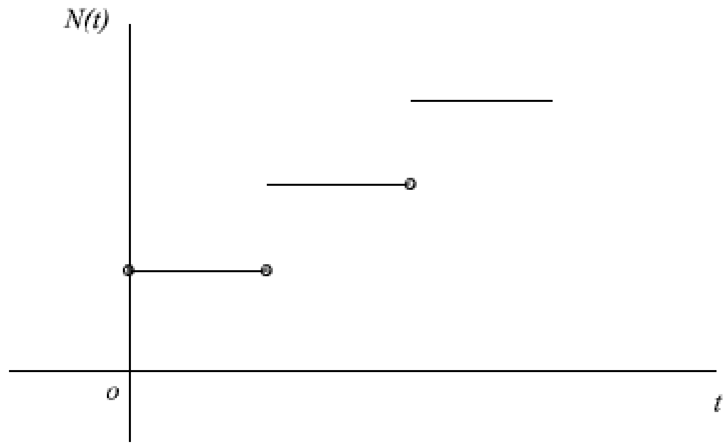
and $\operatorname{var} N(t)=\lambda t$

Thus both the mean and variance of a Poisson process varies linearly with time.

As $N(t)$ is a random process with *independent increment,* we can readily show that

$$C_N(t_1,t_2)=\operatorname{var}(N(\min(t_1,t_2)))$$
$$=\lambda \min(t_1,t_2)$$
$$\therefore R_N(t_1,t_2)= C_N(t_1,t_2)+EN(t_1)EN(t_2)$$
$$=\lambda\min(t_1,t_2)+\lambda^2 t_1 t_2$$

A typical realization of a Poisson process is shown below:



Example: A petrol pump serves on the average 30cars per hour. Find the probability that during a period of 5 minutes (i) no car comes to the station, (ii) exactly 3 cars come to the station and (iii) more than 3 cars come to the station.

Average arrival = 30 cars/hr

= ½ car/min

Probability of no car in 5 minutes

$$\text{(i)}\ P\left\{N(5)=0\right\} = e^{-\frac{1}{2}\times 5} = e^{-2.5} = 0.0821$$

$$\text{(ii)}\ P\left\{N(5)=3\right\} = \frac{\left(\frac{1}{2}\times 5\right)^3}{3!}e^{-2.5}$$

$$\text{(iii)}\ P\{N(5)>3\} = 1-0.21 = 0.79$$

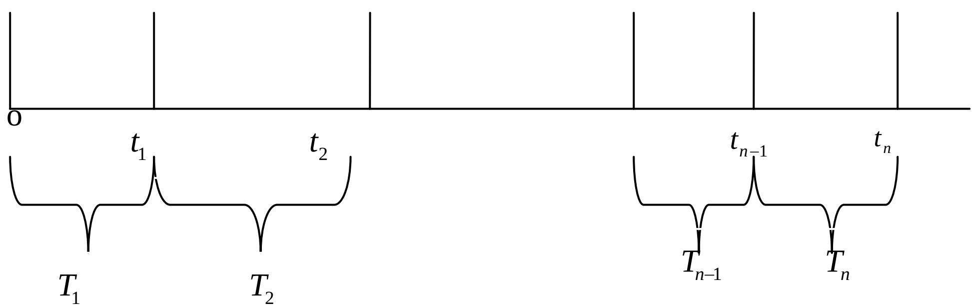
Binomial model:

P = probability of car coming in 1 minute=1/2

n = 5

$\therefore P(X=0) = (1-P)^5 = 0.5^5$

Inter-arrival time and Waiting time of Poisson Process:



Let T_n = time elapsed between (n-1) st event and nth event. The random process $\{T_n, n=1,2,..\}$ represent the *arrival time* of the Poisson process.

T_1= time elapsed before the first event take place. Clearly T_1 is a continuous random variable.

Let us find out the probability $P(\{T_1 > t\})$

$$\begin{aligned} P(\{T_1 > t\}) &= P(\{0 \text{ event upto time } t\}) \\ &= e^{-\lambda t} \end{aligned}$$

$\therefore F_{T_1}(t) = 1 - P\{T_1 > t\} = 1 - e^{-\lambda t}$

$\therefore\ f_{T_1}(t) = \lambda e^{-\lambda t} \quad t \geq 0$

Similarly,

$P(\{T_n > t\}) =$ P({0 event occurs in the interval $(t_{n-1},\ t_{n-1}+t$)/(n-1)th event occurs at (0, t_{n-1}] })

$=$ P({0 event occurs in the interval $(t_{n-1},\ t_{n-1}+t$]})

$= e^{-\lambda t}$

$\therefore\ f_{T_n}(t) = \lambda e^{-\lambda t}$

Thus the inter-arrival times of a Poisson process with the parameter λ are exponentially distributed. with

$$f_{T_n}(t) = \lambda e^{-\lambda t} \qquad n > 0$$

Remark

- We have seen that the inter-arrival times are identically distributed with the exponential pdf. Further, we can easily prove that the inter-arrival times are *independent* also. For example

$$\begin{aligned}
F_{T_1,T_2}(t_1,t_2) &= P(\{T_1 \leq t_1, T_2 \leq t_2\}) \\
&= P(\{T_1 \leq t_1\})P(\{T_2 \leq t_2\}/\{T_1 \leq t_1\}) \\
&= P(\{T_1 \leq t_1\})P(\{\text{zero event occurs in } (T_1,\ T_1+t_2)\}/\{\text{one event occurs in } (0_1,\ T_1]\}) \\
&= P(\{T_1 \leq t_1\})P(\{\text{zero event occurs in } (T_1,\ T_1+t_2)\}) \\
&= P(\{T_1 \leq t_1\})P(\{T_2 \leq t_2\}) \\
&= F_{T_1}(t_1)F_{T_2}(t_2)
\end{aligned}$$

- It is interesting to note that the converse of the above result is also true. If the inter-arrival times between two events of a discrete state $\{N(t), t \geq 0\}$ process are exponentially distributed with mean $\frac{1}{\lambda}$, then $\{N(t), t \geq 0\}$ is a Poisson process with the parameter λ.
- The exponential distribution of the inter-arrival process indicates that the arrival process has *no memory*. Thus
  $P(\{T_n > t_0 + t_1 / T_n > t_1\}) = P(\{T_n > t_0\}) \quad \forall t_0, t_1$

Another important quantity is the *waiting time* W_n. This is the time that elapses before the nth event occurs. Thus

$$W_n = \sum_{i=1}^{n} T_i$$

How to find the first order pdf of W_n is left as an exercise. Note that W_n is the sum of n independent and identically distributed random variables.

$$\therefore EW_n = \sum_{i=1}^{n} ET_i = \sum_{i=1}^{n} ET_i = \frac{n}{\lambda}$$

and

$$var(W_n) = \text{var} \sum_{i=1}^{n} (T_i) = \sum_{i=1}^{n} \text{var}(T_i) = \frac{n}{\lambda^2}$$

Example

The number of customers arriving at a service station is a Poisson process with a rate 10 customers per minute.

(a) What is the mean arrival time of the customers?

(b) What is the probability that the second customer will arrive 5 minutes after the first customer has arrived?

(c) What is the average waiting time before the 10th customer arrives?

Semi-random Telegraph signal

Consider a two-state random process $\{X(t)\}$ with the states $X(t)=1$ and $X(t)=-1$. Suppose $X(t)=1$ if the number of events of the Poisson process $N(t)$ in the interval $(0,\ t]$ is even and $X(t)=-1$ if the number of events of the Poisson process $N(t)$ in the interval $(0,\ t]$ is odd. Such a process $\{X(t)\}$ is called the *semi-random telegraph signal* because the initial value $X(0)$ is always 1.

$$\begin{aligned} p_{X(t)}(1) &= P(\{X(t)=1\}) \\ &= P(\{N(t)=0\} + P(\{N(t)=2\}) + \ldots \\ &= e^{-\lambda t}\{1 + \frac{(\lambda t)^2}{2!} + \ldots.) \\ &= e^{-\lambda t} \cosh \lambda t \end{aligned}$$

and

$$\begin{aligned} p_{X(t)}(-1) &= P(\{X(t)=-1\}) \\ &= P(\{N(t)=1\} + P(\{N(t)=3\}) + \ldots \\ &= e^{-\lambda t}\{\frac{\lambda t}{1!} + \frac{(\lambda t)^3}{3!} + \ldots.) \\ &= e^{-\lambda t} \sinh \lambda t \end{aligned}$$

We can also find the conditional and joint probability mass functions. For example, for $t_1 < t_2$,

$$p_{X(t_1),X(t_2)}(1,1) = P(\{X(t_1)=1\})P(\{X(t_2)=1)\}/\{X(t_1)=1)\}$$
$$= e^{-\lambda t_1}\cosh\lambda t_1 P(\{N(t_2) \text{ is even }\}/\{N(t_1) \text{ is even})\}$$
$$= e^{-\lambda t_1}\cosh\lambda t_1 P(\{N(t_2)-N(t_1) \text{ is even }\}/\{N(t_1) \text{ is even})\}$$
$$= e^{-\lambda t_1}\cosh\lambda t_1 P(\{N(t_2)-N(t_1) \text{ is even }\}$$
$$= e^{-\lambda t_1}\cosh\lambda t_1 e^{-\lambda(t_2-t_1)}\cosh\lambda(t_2-t_1)$$
$$= e^{-\lambda t_2}\cosh\lambda t_1 \cosh\lambda(t_2-t_1)$$

Similarly

$$p_{X(t_1),X(t_2)}(1,-1) = e^{-\lambda t_2}\cosh\lambda t_1 \sinh\lambda(t_2-t_1),$$
$$p_{X(t_1),X(t_2)}(-1,1) = e^{-\lambda t_2}\sinh\lambda t_1 \sinh\lambda(t_2-t_1)$$
$$p_{X(t_1),X(t_2)}(-1,-1) = e^{-\lambda t_2}\sinh\lambda t_1 \cosh\lambda(t_2-t_1)$$

Mean, autocorrelation and autocovariance function of $X(t)$

$$EX(t) = 1\times e^{-\lambda t}\cosh\lambda t - 1\times e^{-\lambda t}\sinh\lambda t$$
$$= e^{-\lambda t}(\cosh\lambda t - \sinh\lambda t)$$
$$= e^{-2\lambda t}$$
$$EX^2(t) = 1\times e^{-\lambda t}\cosh\lambda t + 1\times e^{-\lambda t}\sinh\lambda t$$
$$= e^{-\lambda t}(\cosh\lambda t + \sinh\lambda t)$$
$$= e^{-\lambda t}e^{\lambda t}$$
$$= 1$$
$$\therefore \operatorname{var}(X(t)) = 1 - e^{-4\lambda t}$$

For $t_1 < t_2$

$$R_X(t_1,t_2) = EX(t_1)X(t_2)$$
$$= 1\times 1\times p_{X(t_1),X(t_2)}(1,1) + 1\times(-1)\times p_{X(t_1),X(t_2)}(1,-1)$$
$$+ (-1)\times 1\times p_{X(t_1),X(t_2)}(-1,1) + (-1)\times(-1)\times p_{X(t_1),X(t_2)}(-1,-1)$$
$$= e^{-\lambda t_2}\cosh\lambda t_1\cosh\lambda(t_2-t_1) - e^{-\lambda t_2}\cosh\lambda t_1\sinh\lambda(t_2-t_1)$$
$$- e^{-\lambda t_2}\sinh\lambda t_1\sinh\lambda(t_2-t_1) + e^{-\lambda t_2}\sinh\lambda t_1\cosh\lambda(t_2-t_1)$$
$$= e^{-\lambda t_2}\cosh\lambda t_1(\cosh\lambda(t_2-t_1) - \sinh\lambda(t_2-t_1)) + e^{-\lambda t_2}\sinh\lambda t_1(\cosh\lambda(t_2-t_1) -_1 \sinh\lambda(t_2-t_1))$$
$$= e^{-\lambda t_2}e^{-\lambda(t_2-t_1)}(\cosh\lambda t_1 + \sinh\lambda t_1)$$
$$= e^{-\lambda t_2}e^{-\lambda(t_2-t_1)}e^{\lambda t_1}$$
$$= e^{-2\lambda(t_2-t_1)}$$

Similarlrly for $t_1 > t_2$

$$R_X(t_1,t_2) = e^{-2\lambda(t_1-t_2)}$$
$$\therefore R_X(t_1,t_2) = e^{-2\lambda|t_1-t_2|}$$

Random Telegraph signal

Consider a two-state random process $\{Y(t)\}$ with the states $Y(t)=1$ and $Y(t)=-1$.

Suppose $P(\{Y(0)=1\})=\frac{1}{2}$ and $P(\{Y(0)=-1\})=\frac{1}{2}$ and $Y(t)$ changes polarity with an equal probability with each occurrence of an event in a Poisson process of parameter λ.

Such a random process $\{Y(t)\}$ is called a *random telegraph signal* and can be expressed as

$$Y(t)=AX(t)$$

where $\{X(t)\}$ is the semirandom telegraph signal and A is a random variable independent of $X(t)$ with $P(\{A=1\})=\frac{1}{2}$ and $P(\{A=-1\})=\frac{1}{2}$.

Clearly,

$$EA=(-1)\times\frac{1}{2}+1\times\frac{1}{2}=0$$

and

$$EA^2=(-1)^2\times\frac{1}{2}+1^2\times\frac{1}{2}=1$$

Therefore,

$$\begin{aligned}EY(t)&=EAX(t)\\&=EAEX(t) \qquad \because A \text{ and } X(t) \text{ are independent}\\&=0\end{aligned}$$

$$\begin{aligned}R_Y(t_1,t_2)&=EAX(t_1)AX(t_2)\\&=EA^2EX(t_1)X(t_2)\\&=e^{-2\lambda|t_1-t_2|}\end{aligned}$$

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 random process. 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

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

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

Thus the joint distribution functions of any set of random variables $X(t_1), X(t_2), ..., X(t_n)$ 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 $F_{X(t_1),X(t_2).....X(t_n)}(x_1, x_2.....x_n) = F_{X(t_1+t_0),X(t_2+t_0).....X(t_n+t_0)}(x_1, x_2.....x_n)$ for $n = 1, 2, .., k,$ then $\{X(t)\}$ is called *kth order stationary*.

- 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

$$F_{X(t_1),X(t_2)}(x_1,x_2.) = F_{X(t_1+t_0),X(t_2+t_0)}(x_1,x_2)$$

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 stationarity 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 process $\{Z(t) = X(t) + jY(t)\}$ is called SSS if $\{X(t)\}$ and $\{Y(t)\}$ are jointly SSS.

Example An iid process is SSS. This is because, $\forall n$,

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

Example The Poisson process is $\{N(t), t \geq 0\}$ not stationary, because

$$EN(t) = \lambda t$$

which varies with time.

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

$EX(t) = \mu_X$ = constant

and

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

(Equivalently, $Cov(X(t_1)X(t_2)) = C_X(t_1 - t_2)$ is a function of time lag $t_1 - t_2$)

Remark

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

$\therefore EX^2(t) = R_X(0) = \text{constant}$

$\text{var}(X(t) = EX^2(t) - (EX(t))^2 = \text{constant}$

$$
\begin{aligned}
C_X(t_1, t_2) &= EX(t_1)X(t_2) - EX(t_1)EX(t_2) \\
&= R_X(t_2 - t_1) - \mu_X^2
\end{aligned}
$$

$\therefore C_X(t_1, t_2)$ is a function of lag $(t_2 - t_1)$.

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

Example: Sinusoid with random phase

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

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

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

Note that

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

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

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

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

Note that

$$EX(t) = EA\cos(w_0 t + \phi)$$
$$= \int_0^{2\pi} A\cos(w_0 t + \phi)\frac{1}{2\pi}d\phi$$
$= 0$ which is a constant

and

$$R_X(t_1, t_2) = EX(t_1)X(t_2)$$
$$= EA\cos(w_0 t_1 + \phi)A\cos(w_0 t_2 + \phi)$$
$$= \frac{A^2}{2}E[c\,os(w_0 t_1 + \phi + w_0 t_2 + \phi) + c\,os(w_0 t_1 + \phi - w_0 t_2 - \phi)]$$
$$= \frac{A^2}{2}E[c\,os(w_0(t_1 + t_2) + 2\phi) + c\,os(w_0(t_1 - t_2)]$$
$= \frac{A^2}{2}c\,os(w_0(t_1 - t_2)$ which is a function of the lag $t_1 - t_2$.

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

Example: Sinusoid with random amplitude

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

$X(t) = A\cos(w_0 t + \phi)$ where ϕ and w_0 are constants and A is a random variable. Here,

$$EX(t) = EA \times \cos(w_0 t + \phi)$$

which is independent of time only if $EA = 0$.

