EE 203 PROBABILITY AND RANDOM VARIABLES

COURSE CONTENT

1. Description of Probability
2. Set Theory
3. Probability Space
4. Conditional Probability
5. Repeated Trials (Bernoulli's Theorem)
6. Random Variables
7. Probability Distribution Functions
8. Probability Density Functions
9. Some of the Important Random Variables
10. Conditional Distributions
11. Functions of One Random Variable
12. Mean and Variance
13. Moments
14. Characteristic Functions
15. Two Random Variables
16. Joint Moments
17. Conditional Distributions
18. Sequences of Random Variables
19. Mean Square Estimation
20. Correlation and Covariance
21. Noise (White and Colored)
22. Power Spectrum
23. Limit Theorems
24. Statistics

TEXT BOOK:

1. NAME : Probability, Random Variables and Stochastic Processes
   AUTHOR : Athanasios Papoulis, S. Unnikrishna Pillai
   PUBLISHER : McGraw-Hill
   ISBN : 0-07-112256-7 (ISE)
   EDITION : 2002 (Fourth Edition)
   LIBRARY CODE : QA 273 P218 2002

REFERENCE BOOKS:

1. NAME : Probability and Random Processes for Electrical Engineers
   AUTHOR : Yannis Viniotis
   PUBLISHER : McGraw-Hill
   ISBN : 0-07-067491-4
   EDITION : 1998
   LIBRARY CODE : QA 273 V56 1998

2. NAME : Probability, Random Variables and Stochastic Processes
AUTHOR : Athanasios Papoulis
PUBLISHER : McGraw-Hill
ISBN : 0-07-100870-5
EDITION : 1991
LIBRARY CODE : QA 273 P218 1991

3. NAME : A First Course in Probability
AUTHORS : Sheldon Ross
PUBLISHER : Prentice-Hall, Inc.
ISBN : 0-13-896523-4
EDITION : 1998 (Fifth Edition)
LIBRARY CODE : QA 273 R826 1998

4. NAME : Probability and Random Processes with Applications to Signal Processing
AUTHORS : Sheldon Ross
PUBLISHER : Prentice-Hall, Inc.
ISBN : 0-13-896523-4
EDITION : 1998 (Fifth Edition)
LIBRARY CODE : QA 273 R826 1998

GRADING:

HOMEWORKS:		0 %
QUIZ:		5 x 3% =15 %
ATTENDANCE:		5%
1 MID TERM EXAM (IN CLASS)	:	40 %
1 FINAL EXAM (IN CLASS)	:	40 %
:		
	TOTAL :	100 %

Note: It is essential that students show at least 70 % attendance in lectures.

1. Basics

Probability theory deals with the study of random phenomena

Under repeated experiments, random phenomena yield different outcomes that have certain underlying patterns

An experiment assumes a set of repeatable conditions that allow any number of identical repetitions.

When an experiment is performed under these conditions, certain elementary events ξ_i occur in different but *completely uncertain* ways.

We can assign nonnegative number $P(\xi_i)$, as the probability of the event ξ_i in various ways

Laplace's Classical Definition:

The Probability of an event *A* is defined a-priori without actual experimentation as

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}}$$

provided all these outcomes are *equally likely*.

Consider a box with *n* white and *m* red balls.

In this case, there are two elementary outcomes: white ball or red ball.

Probability of "selecting a white ball" = $n / (n+m)$

Relative Frequency Definition:

The probability of an event *A* is defined as

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

where

n_A is the number of occurrences of *A*

n is the total number of trials.

The totality of all ξ_i *known a priori*, constitutes a set Ω

The set of all experimental outcomes

$\Omega = \{ \xi_1, \xi_2, \ldots , \xi_k, \ldots \}$

Ω has subsets A, B, C, ...

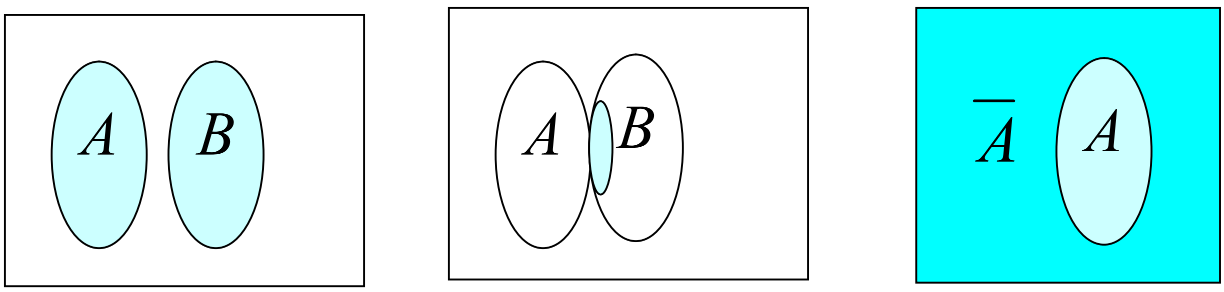
If *A* is a subset of Ω, then $\xi \in A$ implies $\xi \in \Omega$