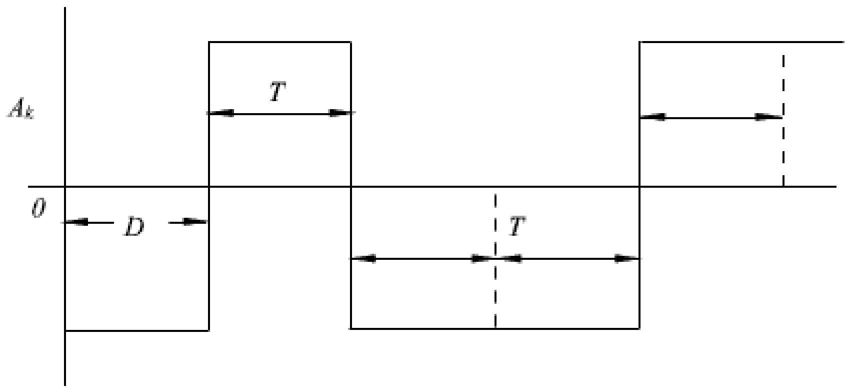
$$R_X(t_1, t_2) = EX(t_1)X(t_2)$$
$$= EA\cos(w_0 t_1 + \phi)A\cos(w_0 t_2 + \phi)$$
$$= EA^2 \times \cos(w_0 t_1 + \phi)\cos(w_0 t_2 + \phi)$$
$$= EA^2 \times \frac{1}{2}[c\,os(w_0(t_1 + t_2) + 2\phi) + c\,os(w_0(t_1 - t_2)]$$

which will not be function of $(t_1 - t_2)$ only.

Example: Random binary wave

Consider a binary random process $\{X(t)\}$ consisting of a sequence of random pulses of duration T with the following features:

- Pulse amplitude A_K is a random variable with two values $p_{A_K}(1) = \frac{1}{2}$ and $p_{A_K}(-1) = \frac{1}{2}$



- Pulse amplitudes at different pulse durations are independent of each other.
- The start time of the pulse sequence can be any value between 0 to *T*. Thus the random start time *D* (Delay) is uniformly distributed between 0 and *T*.

A realization of the random binary wave is shown in Fig. above. Such waveforms are used in binary munication- a pulse of amplitude 1is used to transmit '1' and a pulse of amplitude -1 is used to transmit '0'.

The random process *X* (*t*) can be written as,

$$X(t) = \sum_{n=-\infty}^{\infty} A_n rect \frac{(t - nT - D)}{T}$$

For any *t*,

$$EX(t) = 1 \times \frac{1}{2} + (-1)(\frac{1}{2}) = 0$$

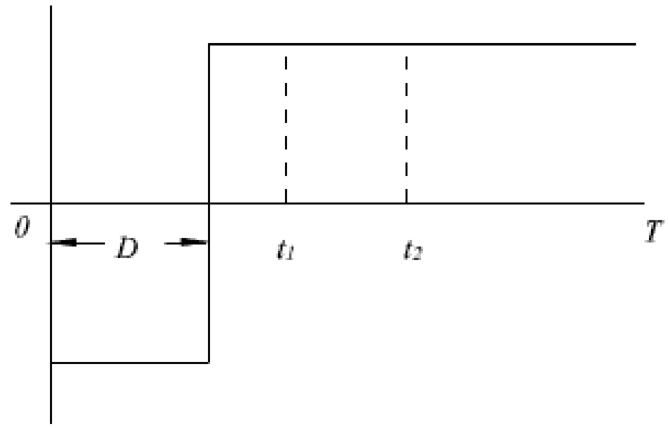
$$EX^2(t) = 1^2 \times \frac{1}{2} + (-1)^2 \frac{1}{2} = 1$$

Thus mean and variance of the process are constants.

To find the autocorrelation function $R_X(t_1, t_2)$ let us consider the case $0 < t_1 < t_1 + \tau < T$. Depending on the delay *D*, the points t_1 and t_2 may lie on one or two pulse intervals.

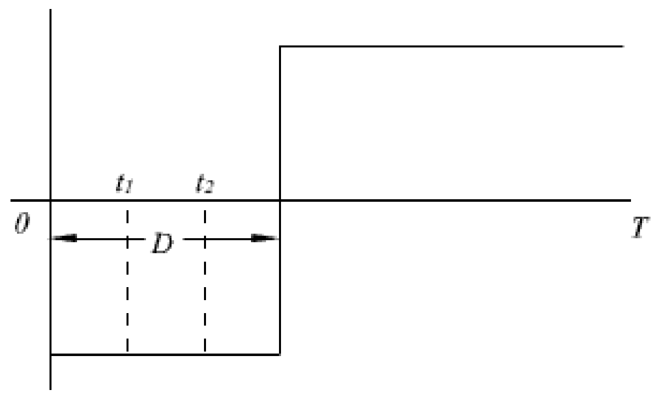
Case 1:

$X(t_1)X(t_2) = 1$



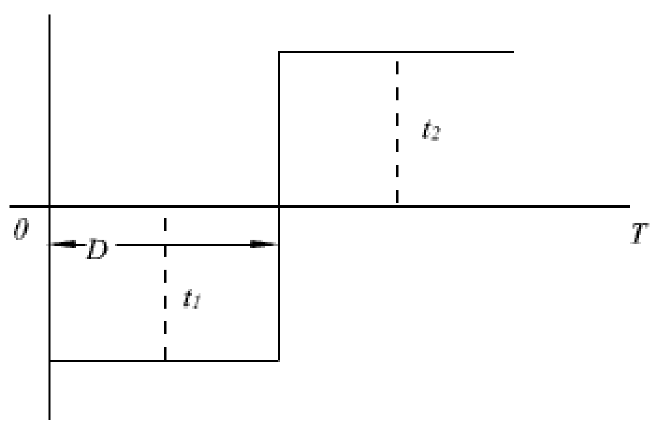
Case 2:

$X(t_1)X(t_1) = (-1)(-1) = 1$



Case 3:

$X(t_1)X(t_2) = -1$



Thus,

$$X(t_1)X(t_2) = \begin{cases} 1 & 0 < D < t_1 \quad or \quad t_2 < D < T \\ -1 & t_1 < D < t_2 \end{cases}$$

$$\begin{aligned}
\therefore R_X(t_1, t_2) &= EEX(t_1)X(t_2) \mid D \\
&= E(X(t_1)X(t_2) \mid P(0 < D < t_1 \;\; or \;\; t_2 < D < T)) + E(X(t_1)X(t_2)) \mid t_1 < D < t_2).P(t_1 < D < t_2) \\
&= 1 \times \left(1 - \frac{t_2 - t_1}{T}\right) + (1 \times \frac{1}{2} - 1 \times \frac{1}{2})\frac{t_2 - t_1}{T} \\
&= 1 - \frac{t_2 - t_1}{T}
\end{aligned}$$

We also have, $R_X(t_2, t_1) = EX(t_2)X(t_1) = EX(t_1)X(t_2) = R_X(t_1, t_2)$

So that $R_X(t_1, t_2) = 1 - \frac{|t_2 - t_1|}{T} \qquad |t_2 - t_1| \le T$

For $|t_2 - t_1| > T$, t_1 and t_2 are at different pulse intervals.

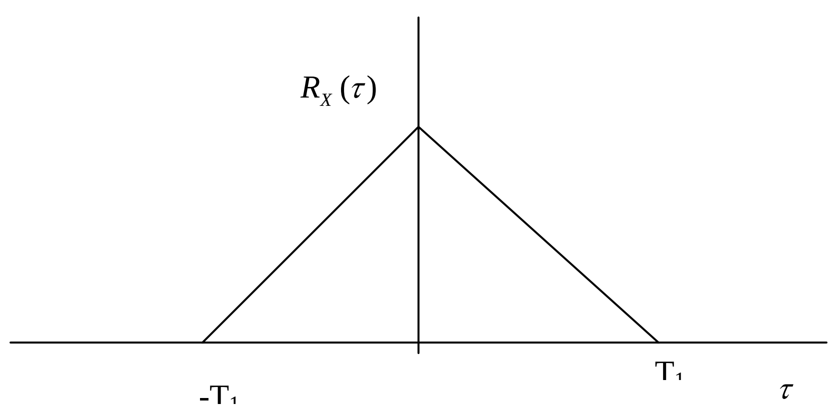
$\therefore EX(t_1)X(t_2) = EX(t_1)EX(t_2) = 0$

Thus the autocorrelation function for the random binary waveform depends on

$\tau = t_2 - t_1$, and we can write

$$R_X(\tau) = 1 - \frac{|\tau_1|}{T} \qquad |\tau| \le T$$

The plot of $R_X(\tau)$ is shown below.



Example Gaussian Random Process

Consider the Gaussian process $\{X(t)\}$ discussed earlier. For any positive integer $n,\ X(t_1), X(t_2),.....X(t_n)$ is jointly Gaussian with the joint density function given by

$$f_{X(t_1),X(t_2)...X(t_n)}(x_1,x_2,...x_n) = \frac{e^{-\frac{1}{2}\mathbf{X}'\mathbf{C}_X^{-1}\mathbf{X}}}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(\mathbf{C_X})}}$$

where $\mathbf{C_X} = E(\mathbf{X}-\boldsymbol{\mu}_\mathbf{X})(\mathbf{X}-\boldsymbol{\mu}_\mathbf{X})'$

and $\boldsymbol{\mu}_\mathbf{X} = E(\mathbf{X}) = \left[E(X_1), E(X_2)......E(X_n)\right]'.$

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

$$\boldsymbol{\mu}_\mathbf{X} = \begin{bmatrix} EX(t_1) \\ EX(t_2) \\ .. \\ .. \\ EX(t_n) \end{bmatrix} = \begin{bmatrix} \mu_X \\ \mu_X \\ .. \\ .. \\ \mu_X \end{bmatrix}$$

$$C_\mathbf{X} = E\left(\begin{bmatrix} X(t_1)-\mu_X \\ X(t_2)-\mu_X \\ .. \\ .. \\ X(t_n)-\mu_X \end{bmatrix} \begin{bmatrix} X(t_1)-\mu_X \\ X(t_2)-\mu_X \\ .. \\ .. \\ X(t_n)-\mu_X \end{bmatrix}' \right)$$

$$= \begin{bmatrix} C_X(0) & C_X(t_2-t_1)... & C_X(t_n-t_1) \\ C_X(t_2-t_1) & C_X(0)... & C_X(t_2-t_n) \\ \vdots & & \\ C_X(t_n-t_1) & C_X(t_n-t_1)... & C_X(0) \end{bmatrix}$$

We see that $f_{X(t_1),X(t_2)...X(t_n)}(x_1,x_2,...x_n)$ depends on the time-lags. Thus, for a Gaussian random process, WSS implies strict sense stationarity, because this process is completely described by the mean and the autocorrelation functions.

Properties 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 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 $\tau = 0, \pm\frac{2\pi}{\omega}, \pm\frac{2\pi}{\omega},$ 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 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 charactersing 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. If $X(t)$ is a voltage signal applied across a 1 ohm resistance, then $R_X(0)$ is the ensemble average power delivered to the resistance. Thus,

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

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

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

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

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

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

$$R_X^2(\tau) = \{EX(t)X(t+\tau)\}^2$$
$$= EX^2(t)EX^2(t+\tau)$$
$$= R_X(0)R_X(0)$$
$$\therefore |R_X(\tau)| \le R_X(0)$$

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

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.*

5. 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

$$E((X(t+\tau+T_p)-X(t+\tau))X(t))^2 \leq \; E(X(t+\tau+T_p)-X(t+\tau))^2 EX^2(t)$$
$$\Rightarrow (R_X(\tau+T_p)-R_X(\tau))^2 \leq 2(R_X(0)-R_X(T_p))R_X(0)$$
$$\Rightarrow (R_X(\tau+T_p)-R_X(\tau))^2 \leq 0 \qquad\qquad \because R_X(0)=R_X(T_p)$$
$$\therefore R_X(\tau+T_p)=R_X(\tau)$$

For example, $X(t)=A\cos(w_0 t+\phi)$ where A and w_0 are constants and $\phi \sim U[0,\ 2\pi]$, is MS periodic random process with a period $\frac{2\pi}{\omega_0}$. Its autocorrelation function
$R_X(\tau)=\frac{A^2\cos\omega_0\tau}{2}$ is periodic with the same period $\frac{2\pi}{\omega_0}$.
The converse of this result is also true. If $R_X(\tau)$ is periodic with period T_p then $X(t)$ is MS periodic with a period T_p. This property helps us in determining time period of a MS periodic random process.

6. Suppose $X(t)=\mu_X+V(t)$
   where $V(t)$ is a zero-mean WSS process and $\lim_{\tau\to\infty} R_V(\tau)=0.$ Then
   $\lim_{\tau\to\infty} R_X(\tau)=\mu_X^2$

Interpretation of the autocorrelation function of a WSS process

The autocorrelation function $R_X(\tau)$ measures the correlation between two random variables $X(t)$ and $X(t+\tau)$. If $R_X(\tau)$ drops quickly with respect to τ, then the $X(t)$ and $X(t+\tau)$ will be less correlated for large τ. This in turn means that the signal has lot of changes with respect to time. Such a signal has high frequency components. If $R_X(\tau)$ drops slowly, the signal samples are highly correlated and such a signal has less high frequency components. Later on we see that $R_X(\tau)$ is directly related to the frequency -domain representation of a WSS process.

Fig.

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:

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

This is because

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

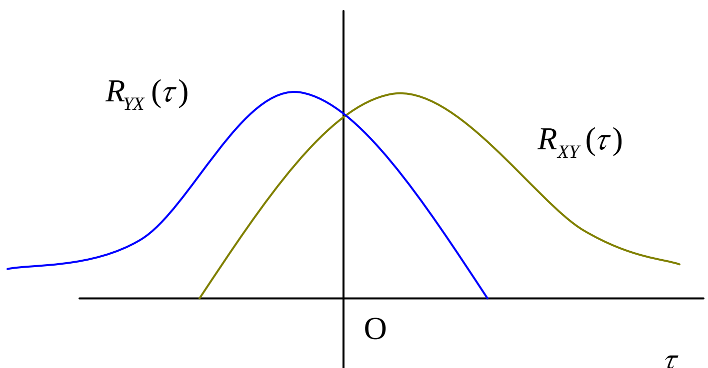


Fig. $R_{XY}(\tau) = R_{YX}(-\tau)$

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

We have

$$\begin{aligned} \left|R_{XY}(\tau)\right|^2 &= \left|EX(t+\tau)Y(t)\right|^2 \\ &\leq EX^2(t+\tau)EY^2(t) \quad \text{using Cauch-Schwarts Inequality} \\ &= R_X(0)R_Y(0) \end{aligned}$$

$\therefore \left|R_{XY}(\tau)\right| \leq \sqrt{R_X(0)R_Y(0)}$

Further,

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

$\therefore \left|R_{XY}(\tau)\right| \leq \sqrt{R_X(0)R_Y(0)} \leq \frac{1}{2}\left(R_X(0)+R_Y(0)\right)$

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

(iv) If $X(t)$ and Y (t) is orthogonal process, $R_{XY}(\tau) = EX\left(t+\tau\right)Y\left(t\right) = 0$

Example

Consider a random process $Z(t)$ which is sum of two real jointly WSS random processes $X(t)$ and $Y(t)$. We have

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

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

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

Continuity and Differentiation of Random Processes

- We know that the dynamic behavior of a system is described by differential equations involving input to and output of the system. For example, the behavior of the RC network is described by a first order linear differential equation with the source voltage as the input and the capacitor voltage as the output. *What happens if the input voltage to the network is a random process?*

- Each realization of the random process is a deterministic function and the concepts of differentiation and integration can be applied to it. Our aim is to extend these concepts to the ensemble of realizations and apply calculus to random process.

We discussed the convergence and the limit of a random sequence. The continuity of the random process can be defined with the help of convergence and limits of a random process. We can define continuity with probability 1, mean-square continuity, and continuity in probability etc. We shall discuss the *mean-square continuity* and the elementary concepts of corresponding *mean-square calculus*.

Mean-square continuity of a random process

Recall that a sequence of random variables $\{X_n\}$ converges to a random variable X if

$$\lim_{n\to\infty} E\left[X_n - X\right]^2 = 0$$

and we write

$$\underset{n\to\infty}{\text{l.i.m.}} X_n = X$$

A random process $\{X(t)\}$ is said to be continuous at a point $t = t_0$ in the mean-square sense if $\underset{t\to t_0}{\text{l.i.m.}} X(t) = X(t_0)$ or equvalently

$$\lim_{t\to t_0} E\left[X(t) - X(t_0)\right]^2 = 0$$

Mean-square continuity and autocorrelation function

(1) A random process $\{X(t)\}$ is MS continuous at t_0 if its auto correlation function $R_X(t_1,t_2)$ is continuous at (t_0,t_0).

Proof:

$$\begin{aligned} E\left[X(t) - X(t_0)\right]^2 &= E\left(X^2(t) - 2X(t)X(t_0) + X^2(t_0)\right) \\ &= R_X(t,t) - 2R_X(t,t_0) + R_X(t_0,t_0) \end{aligned}$$

If $R_X(t_1,t_2)$ is continuous at (t_0,t_0), then

$$\lim_{t\to t_0} E\left[X(t)-X(t_0)\right]^2 = \lim_{t\to t_0} R_X(t,t)-2R_X(t,t_0)+R_X(t_0,t_0)$$
$$= R_X(t_0,t_0)-2R_X(t_0,t_0)+R_X(t_0,t_0)$$
$$= 0$$

(2) If $\{X(t)\}$ is MS continuous at t_0 its mean is continuous at t_0.

This follows from the fact that

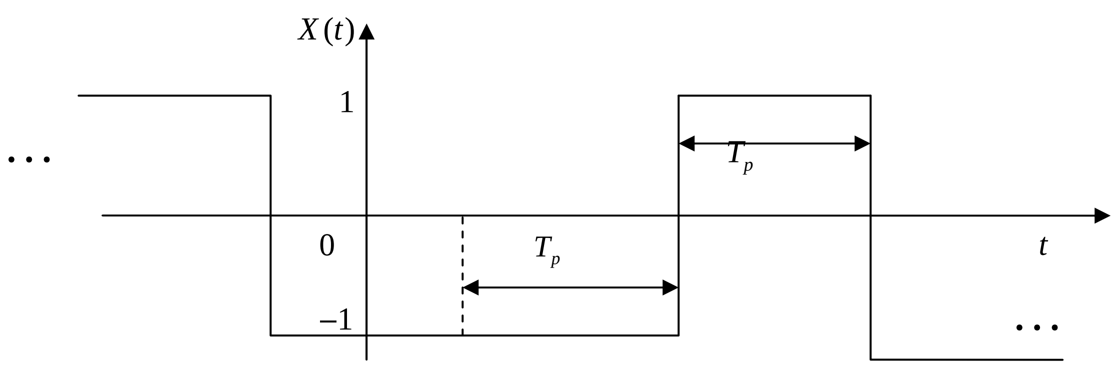
$$\left(E\left[X(t)-X(t_0)\right]\right)^2 \le E\left[\left(X(t)-X(t_0)\right)\right]^2$$

$$\therefore \lim_{t\to t_0}\left[E\left[X(t)-X(t_0)\right]\right]^2 \le \lim_{t\to t_0} E\left[X(t)-X(t_0)\right]^2 = 0$$

$\therefore EX(t)$ is continuous at t_0.

Example

Consider the random binary wave $\{X(t)\}$ discussed in Eaxmple . Atypical realization of the process is shown in Fig. below. The realization is a discontinuous function.



The process has the autocorrelation function given by

$$R_X(\tau) = \begin{cases} 1-\dfrac{|\tau|}{T_p} & |\tau| \le T_p \\ 0 & \text{otherwise} \end{cases}$$

We observe that $R_X(\tau)$ is continuous at $\tau = 0$. Therefore, $R_X(\tau)$ is continuous at all τ.

Example For a Wiener process $\{X(t)\}$,

$R_X(t_1, t_2) = \alpha \min(t_1, t_2)$

where α is a constant.

$\therefore R_X(t,t) = \alpha \min(t,t) = \alpha t$

Thus the autocorrelation function of a Wiener process is continuous everywhere implying that a Wiener process is m.s. continuous everywhere. We can similarly show that the Poisson process is m.s. continuous everywhere.

Mean-square differentiability

The random process $\{X(t)\}$ is said to have the mean-square derivative $X'(t)$ at a point $t \in \Gamma$, provided $\frac{X(t+\Delta t) - X(t)}{\Delta t}$ approaches $X'(t)$ in the mean square sense as $\Delta t \to 0$. In other words, the random process $\{X(t)\}$ has a m-s derivative $X'(t)$ if

$$\lim_{\Delta t \to 0} E\left[\frac{X(t+\Delta t) - X(t)}{\Delta t} - X'(t)\right]^2 = 0$$

Remark

(1) If all the sample functions of a random process $X(t)$ is differentiable, then the above condition is satisfied and the m-s derivative exists.

Example Consider the random-phase sinusoid $\{X(t)\}$ given by

$X(t) = A\cos(w_0 t + \phi)$ where A and w_0 are constants and $\phi \sim U[0,\ 2\pi]$. Then for each ϕ, $X(t)$ is differentiable. Therefore, the m.s. derivative is

$$X'(t) = -Aw_0 \sin(w_0 t + \phi)$$

M.S. Derivative and Autocorrelation functions

The m-s derivative of a random process $X(t)$ at a point $t \in \Gamma$ exists if $\frac{\partial^2 R_X(t_1, t_2)}{\partial t_1 \partial t_2}$ exists at the point (t,t).

Applying the Cauchy criterion, the condition for existence of m-s derivative is

$$\lim_{\Delta t_1, \Delta t_2 \to 0} E\left[\frac{X(t+\Delta t_1) - X(t)}{\Delta t_1} - \frac{X(t+\Delta t_2) - X(t)}{\Delta t_2}\right]^2 = 0$$

Expanding the square and taking expectation results,

$$\lim_{\Delta t_1,\Delta t_2\to 0} E\left[\frac{X(t+\Delta t_1)-X(t)}{\Delta t_1}-\frac{X(t+\Delta t_2)-X(t)}{\Delta t_2}\right]^2$$

$$=\lim_{\Delta t_1,\Delta t_2\to 0}\left[\frac{R_X(t+\Delta t_1,t+\Delta t_1)+R_X(t,t)-2R_X(t+\Delta t_1,t)}{\Delta t_1^{\,2}}\right]+\left[\frac{R_X(t+\Delta t_2,t+\Delta t_2)+R_X(t,t)-2R_X(t+\Delta t_2,t)}{\Delta t_2^{\,2}}\right]$$

$$-2\left[\frac{R_X(t+\Delta t_1,t+\Delta t_2)-R_X(t+\Delta t_1,t)-R_X(t,t+\Delta t_2)+R_X(t,t)}{\Delta t_1\Delta t_2}\right]$$

Each of the above terms within square bracket converges to $\left.\frac{\partial^2 R_X(t_1,t_2)}{\partial t_1\partial t_2}\right]_{t_1=t,t_2=t}$ if the second partial derivative exists.

$$\therefore \lim_{\Delta t_1,\Delta t_2\to 0} E\left[\frac{X(t+\Delta t_1)-X(t)}{\Delta t_1}-\frac{X(t+\Delta t_2)-X(t)}{\Delta t_2}\right]^2=\left.\frac{\partial^2 R_X(t_1,t_2)}{\partial t_1\partial t_2}+\frac{\partial^2 R_X(t_1,t_2)}{\partial t_1\partial t_2}-2\frac{\partial^2 R_X(t_1,t_2)}{\partial t_1\partial t_2}\right]_{t=t_1,t=t_2}$$

$$=0$$

Thus, $\{X(t)\}$ is m-s differentiable at $t\in\Gamma$ if $\frac{\partial^2 R_X(t_1,t_2)}{\partial t_1\partial t_2}$ exists at $(t,t)\in\Gamma\times\Gamma$.

Particularly, if $X(t)$ is WSS,

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

Substituting $t_1-t_2=\tau$, we get

$$\frac{\partial^2 R_X(t_1,t_2)}{\partial t_1\partial t_2}=\frac{\partial^2 R_X(t_1-t_2)}{\partial t_1\partial t_2}$$

$$=\frac{\partial}{\partial t_1}\left(\frac{dR_X(\tau)}{d\tau}.\frac{\partial(t_1-t_2)}{\partial t_2}\right)$$

$$=-\frac{d^2R_X(\tau)}{d\tau^2}\frac{\partial\tau}{\partial t_1}$$

$$=-\frac{d^2R_X(\tau)}{d\tau^2}$$

Therefore, a WSS process $X(t)$ is m-s differentiable if $R_X(\tau)$ has second derivative at $\tau = 0$.

Example
Consider a WSS process $\{X(t)\}$ with autocorrelation function

$R_X(\tau) = \exp(-a|\tau|)$

$R_X(\tau)$ does not have the first and second derivative at $\tau = 0$. $\{X(t)\}$ is not mean-square differentiable.

Example The random binary wave $\{X(t)\}$ has the autocorrelation function

$$R_X(\tau) = \begin{cases} 1 - \dfrac{|\tau|}{T_p} & |\tau| \le T_p \\ 0 & \text{otherwise} \end{cases}$$

$R_X(\tau)$ does not have the first and second derivative at $\tau = 0$. Therefore, $\{X(t)\}$ is not mean-square differentiable.

Example For a Wiener process $\{X(t)\}$,

$R_X(t_1, t_2) = \alpha \min(t_1, t_2)$

where α is a constant.

$$\therefore R_X(0, t_2) = \begin{cases} \alpha t_2 & \text{if } t_2 < 0 \\ 0 & \text{other wise} \end{cases}$$

$$\therefore \frac{\partial R_X(0, t_2)}{\partial t_1} = \begin{cases} \alpha & \text{if } t_2 < 0 \\ 0 & \text{if } t_2 > 0 \\ \text{does not exist if if } t_2 = 0 \end{cases}$$

$$\therefore \frac{\partial^2 R_X(t_1, t_2)}{\partial t_1 \partial t_2} \text{ does not exist at } (t_1 = 0, t_2 = 0)$$

Thus a Wiener process is m.s. differentiable nowhere..

Mean and Autocorrelation of the Derivative process

We have,

$$
\begin{aligned}
EX'(t) &= E\lim_{\Delta t\to 0}\frac{X(t+\Delta t)-X(t)}{\Delta t}\\
&= \lim_{\Delta t\to 0}\frac{EX(t+\Delta t)-EX(t)}{\Delta t}\\
&= \lim_{\Delta t\to 0}\frac{\mu_X(t+\Delta t)-\mu_X(t)}{\Delta t}\\
&= \mu_X'(t)
\end{aligned}
$$

For a WSS process $EX'(t)=\mu_X'(t)=0$ as $\mu_X(t)=$ constant.

$$
\begin{aligned}
EX(t_1)X'(t_2) &= R_{XX'}(t_1,t_2)\\
&= EX(t_1)\lim_{\Delta t_2\to 0}\frac{X(t_2+\Delta t_2)-X(t_2)}{\Delta t_2}\\
&= \lim_{\Delta t_2\to 0}\frac{E\left[X(t_1)X(t_2+\Delta t_2)-X(t_1)X(t_2)\right]}{\Delta t_2}\\
&= \lim_{\Delta t_2\to 0}\frac{R_X(t_1,t_2+\Delta t_2)-R_X(t_1,t_2)}{\Delta t_2}\\
&= \frac{\partial R_X}{\partial t_2}(t_1,t_2)
\end{aligned}
$$

Similarly we can show that

$$
EX'(t_1)X'(t_2)=\frac{\partial^2 R_{XX}(t_1,t_2)}{\partial t_1\partial t_2}
$$

For a WSS process

$$
EX(t_1)X'(t_2)=\frac{\partial}{\partial t}R_X(t_1-t_2)=\frac{dR_X(\tau)}{d\tau}
$$

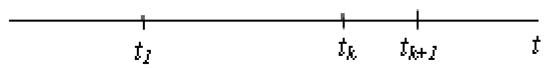
and

$$
\begin{aligned}
EX'(t_1)X'(t_2) &= R_X{}'(t_1-t_2)\\
&= \frac{d^2R_X(\tau)}{d\tau^2}\\
\therefore \operatorname{var}(X'(t) &= \left.\frac{d^2R_X(\tau)}{d\tau^2}\right|_{\tau=0}
\end{aligned}
$$

Mean Square Integral

Recall that the definite integral (Riemannian integral) of a function $x(t)$ over the interval $[t_0,\ t]$ is defined as the limiting sum given by

$$\int_{t_0}^{t} x(\tau)d\tau = \lim_{n\to\infty\,\Delta k\to 0} \sum_{k=0}^{n-1} x(\tau_k)\Delta_k$$



Where $t_0 < t_1 < \ldots\ldots\ldots\ldots < t_{n-1} < t_n = t$ are partitions on the interval $[t_0,\ t]$ and $\Delta_k = t_{k+1} - t_k$ and $\tau_k \in [t_{k-1},\ t_k]$.

For a random process $\{X(t)\}$, the m-s integral can be similarly defined as the process $\{Y(t)\}$ given by

$$Y(t) = \int_{t_0}^{t} X(\tau)d\tau = \underset{n\to\infty\,\Delta k\to 0}{\text{l.i.m.}} \sum_{k=0}^{n-1} X(\tau_k)\Delta_k$$

Existence of M.S. Integral

- It can be shown that the sufficient condition for the m-s integral $\int_{t_0}^{t} X(\tau)d\tau$ to exist is that the double integral $\int_{t_0}^{t}\int_{t_0}^{t} R_X(\tau_1,\tau_2)\,d\tau_1 d\tau_2$ exists.
- If $\{X(t)\}$ is M.S. continuous, then the above condition is satisfied and the process is M.S. integrable.

Mean and Autocorrelation of the Integral of a WSS process

We have

$$EY(t) = E\int_{t_0}^{t} X(\tau)d\tau$$

$$= \int_{t_0}^{t} EX(\tau)d\tau$$

$$= \int_{t_0}^{t} \mu_X d\tau$$

$$= \mu_X (t - t_0)$$

Therefore, if $\mu_X \neq 0$, $\{Y(t)\}$ is necessarily non-stationary.

$$R_Y(t_1, t_2) = EY(t_1)Y(t_2)$$

$$= E\int_{t_0}^{t_1}\int_{t_0}^{t_2} X(\tau_1)X(\tau_2)d\tau_1 d\tau_2$$

$$= \int_{t_0}^{t_1}\int_{t_0}^{t_2} EX(\tau_1)X(\tau_2)d\tau_1 d\tau_2$$

$$= \int_{t_0}^{t_1}\int_{t_0}^{t_2} R_X(\tau_1 - \tau_2)d\tau_1 d\tau_2$$

which is a function of t_1 and t_2.

Thus the integral of a WSS process is always non-stationary.

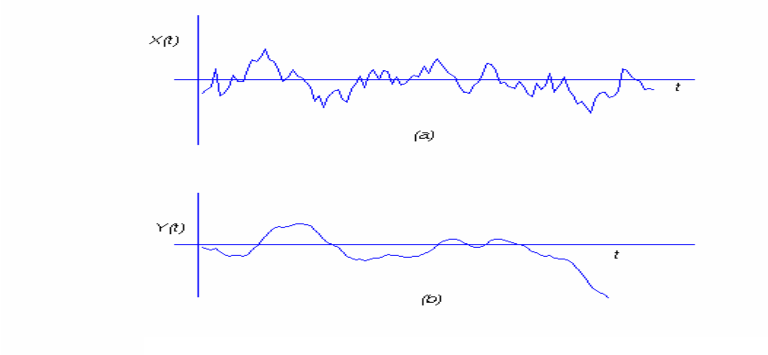


Fig. (a) Realization of a WSS process $X(t)$

(b) corresponding integral $Y(t)$

Remark The nonstatinarity of the M.S. integral of a random process has physical importance – the output of an integrator due to stationary noise rises unboundedly.

Example The random binary wave $\{X(t)\}$ has the autocorrelation function

$$R_X(\tau) = \begin{cases} 1 - \dfrac{|\tau|}{T_p} & |\tau| \le T_p \\ 0 & \text{otherwise} \end{cases}$$

$R_X(\tau)$ is continuous at $\tau = 0$ implying that $\{X(t)\}$ is M.S. continuous. Therefore, $\{X(t)\}$ is mean-square integrable.

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. $\langle \mu_x \rangle_T$ represents the dc value of $x(t)$.
Another important average used in electrical engineering is the rms value given by

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

Can $\langle \mu_x \rangle_T$ and $\langle x_{rms} \rangle_T$ represent μ_X and $\sqrt{EX^2(t)}$ respectively?
To answer such a question we have understand various time averages and their properties.

Time averages of a random process

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

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

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

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

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

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

$$Eg(X(t)) = \int_{-\infty}^{\infty} g(x) f_{X(t)}(x)dx \qquad \text{for continuous case}$$

$$= \sum_{i\in R_{X(t)}} g(x_i) p_{X(t)}(x_i) \qquad \text{for discrete case}$$

The following time averages are of particular interest:

(a) Time-averaged mean

$$\langle \mu_X \rangle_T = \frac{1}{2T}\int_{-T}^{T} X(t)dt \qquad \text{(continuous case)}$$

$$\langle \mu_X \rangle_N = \frac{1}{2N+1}\sum_{i=-N}^{N} X_i \qquad \text{(discrete case)}$$

(b) Time-averaged autocorrelation function

$$\langle R_X(\tau) \rangle_T = \frac{1}{2T}\int_{-T}^{T} X(t)X(t+\tau)dt \qquad \text{(continuous case)}$$

$$\langle R_X[m] \rangle_N = \frac{1}{2N+1}\sum_{i=-N}^{N} X_i X_{i+m} \qquad \text{(discrete case)}$$

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

Mean and Variance of the time averages

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

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

The mean of $\langle \mu_X \rangle_N$

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

and the variance

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

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

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

$$= E\left(\frac{1}{2T}\int_{-T}^{T} (X(t) - \mu_X)dt\right)^2$$

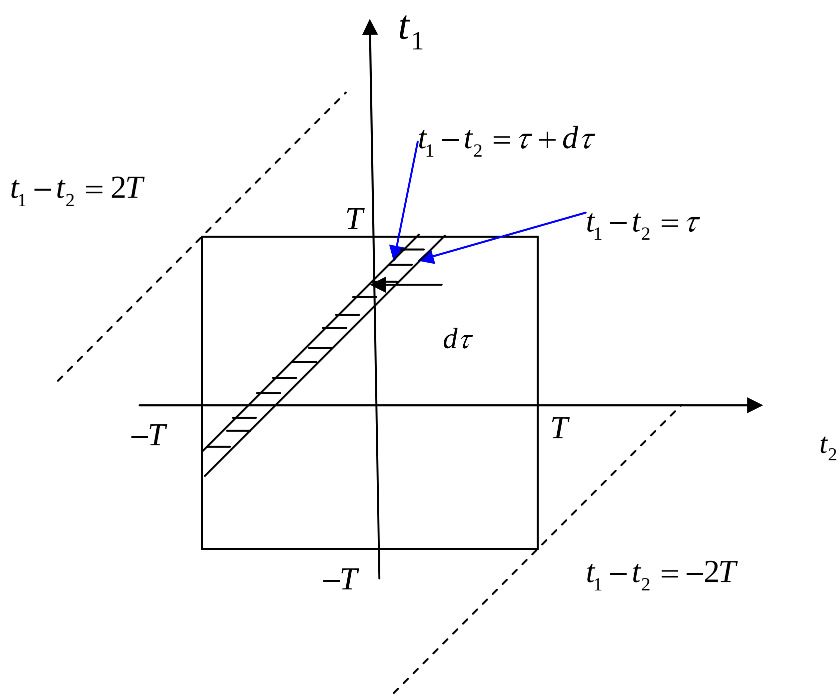
$$= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} E(X(t_1) - \mu_X)(X(t_2) - \mu_X)dt_1dt_2$$

$$= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C_X(t_1 - t_2)dt_1dt_2$$

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 trapizoida strips parallel to $t_1 - t_2 = 0.$ 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:

Therefore,

$$
\begin{aligned}
E\left(\langle \mu_X \rangle_T - \mu_X\right)^2 &= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C_X(t_1 - t_2)dt_1 dt_2 \\
&= \frac{1}{4T^2}\int_{-2T}^{2T}(2T - |\tau|)C_X(\tau)d\tau \\
&= \frac{1}{2T}\int_{-2T}^{2T}\left(1 - \frac{|\tau|}{2T}\right)C_X(\tau)d\tau
\end{aligned}
$$



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

$$\lim_{T\to\infty} \operatorname{var}\langle \mu_X \rangle_T = \sigma_X^2$$

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

$$\boxed{\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T} C_X(\tau)\left[1-\frac{|\tau|}{2T}\right]d\tau = 0}$$

If $C_X(\tau)$ decreases to 0 for $\tau > \tau_0$, then the above condition is satisfied.

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 Consider the random binary waveform $\{X(t)\}$ discussed in Eaxmple .The process has the autocovariance function for $|\tau| \le T_p$ given by

$$C_X(\tau) = \begin{cases} 1-\dfrac{|\tau|}{T_p} & |\tau| \le T_p \\ 0 & \text{otherwise} \end{cases}$$

Here

$$\int_{-2T}^{2T} |C_X(\tau)| d\tau = 2\int_0^{2T} |C_X(\tau)| d\tau$$
$$= 2\int_0^{T_p} \left(1 - \frac{\tau}{T_p}\right) d\tau$$
$$= 2\left(T_p + \frac{T_p^3}{3T_p^2} - \frac{T_p^2}{T_p}\right)$$
$$= \frac{2T_p}{3}$$
$$\therefore \int_{-\infty}^{\infty} |C_X(\tau)| d\tau < \infty$$

Hence $\{X(t)\}$ is not mean ergodic.

Autocorrelation ergodicity

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

If 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

$$C_Z(\tau_1) = EZ(t)Z(t-\tau_1) - EZ(t)EZ(t-\tau_1)$$
$$= EX(t)X(t-\tau)X(t-\tau)X(t-\tau-\tau_1) - R_X^2(\tau)$$

Involves fourth order moment.

Hence found the condition for autocorrelation ergodicity of a jointly Gaussian process.

Thus $X(t)$ will be autocorrelation ergodic if

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

Now $C_Z(\tau) = EZ(t)Z(t+\tau) - R_X^2(\tau)$

Hence, $X(t)$ will be autocorrelation ergodic

If $\lim_{T\to\infty} \frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)\left(Ez(t)z(t+\alpha) - R_X^2(\tau)\right)\mathrm{d}\alpha \to 0$

Example

Consider the random –phased sinusoid given by

$X(t) = A\cos(w_0 t + \phi)$ where A and w_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

$R_X(\tau) = \frac{A^2}{2}\cos w_0\tau$

For any particular realization $x(t) = A\cos(w_0 t + \phi_1)$,

$$\begin{aligned}
\langle \mu_x \rangle_T &= \frac{1}{2T}\int_{-T}^{T} A\cos(w_0 t + \phi_1)dt \\
&= \frac{1}{Tw_0} A\sin(w_0 T)
\end{aligned}$$

and

$$\begin{aligned}
\langle R_x(\tau) \rangle_T &= \frac{1}{2T}\int_{-T}^{T} A\cos(w_0 t + \phi_1) A\cos(w_0(t+\tau) + \phi_1)\mathrm{d}t \\
&= \frac{A^2}{4T}\int_{-T}^{T} [\cos w_0\tau + A\cos(w_0(2t+\tau) + 2\phi_1)]\mathrm{d}t \\
&= \frac{A^2\cos w_0\tau}{2} + \frac{A^2\sin(w_0(2T+\tau))}{4w_0T}
\end{aligned}$$

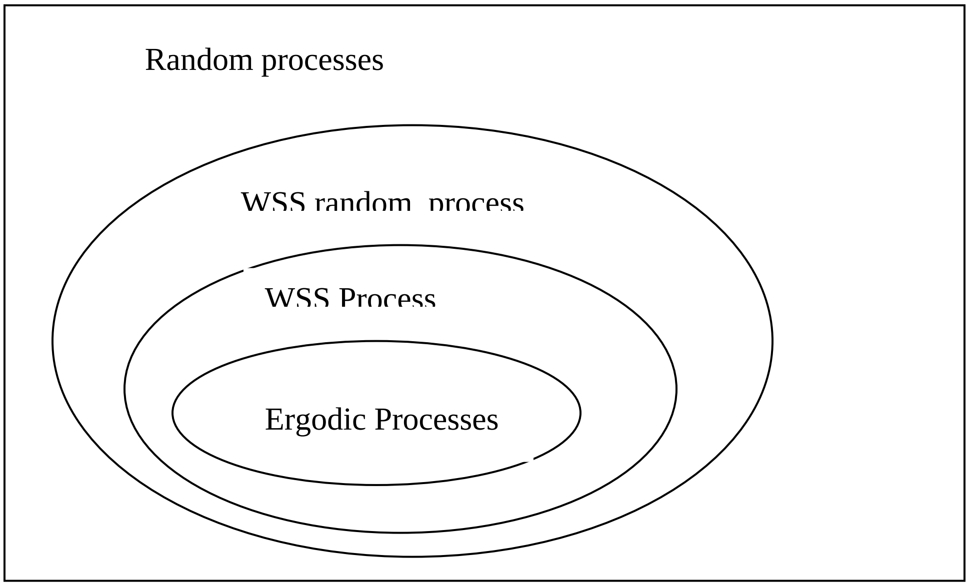
We see that as $T \to \infty$, both $\langle \mu_x \rangle_T \to 0$ and $\langle R_x(\tau) \rangle_T \to \frac{A^2\cos w_0\tau}{2}$

For each realization, both the time-averaged mean and the time-averaged autocorrelation function converges to the corresponding ensemble averages. Thus the random-phased sinusoid is ergodic in both mean and autocorrelation.

Remark

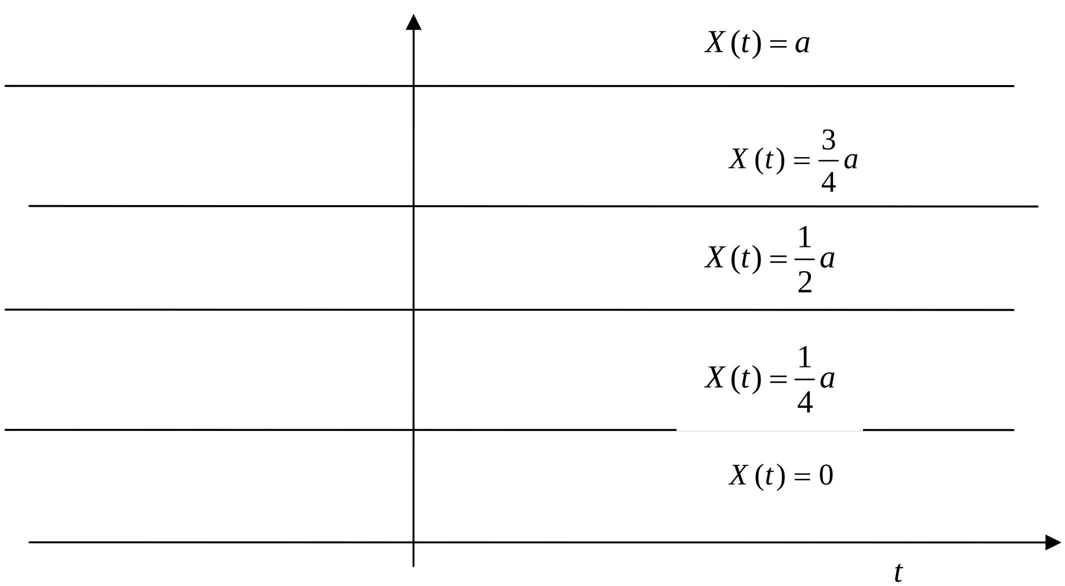
A random process $\{X(t)\}$ is ergodic if its ensemble averages converges in the M.S. sense to the corresponding time averages. This is a stronger requirement than stationarity- the ensemble averages of all orders of such a process are independent of time. This implies that *an ergodic process is necessarily stationary in the strict sense.* The converse is not true- there are stationary random processes which are not ergodic.

Following Fig. shows a hierarchical classification of random processes.



Example

Suppose $X(t) = C$ where $C \sim U[0\ a]$. $\{X(t)\}$ is a family of straight line as illustrated in Fig. below.



Here $\mu_X = \frac{a}{2}$ and

$\langle \mu_X \rangle_T = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} C dt$ is a different constants for different realizations. Hence $\{X(t)\}$ is not mean ergodic.

Spectral Representation of a Wide-sense Stationary Random Process

Recall that the Fourier transform (FT) of a real signal $g(t)$ is given by

$$G(\omega) = FT(g(t)) = \int_{-\infty}^{\infty} g(t)e^{-j\omega t}dt$$

where $e^{-j\omega t} = \cos\omega t + j\sin\omega t$ is the complex exponential.

The Fourier transform $G(\omega)$ exists if $g(t)$ satisfies the following Dirichlet conditions

1) $g(t)$ is absolutely integrable, that is,

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

2) $g(t)$ has only a finite number of discontinuities in any finite interval
3) $g(t)$ has only finite number of maxima and minima within any finite interval.

The signal $g(t)$ can be obtained from $G(\omega)$ by the inverse Fourier transform (IFT) as follows:

$$g(t) = IFT(G(\omega)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)e^{j\omega t}d\omega$$

The existence of the inverse Fourier transform implies that we can represent a function $g(t)$ as a superposition of continuum of complex sinusoids. The Fourier transform $G(\omega)$ is the strength of the sinusoids of frequency ω present in the signal. If $g(t)$ is a voltage signal measured in *volt*, $G(\omega)$ has the unit of *volt/radian.* The function $G(\omega)$ is also called the *spectrum* of $g(t)$.

We can define the Fourier transform also in terms of the frequency variable $f = \frac{\omega}{2\pi}$. In this case, we can define the Fourier transform and the inverse Fourier transform as follows:

$$G(f) = \int_{-\infty}^{\infty} g(t)e^{-j2\pi ft}dt$$

and

$$g(t) = \int_{-\infty}^{\infty} G(f)e^{j2\pi ft}df$$

The Fourier transform is a linear transform and has many interesting properties. Particularly, the energy of the energy of the signal $f(t)$ is related by the Parseval's theorem

$$\int_{-T}^{T} g^2(t)dt = \int_{-\infty}^{\infty} |G(\omega)|^2 d\omega$$

How to have the frequency-domain representation of a random process, particularly a WSS process?

The answer is the spectral representation of WSS random process. Wiener (1930) and Khinchin (1934) independently discovered it. Einstein (1914) also used the concept.

Difficulty in Fourier Representation of a Random Process

We cannot define the Fourier transform of a WSS process $X(t)$ by the mean-square integral

$$FT(X(t)) = \int_{-\infty}^{\infty} X(t)e^{-j\omega t}dt$$

The existence of the above integral would have implied the existence the Fourier transform of every realization of $X(t)$. But the very notion of stationarity demands that the realization does not decay with time and the first condition of Dirichlet is violated.

This difficulty is avoided by a frequency-domain representation of $X(t)$ in terms of the *power spectral density (PSD)*. Recall that the power of a WSS process $X(t)$ is a constant and given by $EX^2(t)$. The PSD denotes the distribution of this power over frequencies.

Definition of Power Spectral Density of a WSS Process

Let us define

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

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

$$\int_{-T}^{T} X_T^2(t)dt = \int_{-\infty}^{\infty} |FTX_T(\omega)|^2 d\omega \quad .$$

Therefore, the power associated with $X_T(t)$ is

$$\frac{1}{2T}\int_{-T}^{T} X_T^2(t)dt = \frac{1}{2T}\int_{-\infty}^{\infty} |FTX_T(\omega)|^2 d\omega \text{ and}$$

The average power is given by

$$\frac{1}{2T} E \int_{-T}^{T} X_T^2(t)dt = \frac{1}{2T} E \int_{-\infty}^{\infty} \left|FTX_T(\omega)\right|^2 d\omega = E \int_{-\infty}^{\infty} \frac{\left|FTX_T(\omega)\right|^2}{2T} d\omega$$

where $\frac{E\left|FTX_T(\omega)\right|^2}{2T}$ is the contribution to the average power at frequency ω and represents the power spectral density for $X_T(t)$. 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

$$S_X(\omega) = \lim_{T\to\infty} \frac{E\left|FTX_T(\omega)\right|^2}{2T}$$

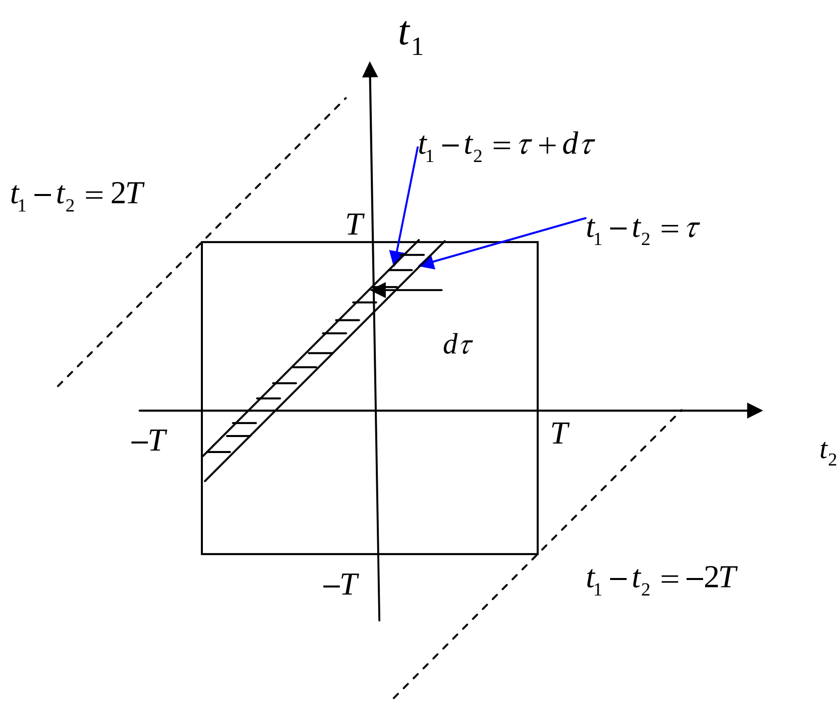
Relation between the autocorrelation function and PSD: Wiener-Khinchin-Einstein theorem

We have

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

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

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



Note that the above integral is to be performed on a square region bounded by $t_1 = \pm T$ and $t_2 = \pm T$. Substitute $t_1 - t_2 = \tau$ so that $t_2 = t_1 - \tau$ is a family of straight lines parallel to $t_1 - t_2 = 0$. The differential area in terms of τ is given by the shaded area and equal to $(2T - |\tau|)d\tau$. The double integral is now replaced by a single integral in τ. Therefore,

$$E\frac{FTX_T(\omega)X_T{}^*(\omega)}{2T} = \frac{1}{2T}\int_{-2T}^{2T} R_X(\tau)e^{-j\omega\tau}(2T - |\tau|)d\tau$$
$$= \int_{-2T}^{2T} R_x(\tau)e^{-j\omega\tau}\left(1 - \frac{|\tau|}{2T}\right)d\tau$$

If $R_X(\tau)$ is integrable then the right hand integral converges to $\int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau$ as $T \to \infty$,

$$\therefore \lim_{T\to\infty}\frac{E\left|FTX_T(\omega)\right|^2}{2T} = \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau$$

As we have noted earlier, the power spectral density $S_X(\omega) = \lim_{T\to\infty}\frac{E\left|FTX_T(\omega)\right|^2}{2T}$ is the contribution to the average power at frequency ω and is called the power spectral density $X(t)$. Thus

$$\boxed{S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau}$$

and using the inverse Fourier transform

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

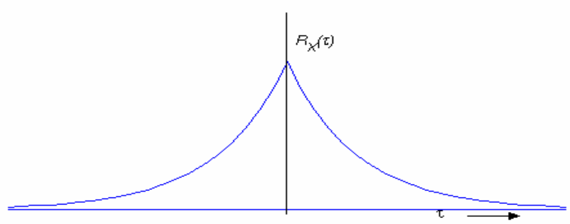
Example 1 The autocorrelation function of a WSS process $X(t)$ is given by

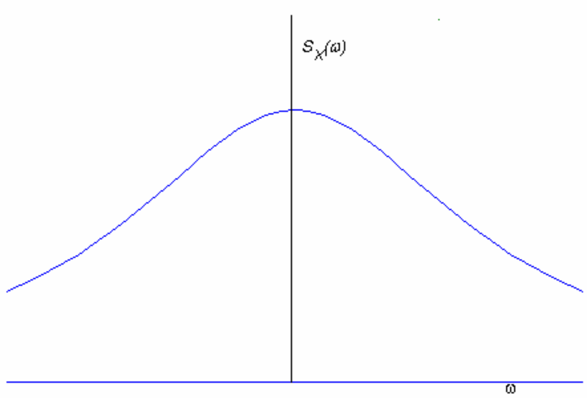
$$R_X(\tau) = a^2 e^{-b|\tau|} \quad b > 0$$

Find the power spectral density of the process.

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

The autocorrelation function and the PSD are shown in Fig.

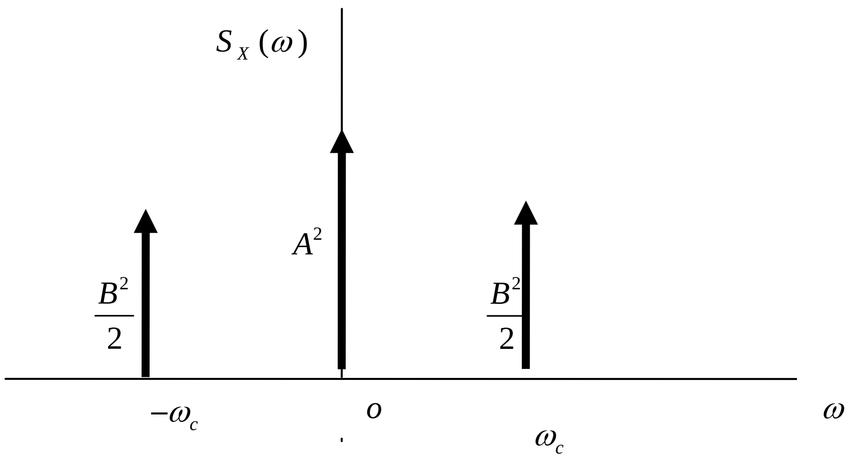




Example: Suppose $X(t) = A + B\sin(\omega_c t + \Phi)$ where A is a constant bias and $\Phi \sim U[0,\ 2\pi]$. Find $R_X(\tau)$ and $S_X(\omega)$..

$$R_x(\tau) = EX(t+\tau)X(t)$$
$$= E(A + B\sin(\omega_c(t+\tau) + \Phi))(A + B\sin(\omega_c t + \Phi))$$
$$= A^2 + \frac{B^2}{2}\cos\omega_c\tau$$
$$\therefore\ S_X(\omega) = A^2\delta(\omega) + \frac{B^2}{4}\left(\delta(\omega+\omega_c) + \delta(\omega-\omega_c)\right)$$

where $\delta(\omega)$ is the Dirac Delta function.



Example 3 PSD of the amplitude-modulated random-phase sinusoid

$$X(t) = M(t)\cos(\omega_c t + \Phi), \qquad \Phi \sim U(0, 2\pi)$$

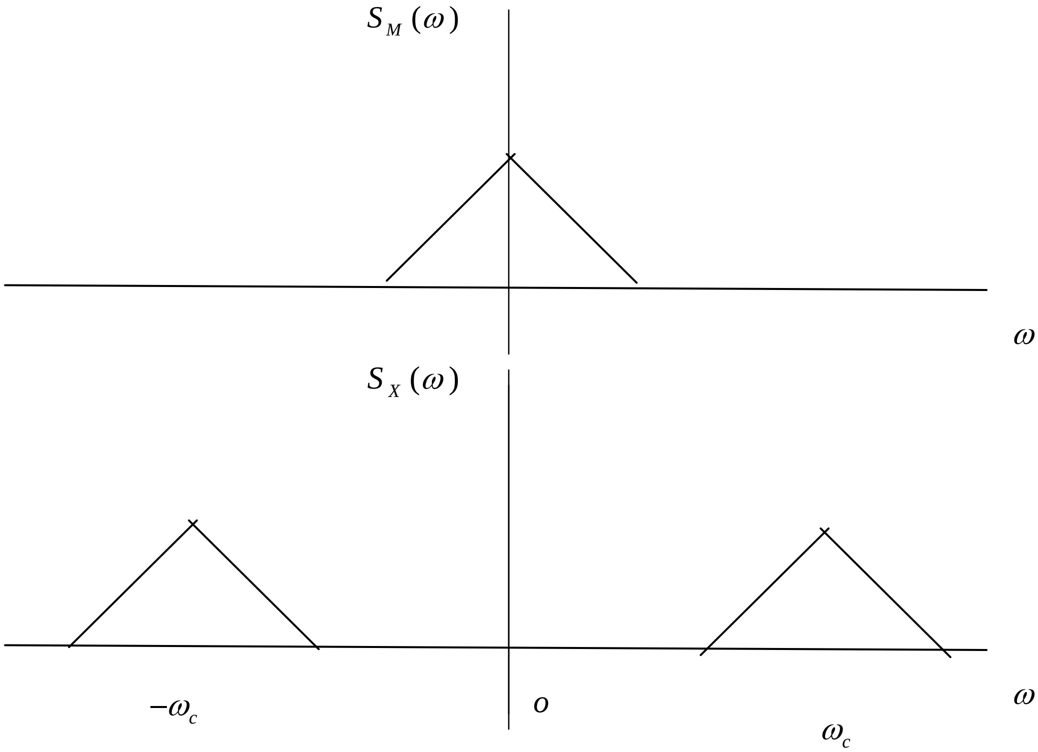
where $M(t)$ is a WSS process independent of Φ.

$$\Longrightarrow \quad R_X(\tau) = E\ M(t+\tau)\ \cos(\omega_c(t+\tau) + \Phi)\ \ M(t)\cos(\omega_c\tau + \Phi)$$
$$= E\ M(t+\tau)M(t)\ E\cos(\omega_c(t+\tau) + \Phi)\cos(\omega_c\tau + \Phi)$$

(Using the independence of $M(t)$ and the sinusoid)

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

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



Example 4 The PSD of a noise process is given by

$$S_N(\omega) = \frac{N_0}{2} \qquad |\omega \pm \omega_c| \le \frac{\omega}{2}$$
$$= 0 \qquad \text{Otherwise}$$

Find the autocorrelation of the process.

$$\therefore\ R_N(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_N(\omega)\ e^{j\omega\tau}\, d\omega$$

$$= \frac{1}{2\pi} \times 2 \times \int_{\omega_c - \frac{\omega}{2}}^{\omega_c + \frac{\omega}{2}} \frac{No}{2} \cos\omega\tau\ d\tau$$

$$= \frac{No}{2\pi} \left[\frac{\sin\left(\omega_c + \frac{\omega}{2}\right) - \sin\left(\omega_c - \frac{\omega}{2}\right)}{\tau} \right]$$

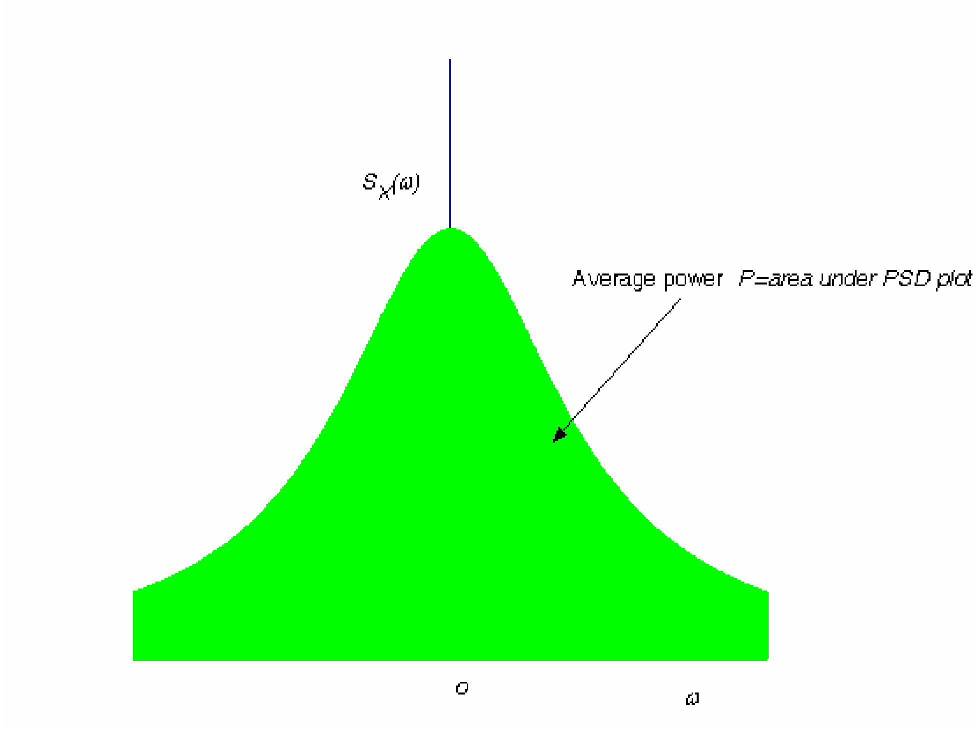
$$= \frac{No}{2\pi}\ \frac{\sin\ \omega\tau/2}{\omega\tau/2}\ \cos\ \omega_o\tau$$

Properties of the PSD

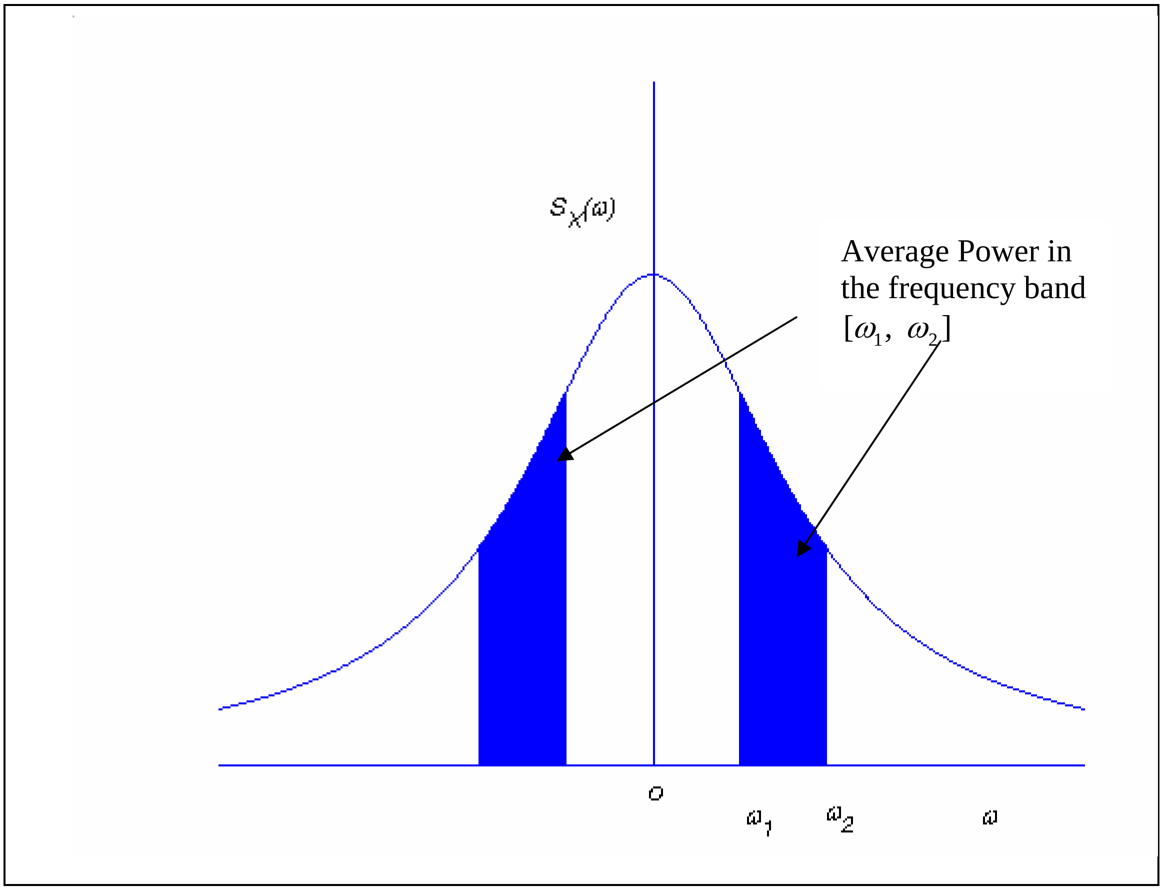
$S_X(\omega)$ being the Fourier transform of $R_X(\tau)$, it shares the properties of the Fourier transform. Here we discuss important properties of $S_X(\omega)$.

The average power of a random process $X(t)$ is

$$\begin{aligned} EX^2(t) &= \mathrm{R}_X(0) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)dw \end{aligned}$$



The average power in the band $[\omega_1, \omega_2]$ is $2\int_{\omega_1}^{\omega_2} S_X(\omega)d\omega$



- 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 $S_X(\omega)$ is a real and even function of ω.

- From the definition $S_X(w) = \lim_{T\to\infty} \frac{E\left|X_T(\omega)\right|^2}{2T}$ is always non-negative. Thus $S_X(\omega) \geq 0$.
- If $\{X(t)\}$ has a periodic component, $\mathrm{R}_X(\tau)$ is periodic and so $S_X(\omega)$ will have impulses.

Remark

1) The function $S_X(\omega)$ is the PSD of a WSS process $\{X(t)\}$ if and only if $S_X(\omega)$ is a non-negative, real and even function of ω and $\int_{-\infty}^{\infty} S_X(\omega) < \infty$
2) The above condition on $S_X(\omega)$ also ensures that the corresponding autocorrelation function $\mathrm{R}_X(\tau)$ is non-negative definite. Thus the non-negative definite property of an autocorrelation function can be tested through its power spectrum.
3) Recall that a periodic function has the *Fourier series* expansion. If $\{X(t)\}$ is M.S. periodic we can have an equivalent *Fourier series* expansion $\{X(t)\}$.

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

$$\boxed{S_{XY}(\omega) = \lim_{T\to\infty} E\,\frac{FTX_T^*(\omega)FTY_T(\omega)}{2T}}$$

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

$$\boxed{S_{YX}(\omega) = \lim_{T\to\infty} E\,\frac{FTY_T^*(\omega)FTX_T(\omega)}{2T}}$$

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 ω. Some properties of the CPSD of two jointly WSS processes *X(t)* and *Y(t)* are listed below:

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

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

$$\begin{aligned} \therefore S_{XY}(\omega) &= \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} R_{YX}(-\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} R_{YX}(\tau) e^{j\omega\tau} d\tau \\ &= S_{YX}^*(\omega) \end{aligned}$$

(2) $\operatorname{Re}(S_{XY}(\omega))$ is an even function of ω and $\operatorname{Im}(S_{XY}(\omega))$ is an odd function of ω

We have

$$\begin{aligned} S_{XY}(\omega) &= \int_{-\infty}^{\infty} R_{XY}(\tau)(\cos\omega\tau + j\sin\omega\tau) d\tau \\ &= \int_{-\infty}^{\infty} R_{XY}(\tau)\cos\omega\tau d\tau + j\int_{-\infty}^{\infty} R_{XY}(\tau)\sin\omega\tau) d\tau \\ &= \operatorname{Re}(S_{XY}(\omega)) + j\operatorname{Im}(S_{XY}(\omega)) \end{aligned}$$

where

$\operatorname{Re}(S_{XY}(\omega)) = \int_{-\infty}^{\infty} R_{XY}(\tau)\cos\omega\tau d\tau$ is an even function of ω and

$\operatorname{Im}(S_{XY}(\omega)) = \int_{-\infty}^{\infty} R_{XY}(\tau)\sin\omega\tau d\tau$ is an odd function of ω and

(3) *X(t)* and *Y(t)* are uncorrelated and have constant means, then

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

Observe that

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

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

(4) If *X(t)* and *Y(t)* are orthogonal, then

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

If *X(t)* and *Y(t)* are orthogonal,

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

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

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

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

Applying Parseval's theorem, we get

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

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

Similarly,

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

Example Consider the random process $Z(t) = X(t) + Y(t)$ discussed in the beginning of the lecture. Here $Z(t)$ is the sum of two jointly WSS orthogonal random processes *X(t)* and *Y(t)*.

We have,

$R_Z(\tau) = R_X(\tau) + R_Y(\tau) + R_{XY}(\tau) + R_{YX}(\tau)$

Taking the Fourier transform of both sides,

$$S_Z(\omega) = S_X(\omega) + S_Y(\omega) + S_{XY}(\omega) + S_{YX}(\omega)$$

$$\therefore \frac{1}{2\pi}\int_{-\infty}^{\infty} S_Z(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} S_Y(\omega)d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XY}(\omega)d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{YX}(\omega)d\omega$$

Therefore,

$$\boxed{P_Z(\omega) = P_X(\omega) + P_Y(\omega) + P_{XY}(\omega) + P_{YX}(\omega)}$$

Remark

- $P_{XY}(\omega) + P_{YX}(\omega)$ is the additional power contributed by $X(t)$ and $Y(t)$ to the resulting power of $X(t) + Y(t)$
- If *X(t)* and *Y(t)* are orthogonal, then

$$\begin{aligned} S_Z(\omega) &= S_X(\omega) + S_Y(\omega) + 0 + 0 \\ &= S_X(\omega) + S_Y(\omega) \end{aligned}$$

Consequently

$$P_Z(\omega) = P_X(\omega) + P_Y(\omega)$$

Thus in the case of two jointly WSS orthogonal processes, the power of the sum of the processes is equal to the sum of respective powers.

Power spectral density of a discrete-time WSS random process

Suppose $g[n]$ is a discrete-time real signal. Assume $g[n]$ to be obtained by sampling a continuous-time signal $g(t)$ at an uniform interval T such that

$$g[n] = g(nT), \qquad n = 0, \pm 1, \pm 2, \ldots$$

The *discrete-time Fourier transform (DTFT)* of the signal $g[n]$ is defined by

$$G(\omega) = \sum_{n=-\infty}^{\infty} g[n]e^{-j\omega n}$$

$G(\omega)$ exists if $\{g[n]\}$ is absolutely summable, that is, $\sum_{n=-\infty}^{\infty} |g[n]| < \infty$. The signal $g[n]$ is obtained from $G(\omega)$ by the *inverse discrete-time Fourier transform*

$$g[n]) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\omega) e^{j\omega n} dw$$

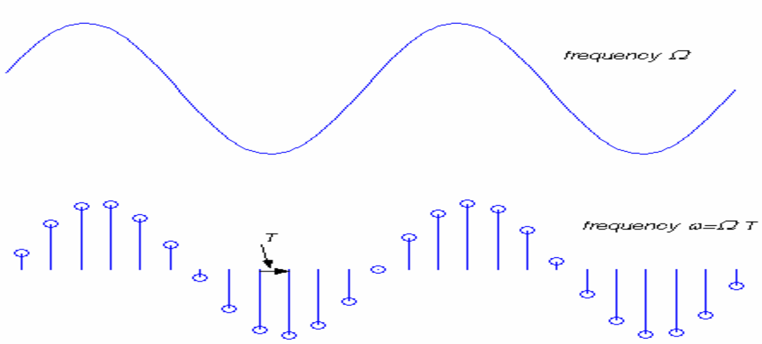
Following observations about the DTFT are important:

- ω is a frequency variable representing the frequency of a discrete sinusoid. Thus the signal $g[n] = A\cos(\omega_0 n)$ has a frequency of ω_0 radian/samples.
- $G(\omega)$ is always periodic in ω with a period of 2π. Thus $G(\omega)$ is uniquely defined in the interval $-\pi \le \omega \le \pi$.
- Suppose $\{g[n]\}$ is obtained by sampling a continuous-time signal $g_a(t)$ at a uniform interval T such that

  $g[n] = g_a(nT), \quad n = 0, \pm 1, \pm 2, ...$

  The frequency ω of the discrete-time signal is related to the frequency Ω of the continuous time signal by the relation $\Omega = \frac{\omega}{T}$

  where T is the uniform sampling interval. The symbol Ω for frequency of a continuous signal is used in the signal-processing literature just to distinguish it from the corresponding frequency of the discrete-time signal. This is illustrated in the Fig. below.



- We can define the $Z-transform$ of the discrete-time signal by the relation

$$G(z) = \sum_{n=-\infty}^{\infty} g[n]z^{-n}$$

where z is a complex variable. $G(\omega)$ is related to $G(z)$ by

$$G(\omega) = G(z)\big|_{z=e^{j\omega}}$$

Power spectrum of a discrete-time real WSS process $\{X[n]\}$

Consider a discrete-time real WSS process $\{X[n]\}$. The very notion of stationarity poses problem in frequency-domain representation of $\{X[n]\}$ through the Discrete-time Fourier transform. The difficulty is avoided similar to the case of the continuous-time WSS process by defining the truncated process

$$X_N[n] = \begin{cases} X[n] & \text{for } |n| \le N \\ 0 & \text{otherwise} \end{cases}$$

The power spectral density $S_X(\omega)$ of the process $\{X[n]\}$ is defined as

$$S_X(\omega) = \lim_{N\to\infty} \frac{1}{2N+1} E\left|DTFTX_N(\omega)\right|^2$$

where

$$DTFTX_N(\omega) = \sum_{n=-\infty}^{\infty} X_N[n]e^{-jwn} = \sum_{n=-N}^{N} X[n]e^{-jwn}$$

Note that the average power of $\{X[n]\}$ is $R_X[0] = E X^2[n]$ and the power spectral density $S_X(\omega)$ indicates the contribution to the average power of the sinusoidal component of frequency ω.

Wiener-Einstein-Khinchin theorem

The Wiener-Einstein-Khinchin 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

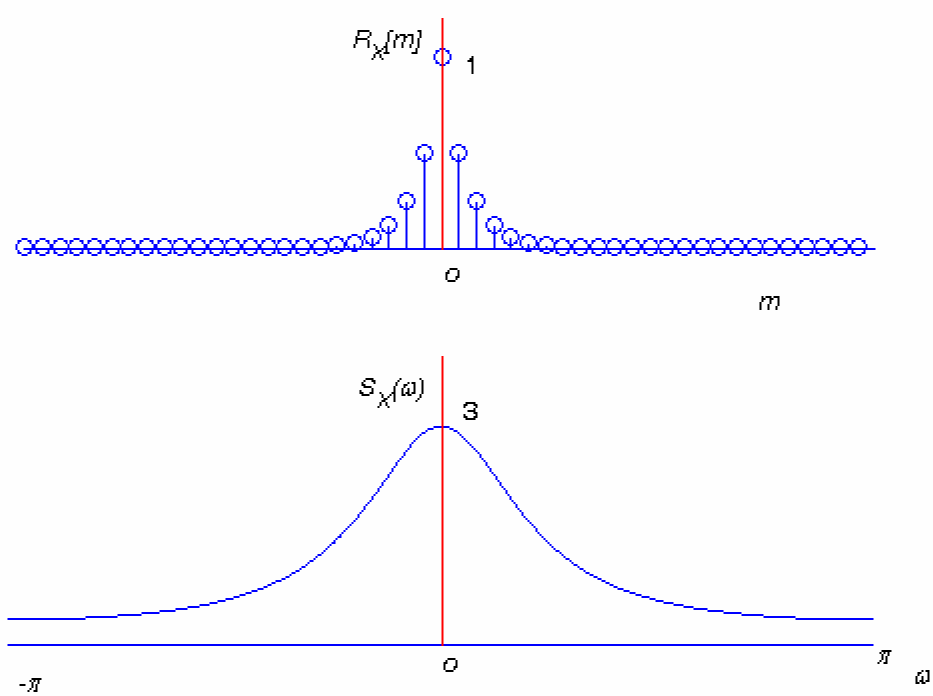
$$S_X(z)=\sum_{m=-\infty}^{\infty} R_x[m]z^{-m}$$

Clearly, $S_X(\omega)=S_X(z)\big|_{z=e^{j\omega}}$

Example Suppose $R_X[m]=2^{-|m|}\qquad m=0,\pm1,\pm2,\pm3....$ Then

$$\begin{aligned} S_X(\omega)&=\sum_{m=-\infty}^{\infty} R_X[m]e^{-j\omega m}\\ &=1+\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\left(\frac{1}{2}\right)^{|m|}e^{-j\omega m}\\ &=\frac{3}{5-4\cos\omega} \end{aligned}$$

The plot of the autocorrelation sequence and the power spectral density is shown in Fig. below.



Example

Properties of the PSD of a discrete-time WSS process

- For the real discrete-time process $\{X[n]\}$, the autocorrelation function $R_X[m]$ is real and even. Therefore, $S_X(\omega)$ is real and even.
- $S_X(\omega) \geq 0.$
- The average power of $\{X[n]\}$ is given by

  $$EX^2[n] = R_X[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_X(\omega)d\omega$$

  Similarly the average power in the frequency band $[\omega_1, \ \omega_2]$ is given by

  $$2\int_{\omega_1}^{\omega_2} S_X(\omega)d\omega$$

- $S_X(\omega)$ is periodic in ω with a period of 2π.

Interpretation of the power spectrum of a discrete-time WSS process

Assume that the discrete-time WSS process $\{X[n]\}$ is obtained by sampling a continuous-time random process $\{X_a(t)\}$ at an uniform interval, that is,

$$X[n] = X_a(nT), \qquad n = 0, \pm 1, \pm 2, \ldots$$

The autocorrelation function $R_X[m]$ is defined by

$$\begin{aligned} R_X[m] &= E\, X[n+m]\, X[n] \\ &= E\, X_a(nT + mT)\, X_a(nT) \\ &= R_{X_a}(mT) \end{aligned}$$

$$\therefore R_X[m] = R_{X_a}(mT) \qquad m = 0, \pm 1, \pm 2, \ldots$$

Thus the sequence $\{R_X[m]\}$ is obtained by sampling the autocorrelation function $R_{X_a}(\tau)$ at a uniform interval T.

The frequency ω of the discrete-time WSS process is related to the frequency Ω of the continuous time process by the relation $\Omega = \dfrac{\omega}{T}$

White noise process

A white noise process $\{W(t)\}$ is defined by

$$S_W(\omega) = \frac{N_0}{2} \qquad -\infty < \omega < \infty$$

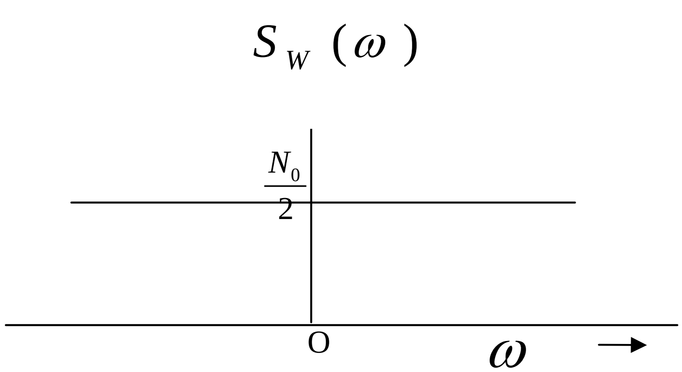
where N_0 is a real constant and called the *intensity* of the white noise. The corresponding autocorrelation function is given by

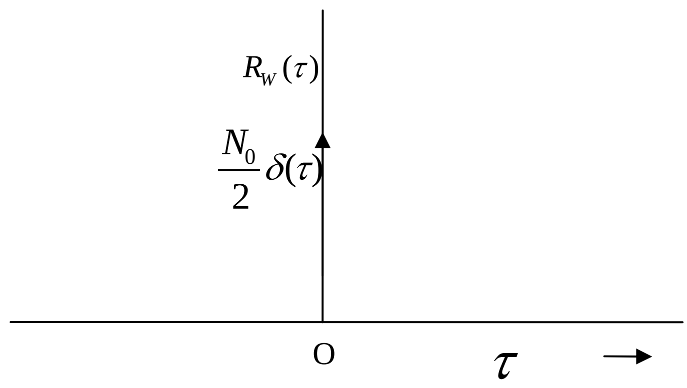
$$R_W(\tau) = \frac{N}{2}\delta(\tau)$$ where $\delta(\tau)$ is the Dirac delta.

The average power of white noise

$$P_{avg} = EW^2(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{N}{2}d\omega \rightarrow \infty$$

The autocorrelation function and the PSD of a white noise process is shown in Fig. below.





Remarks

- The term *white noise* is analogous to *white light* which contains all visible light frequencies.
- We generally consider *zero-mean* white noise process.
- A white noise process is unpredictable as the noise samples at different instants of time are uncorrelated.:

$$C_W(t_i, t_j) = 0 \quad \text{for} \quad t_i \neq t_j.$$

- White noise is a mathematical abstraction, it cannot be realized since it has infinite average power.
- If the system band-width (BW) is sufficiently narrower than the noise BW and noise PSD is flat , we can model it as a white noise process. Thermal noise, which is the noise generated in resistors due to random motion electrons, is well modelled as white Gaussian noise, since they have very flat psd over very wide band (GHzs)

- A white noise process can have any probability density function. Particularly, if the white noise process $\{W(t)\}$ is a Gaussian random process, then $\{W(t)\}$ is called a white Gaussian random process.
- A white noise process is called *strict-sense white noise process* if the noise samples at distinct instants of time are independent. A white Gaussian noise process is a strict-sense white noise process. Such a process represents a 'purely' random process, because its samples at arbitrarily close intervals also will be independent.

Example A random-phase sinusoid corrupted by white noise

Suppose $X(t) = B\sin(\omega_c t + \Phi) + W(t)$ where A is a constant bias and $\Phi \sim U[0,\ 2\pi]$**.** and $\{W(t)\}$ is a zero-mean WGN process with PSD of $\frac{N_0}{2}$ and independent of Φ. Find $R_X(\tau)$ and $S_X(\omega)$..

$$
\begin{aligned}
R_x(\tau) &= EX(t+\tau)X(t) \\
&= E(B\sin(\omega_c(t+\tau) + \Phi) + W(t+\tau))(W(t) + B\sin(\omega_c t + \Phi)) \\
&= \frac{B^2}{2}\cos\omega_c\tau + R_W(\tau)
\end{aligned}
$$

$$
\therefore\ S_X(\omega) = \frac{B^2}{4}\left(\delta(\omega+\omega_c) + \delta(\omega-\omega_c)\right) + \frac{N_0}{2}
$$

where $\delta(\omega)$ is the Dirac Delta function.

Band-limited white noise

A noise process which has non-zero constant PSD over a finite frequency band and zero elsewhere is called band-limited white noise. Thus the WSS process $\{X(t)\}$ is band-limited white noise if

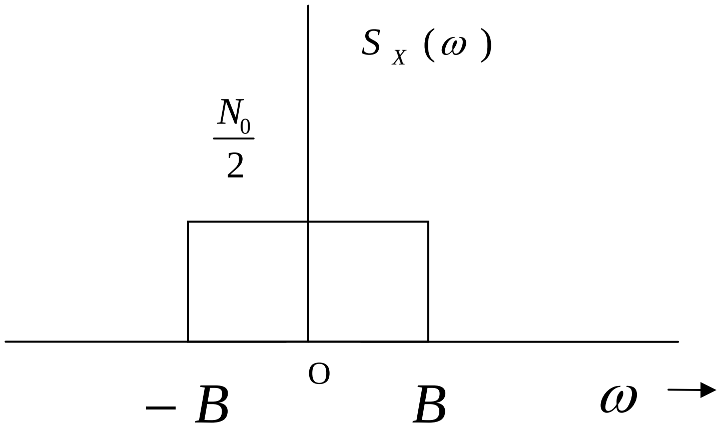
$$
S_X(\omega) = \frac{N_0}{2} \qquad\qquad -B < \omega < B
$$

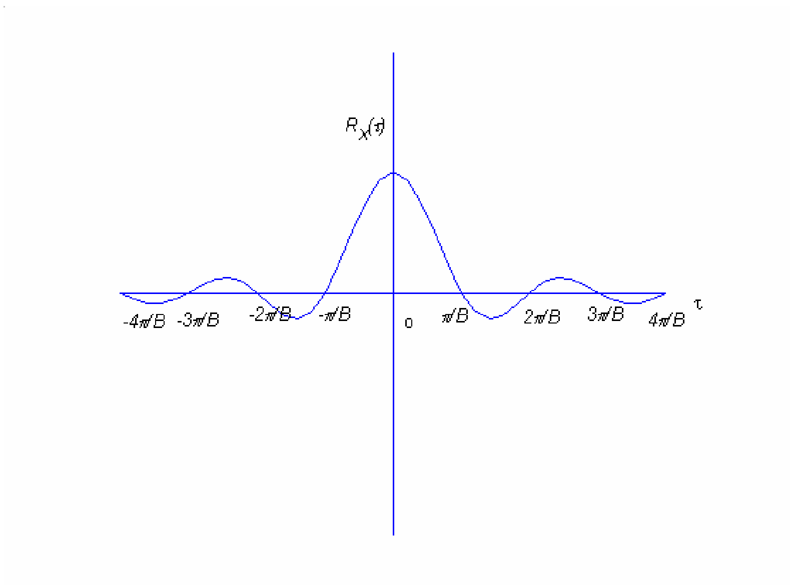
For example, thermal noise which has constant PSD up to very high frequency is better modeled by a band-limited white noise process.

The corresponding autocorrelation function $R_X(\tau)$ is given by

$$
R_X(\tau) = \frac{N_0 B}{2\pi}\frac{\sin B\tau}{B\tau}
$$

The plot of $S_X(\omega)$ and $R_X(\tau)$ of a band-limited white noise process is shown in Fig. below. Further assume that $\{X(t)\}$ is a zero-mean process.





Observe that

- The average power of the process is $EX^2(t) = R_X(0) = \dfrac{N_0 B}{2\pi}$
- $R_X(\tau) = 0$ for $\tau = \pm\dfrac{\pi}{B}, \pm\dfrac{2\pi}{B}, \pm\dfrac{3\pi}{B}, \ldots$ This means that $X(t)$ and $X(t+\dfrac{n\pi}{B})$ where n is a non-zero integer are uncorrelated. Thus we can get uncorrelated

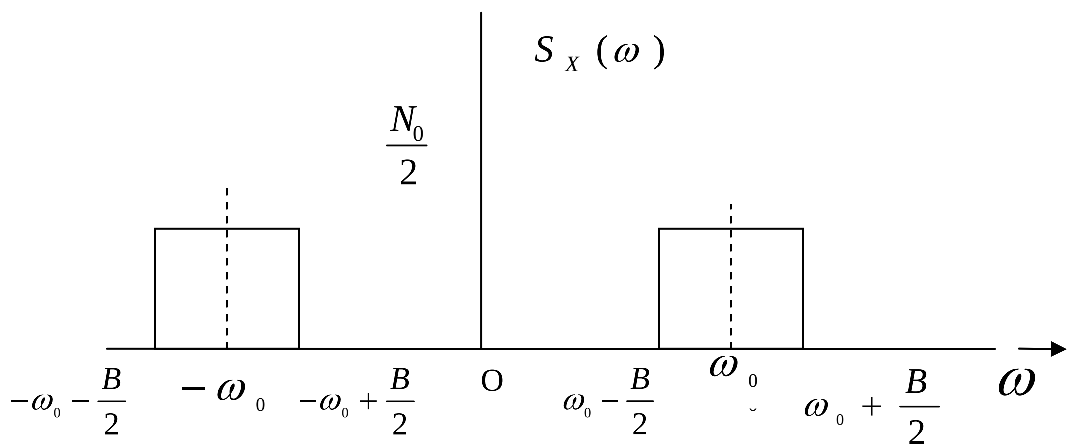
samples by sampling a band-limited white noise process at a uniform interval of $\frac{\pi}{B}$.

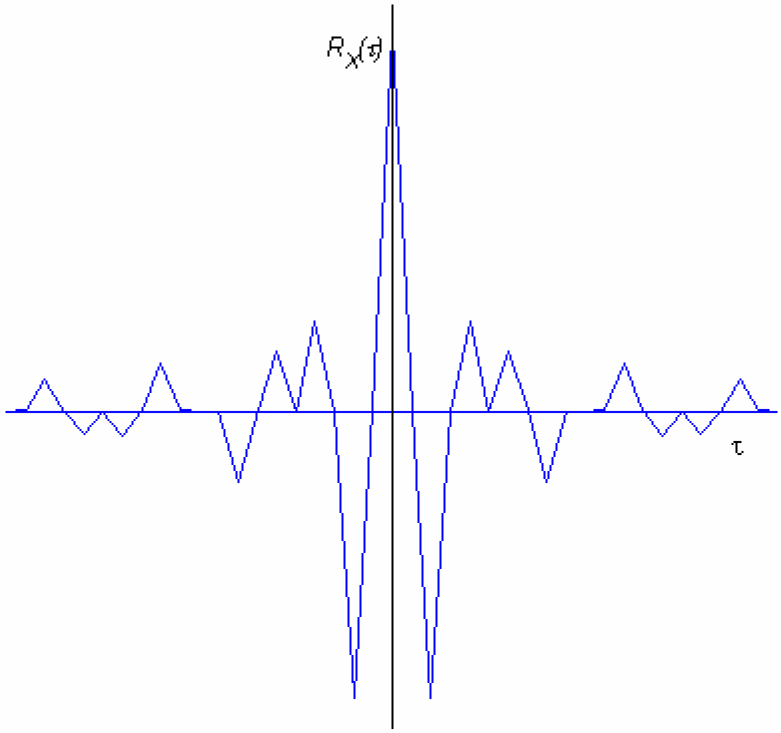
- A band-limited white noise process may also be a band-pass process with PSD as shown in the Fig. and given by

$$S_X(\omega) = \begin{cases} \frac{N_0}{2} & |\omega - \omega_0| < \frac{B}{2} \\ 0 & \text{otherwise} \end{cases}$$

The corresponding autocorrelation function is given by

$$R_X(\tau) = \frac{N_0 B}{2\pi} \frac{\sin \frac{B\tau}{2}}{\frac{B\tau}{2}} \cos \omega_0 \tau$$





Coloured Noise

A noise process which is not white is called *coloured noise*. Thus the noise process $\{X(t)\}$ with $R_X(\tau) = a^2 e^{-b|\tau|}$ $\quad b > 0$ and PSD $S_X(\omega) = \dfrac{2a^2 b}{b^2 + \omega^2}$ is an example of a coloured noise.

White Noise Sequence

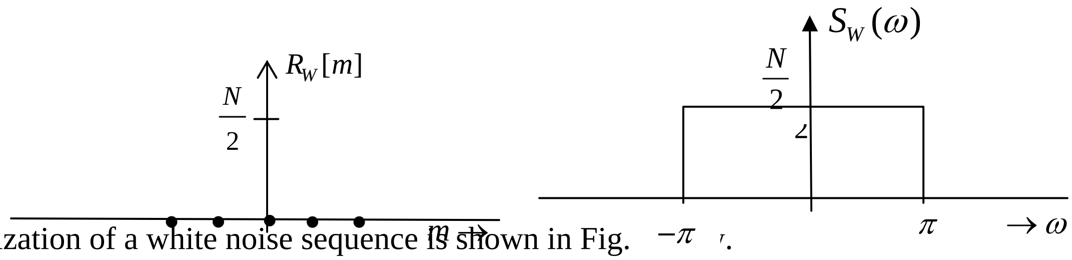
A random sequence $\{W[n]\}$ is called a white noise sequence if

$$S_W(\omega) = \frac{N}{2} \qquad\qquad -\pi \le \omega \le \pi$$

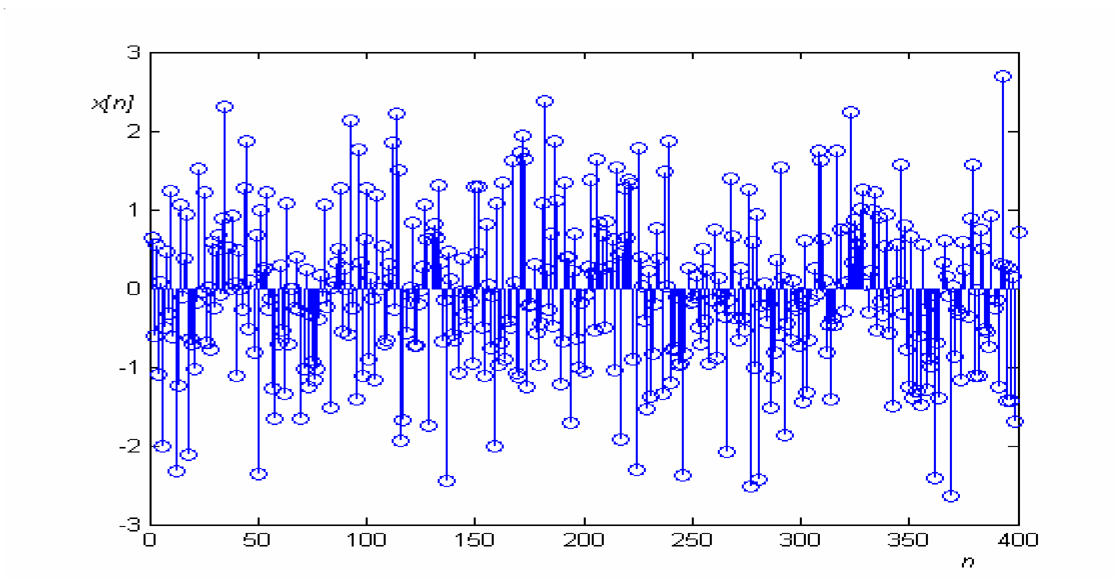
Therefore

$$R_W(m) = \frac{N}{2}\delta(m)$$

where $\delta(m)$ is the unit impulse sequence. The autocorrelation function and the PSD of a whitenoise sequence is shown in Fig.



A realization of a white noise sequence is shown in Fig.



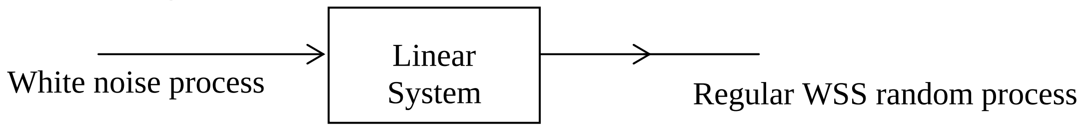
Remark

- The average power of the white noise sequence is $EX^2[n] = \frac{1}{2\pi}\frac{N}{2} \times 2\pi = \frac{N}{2}$

  The average power of the white noise sequence is finite and uniformly distributed over all frequencies.
- If the white noise sequence $\{W[n]\}$ is a Gaussian sequence, then $\{W[n]\}$ is called a *white Gaussian noise* (WGN) sequence.
- An i.i.d. random sequence is always white. Such a sequence may be called strict-sense white noise sequence. A WGN sequence is is a strict-sense stationary white noise sequence.
- The model white noise sequence looks artificial, but it plays a key role in random signal modelling. It plays the similar role as that of the impulse function in the modeling of deterministic signals. A class of WSS processes

called regular processes can be considered as the output of a linear system with white noise as input as illustrated in Fig.

- The notion of the sequence of i.i.d. random variables is also important in *statistical inference*.



Response of Linear time-invariant system to WSS input:

In many applications, physical systems are modeled as linear time invariant (LTI) system. The dynamic behavior of an LTI system to deterministic inputs is described by linear differential equations. We are familiar with time and transfer domain (such as Laplace transform and Fourier transform) techniques to solve these equations. In this lecture, we develop the technique to analyse the response of an LTI system to WSS random process.

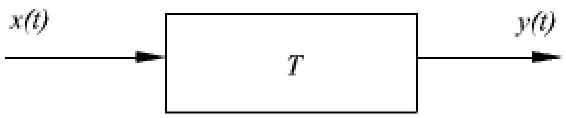
The purpose of this study is two-folds:

(1) Analysis of the response of a system
(2) Finding a LTI system that can optionally estimate an unobserved random process from an observed process. The observed random process is statistically related to the unobserved random process. For example, we may have to find LTI system (also called a filter) to estimate the signal from the noisy observations.

Basics of Linear Time Invariant Systems:

A system is modeled by a transformation T that maps an input signal $x(t)$ to an output signal y(t). We can thus write,

$$y(t) = T\left[x(t)\right]$$



Linear system
The system is called linear if superposition applies: **the weighted sum of inputs results in the weighted sum of the corresponding outputs**. Thus for a linear system

$$T\left[a_1 x_1(t) + a_2 x_2(t)\right] = a_1 T\left[x_1(t)\right] + a_2 T\left[x_2(t)\right]$$

Example : Consider the output of a linear differentiator, given by

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

Then, $\frac{d}{dt}\left(a_1 x_1(t) + a_2 x_2(t) \right)$

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

Hence the linear differentiator is a linear system.

Linear time-invariant system

Consider a linear system with $y(t) = T\,x(t)$. The system is called time-invariant if

$$T\,x\left(t-t_0\right) = y\left(t-t_0\right) \qquad \forall \quad t_0$$

It is easy to check that that the differentiator in the above example is a linear time-invariant system.

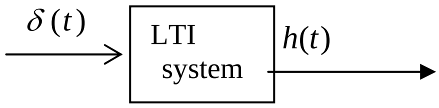
Causal system

The system is called causal if the output of the system at $t = t_0$ depends only on the present and past values of input. Thus for a causal system

$$y\left(t_0\right) = T\left(x(t), t \le t_0\right)$$

Response of a linear time-invariant system to deterministic input

A linear system can be characterised its impulse response $h(t) = T\delta(t)$ where $\delta(t)$ is the *Dirac delta function.*



Recall that any function $x(t)$ can be represented in terms of the Dirac delta function as follows

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\,\delta\left(t-s\right)\,ds$$

If $x(t)$ is input to the linear system $y(t) = T\,x(t)$, then

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

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

If the system is time invariant,

$$h\left(t,s\right) = h\left(t-s\right)$$

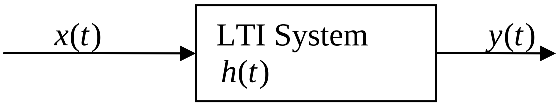
Therefore for a linear invariant system,

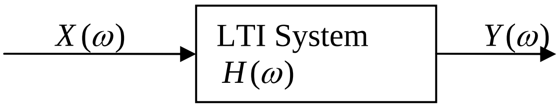
$$y(t) = \int_{-\infty}^{\infty} x\left(s\right)\,h\left(t-s\right)ds = x(t) * h(t)$$

where * denotes the convolution operation.

Thus for a LTI System,

$$\boxed{y(t) = x(t) * h(t) = h(t) * x(t)}$$





Taking the Fourier transform, we get

$$Y(\omega) = H(\omega)\ X(\omega)$$

where $\quad H(\omega) = FT\, h(t) = \int_{-\infty}^{\infty} h(t)\ e^{-j\omega t}\ dt \quad$ is the frequency response of the system

Response of an LTI System to WSS input

Consider an LTI system with impulse response $h(t)$. Suppose $\{X(t)\}$ is a WSS process input to the system. The output $\{Y(t)\}$ of the system is given by

$$Y(t) = \int_{-\infty}^{\infty} h(s)\ X(t-s)\ ds = \int_{-\infty}^{\infty} h(t-s)\ X(s)\ \ ds$$

where we have assumed that the integrals exist in the mean square (m.s.) sense.

Mean and autocorrelation of the output process $\{Y(t)\}$

The mean of the output process is given by,

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

where $H(0)$ is the frequency response $H(\omega)$ at 0 frequency ($\omega = 0$) given by

$$H(\omega)\Big|_{\omega=0} = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}dt\Big|_{\omega=0} = \int_{-\infty}^{\infty} h(t)dt$$

$\boxed{\text{Therefore, the mean of the output process } \{Y(t)\} \text{ is a constant}}$

The Cross correlation of the input *X(t)* and the out put *Y(t)* is given by

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

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

Thus we see that $R_{XY}(\tau)$ is a function of lag τ only. Therefore, $X(t)$ and $Y(t)$ are jointly wide-sense stationary.

The autocorrelation function of the output process Y(t) is given by,

$$\begin{aligned}
\therefore\qquad EY(t+\tau)Y(t)) &= E\int_{-\infty}^{\infty} h(s)\;X(t+\tau-s)ds\,Y(t) \\
&= \int_{-\infty}^{\infty} h(s)\;E\;X(t+\tau-s)\;Y(t)\;ds \\
&= \int_{-\infty}^{\infty} h(s)\;R_{XY}(\tau-s)\;ds \\
&= h(\tau) * R_{XY}(\tau) = h(\tau) * h(-\tau) * R_X(\tau)
\end{aligned}$$

Thus the autocorrelation of the output process $\{Y(t)\}$ depends on the time-lag τ, i.e., $EY(t)Y(t+\tau) = R_Y(\tau)$.

Thus

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

The above analysis indicates that for an LTI system with WSS input

(1) the output is WSS and

(2) the input and output are jointly WSS.

The average power of the output process $\{Y(t)\}$ is given by

$$
\begin{aligned}
P_Y &= R_Y(0) \\
&= R_X(0) * h(0) * h(0)
\end{aligned}
$$

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

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

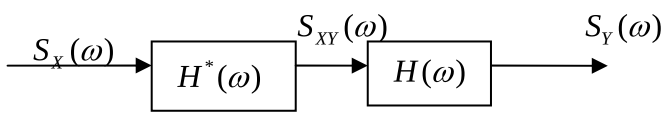
$$\text{and } R_{YX}(\tau) = h(\tau) * R_X(\tau)$$

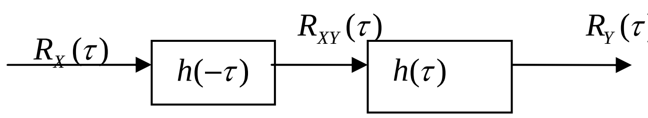
Taking the Fourier transform of $R_{XY}(\tau)$ we get the cross power spectral density $S_{XY}(\omega)$ given by

$$S_{XY}(\omega) = H^*(\omega) S_X(\omega)$$

and

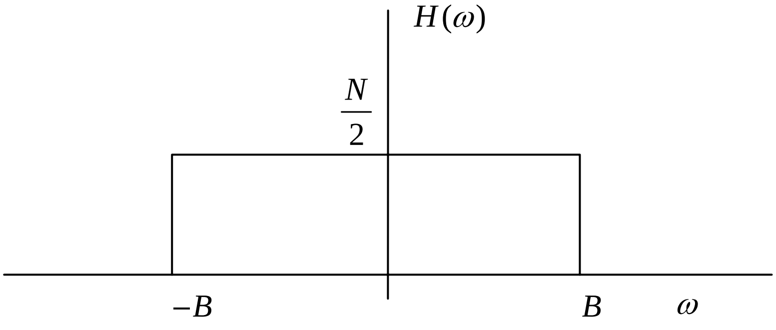
$$S_{YX}(\omega) = H(\omega) S_X(\omega)$$





Example:

(a) White noise process $X(t)$ with power spectral density $\frac{N_0}{2}$ is input to an ideal low pass filter of band-width B. Find the PSD and autocorrelation function of the output process.



The input process $X(t)$ is white noise with power spectral density $S_X(\omega) = \frac{N_0}{2}$.

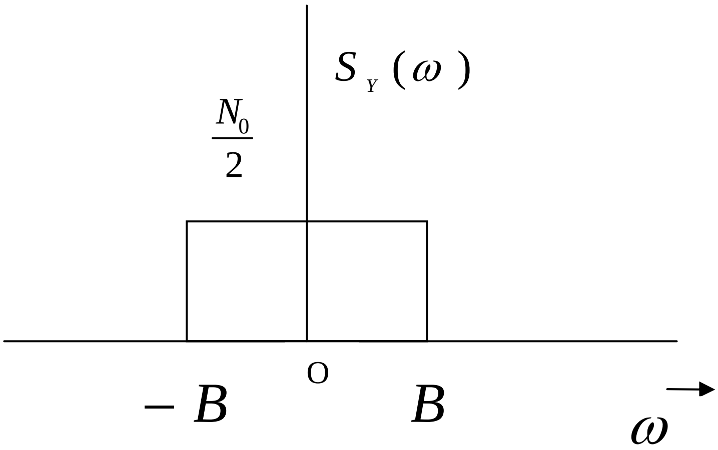
The output power spectral density $S_Y(\omega)$ is given by,

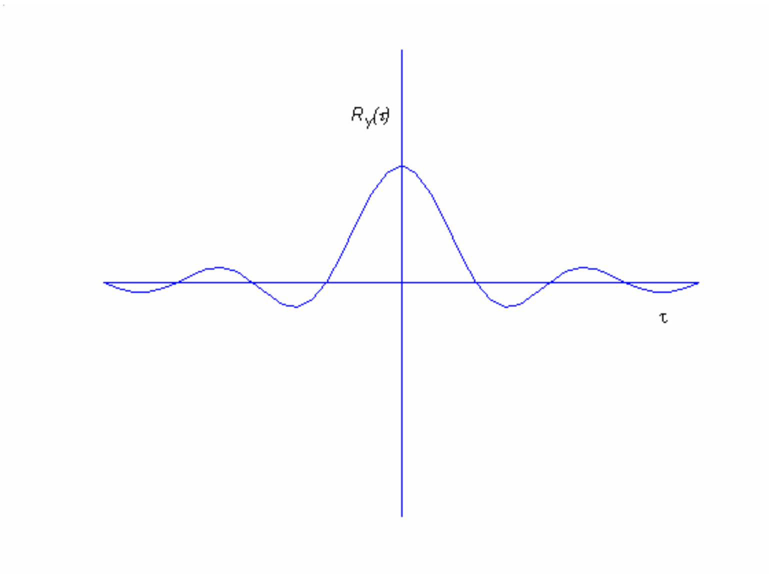
$$\begin{aligned} S_Y(\omega) &= |H(\omega)|^2 S_X(\omega) \\ &= 1 \times \frac{N_0}{2} \qquad -B \leq \omega \leq B \\ &= \frac{N_0}{2} \qquad -B \leq \omega \leq B \end{aligned}$$

$$\therefore R_Y(\tau) = \text{Inverse Fourier transform of } S_Y(\omega)$$

$$= \frac{1}{2\pi} \int_{-B}^{B} \frac{N_0}{2} e^{j\omega\tau} \, d\omega = \frac{N_0}{2} \frac{\sin 2\pi B\tau}{\pi\tau}$$

The output PSD $S_Y(\omega)$ and the output autocorrelation function $R_Y(\tau)$ are illustrated in Fig. below.





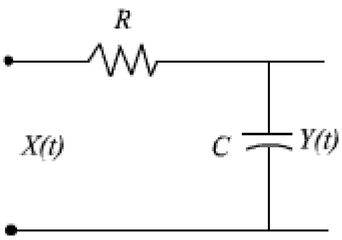
Example 2:

A random voltage modeled by a white noise process $X(t)$ with power spectral density $\frac{N_0}{2}$ is input to a RC network shown in the fig.

Find (a) output PSD $S_Y(\omega)$

(b) output auto correlation function $R_Y(\tau)$

(c) average output power $EY^2(t)R$



The frequency response of the system is given by

$$H(\varpi) = \frac{\frac{1}{jC\omega}}{R + \frac{1}{jC\omega}}$$
$$= \frac{1}{jRC\omega + 1}$$

Therefore,

(a)

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega)$$
$$= \frac{1}{R^2C^2\omega^2 + 1} S_X(\omega)$$
$$= \frac{1}{R^2C^2\omega^2 + 1} \frac{N_0}{2}$$

(b) Taking inverse Fourier transform

$$R_Y(\tau) = \frac{N_0}{4RC} e^{-\frac{|\tau|}{RC}}$$

(c) Average output power

$$EY^2(t) = R_Y(0) = \frac{N_0}{4RC}$$

Discrete-time Linear Shift Invariant System with WSS Random Inputs

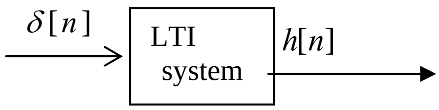
We have seen that the Dirac delta function $\delta(t)$ plays a very important role in the analysis of the response of the continuous-time LTI systems to deterministic and random inputs. Similar role in the case of the discrete-time LTI system is played by the unit sample sequence $\delta[n]$ defined by

$$\delta[n] = \begin{cases} 1 & \text{for } n = 0 \\ 0 & \text{otherwise} \end{cases}$$

Any discrete-time signal $x[n]$ can be expressed in terms of $\delta[n]$ as follows:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k]$$

A discrete-time linear shift-invariant system is characterized by the *unit sample response* $h[n]$ which is the output of the system to the unit sample sequence $\delta[n]$.



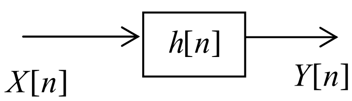
The transfer function of such a system is given by

$$H(\omega) = \sum_{n=-\infty}^{\infty} h[n]e^{-j\omega n}$$

An analysis similar to that for the continuous-time LTI system shows that the response $y[n]$ of a the linear time-invariant system with impulse response $h[n]$ to a deterministic input $x[n]$ is given by

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] = x[n] * h[n]$$

Consider a discrete-time linear system with impulse response $h[n]$ and WSS input $X[n]$



$$Y[n] = X[n] * h[n] = \sum_{k=-\infty}^{\infty} h[k]X[n-k]$$

$$E\, Y[n] = E\, X[n] * h[n]$$

For the WSS input $X[n]$

$$\mu_Y = E\, Y[n] = \mu_X * h[n] = \mu_X \sum_{n=-\infty}^{\infty} h[n] = \mu_X H(0)$$

where $H(0)$ is the dc gain of the system given by

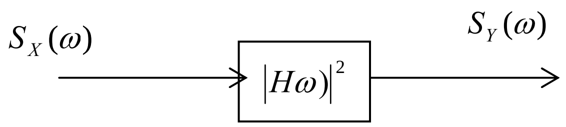
$$
\begin{aligned}
H(0) &= H(\omega)\big|_{\omega=0} \\
&= \sum_{n=-\infty}^{\infty} h[n]e^{-jwn}\Big|_{\omega=0} \\
&= \sum_{n=-\infty}^{\infty} h[n]
\end{aligned}
$$

$$
\begin{aligned}
R_Y[m] &= E\,Y[n]Y[n-m] \\
&= E(X[n]*h[n])(X[n-m]*h[n-m]). \\
&= R_X[m]*h[m]*h[-m]
\end{aligned}
$$

$R_Y[m]$ is a function of lag m only.

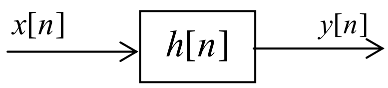
From above we get

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



- Note that though the input is an uncorrelated process, the output is a correlated process.

Consider the case of the discrete-time system with a random sequence $x[n]$ as an input.



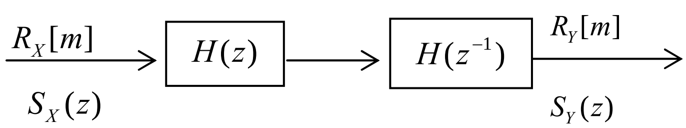
$$R_Y[m] = R_X[m]*h[m]*h[-m]$$

Taking the $z-$transform, we get

$$S_Y(z) = S_X(z)H(z)H(z^{-1})$$

Notice that if $H(z)$ is causal, then $H(z^{-1})$ is anti causal.

Similarly if $H(z)$ is minimum-phase then $H(z^{-1})$ is maximum-phase.



Example

If $H(z) = \dfrac{1}{1-\alpha z^{-1}}$ and $x[n]$ is a unity-variance white-noise sequence, then

Given $EX^2[n] = \sigma_X^2$

$$\therefore S_X(\omega) = \frac{\sigma_X^2}{2\pi}$$

$$S_Y(z) = H(z)H(z^{-1})S_X(z)$$

$$= \left(\frac{1}{1-\alpha z^{-1}}\right)\left(\frac{1}{1-\alpha z}\right)\frac{1}{2\pi}$$

By partial fraction expansion and inverse z – transform, we get

$$R_Y[m] = \frac{1}{1-\alpha^2} a^{|m|}$$

Spectral factorization theorem

A WSS random signal $X[n]$ that satisfies the Paley Wiener condition $\int_{-\pi}^{\pi} |\ln S_X(\omega)|\, d\omega < \infty$ can be considered as an output of a linear filter fed by a white noise sequence.

If $S_X(\omega)$ is an analytic function of ω,

and $\int_{-\pi}^{\pi} |\ln S_X(\omega)|\, d\omega < \infty$, then $S_X(z) = \sigma_v^2 H_c(z) H_a(z)$

where

$H_c(z)$ is the causal minimum phase transfer function

$H_a(z)$ is the anti-causal maximum phase transfer function

and σ_v^2 a constant and interpreted as the variance of a white-noise sequence.

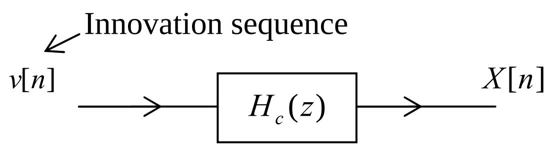


Figure Innovation Filter

Minimum phase filter => the corresponding inverse filter exists.

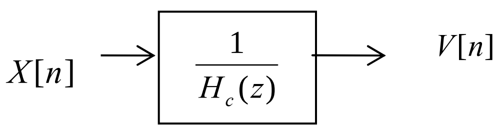


Figure whitening filter

Since $\ln S_X(z)$ is analytic in an annular region $\rho < |z| < \frac{1}{\rho}$,

$$\ln S_X(z) = \sum_{k=-\infty}^{\infty} c[k] z^{-k}$$

where $c[k] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln S_X(\omega) e^{i\omega n} dw$ is the kth order cepstral coefficient.

For a real signal $c[k] = c[-k]$

and $c[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln S_{XX}(w) dw$

$$S_{XX}(z) = e^{\sum_{k=-\infty}^{\infty} c[k] z^{-k}}$$

$$= e^{c[0]} e^{\sum_{k=1}^{\infty} c[k] z^{-k}} e^{\sum_{k=-\infty}^{-1} c[k] z^{-k}}$$

Let $H_C(z) = e^{\sum_{k=1}^{\infty} c[k] z^{-k}} \qquad |z| > \rho$

$$= 1 + h_c(1) z^{-1} + h_c(2) z^{-2} + \ldots\ldots$$

$(\because h_c[0] = Lim_{z\to\infty} H_C(z) = 1$

$H_C(z)$ and $\ln H_C(z)$ are both analytic

=> $H_C(z)$ is a ***minimum phase filter***.

Similarly let

$$H_a(z) = e^{\sum_{k=-\infty}^{-1} c(k) z^{-k}}$$

$$= e^{\sum_{k=1}^{\infty} c(k) z^{k}} = H_C(z^{-1}) \qquad |z| < \frac{1}{\rho}$$

Therefore,

$S_{XX}(z) = \sigma_V^2 H_C(z) H_C(z^{-1})$

where $\sigma_V^2 = e^{c(0)}$

Salient points

- $S_{XX}(z)$ can be factorized into a minimum-phase and a maximum-phase factors i.e. $H_C(z)$ and $H_C(z^{-1})$.
- In general spectral factorization is difficult, however for a signal with rational power spectrum, spectral factorization can be easily done.

- Since is a minimum phase filter, $\frac{1}{H_C(z)}$ exists (=> stable), therefore we can have a filter $\frac{1}{H_C(z)}$ to filter the given signal to get the innovation sequence.
- $X[n]$ and $v[n]$ are related through an invertible transform; so they contain the same information.

Example

Wold's Decomposition

Any WSS signal $X[n]$ can be decomposed as a sum of two mutually orthogonal processes

- a regular process $X_r[n]$ and a predictable process $X_p[n]$, $X[n] = X_r[n] + X_p[n]$
- $X_r[n]$ can be expressed as the output of linear filter using a white noise sequence as input.
- $X_p[n]$ is a predictable process, that is, the process can be predicted from its own past with zero prediction error.