From *A* and *B*, other related subsets can be generated, like:

$$A \cup B = \{ \xi \mid \xi \in A \quad \text{or} \quad \xi \in B \}$$

$$A \cap B = \{ \xi \mid \xi \in A \quad \text{and} \quad \xi \in B \}$$

$$\overline{A} \quad = \{ \xi \mid \xi \notin A \}$$

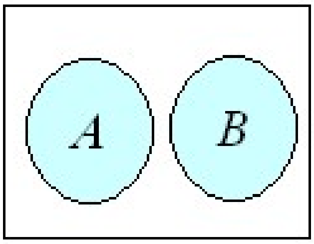


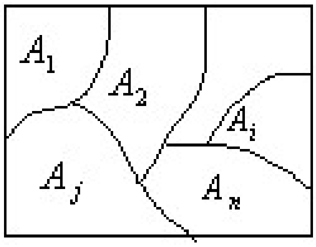
$A \cup B$ $\qquad$ $A \cap B$ $\qquad$ $\overline{A}$

- If $A \cap B = \varnothing$, the empty set, then *A* and *B* are said to be mutually exclusive (M.E).

- A partition of Ω is a collection of mutually exclusive subsets of Ω such that their union is Ω.

$$A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1} A_i = \Omega.$$

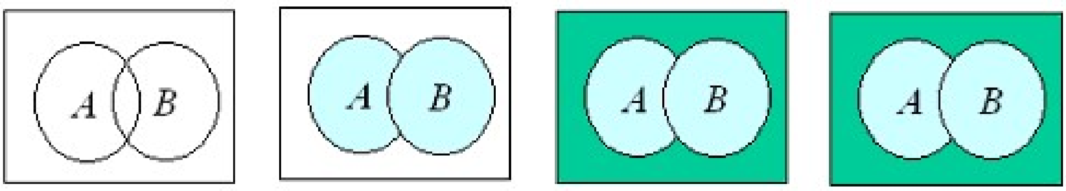




$$A \cap B = \phi$$

<u>De-Morgan's Laws:</u>

$$\overline{A \cup B} = \overline{A} \cap \overline{B}\,; \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}$$



$A \cup B$ $\quad$ $\overline{A \cup B}$ $\quad$ $\overline{A} \cap \overline{B}$

- Event: Some of the subsets of Ω can be considered as events, for which we must have mechanism to compute their probabilities.

Example: Consider the experiment where two coins are simultaneously tossed. The various elementary events are

$$\xi_1 = (H,H), \quad \xi_2 = (H,T), \quad \xi_3 = (T,H), \quad \xi_4 = (T,T)$$

and

$$\Omega = \{\xi_1, \xi_2, \xi_3, \xi_4\}.$$

The subset

$$A = \{\xi_1, \xi_2, \xi_3\}$$

is the same as “Head has occurred at least once” and is also an event.

Suppose two subsets *A* and *B* are both events, then consider

“Does an outcome belong to *A* or *B* = A U B ? ”

“Does an outcome belong to *A* and *B*= A ∩ B ?”

“Does an outcome fall outside *A ?”*

Thus the sets

$$A \cup B \quad , \quad A \cap B \quad , \overline{A} \quad , \quad \overline{B}, \quad \text{etc.}$$

also qualify as events.

Formalize this using the notion of a Field.

- **Field**: A collection of subsets of a nonempty set Ω forms a field *F* if

(i) $\Omega \in F$

(ii) If $A \in F$, then $\overline{A} \in F$

(iii) If $A \in F$ and $B \in F$, then $A \cup B \in F$.

Using (i) - (iii), it can be shown that

$$A \cap B, \overline{A} \cap B, \quad \text{etc.}$$

also belong to *F*.

E.g., from (ii) we have

$$\overline{A} \in F, \overline{B} \in F,$$

and using (iii) this gives

$$\overline{A} \cup \overline{B} \in F \ ;$$

applying (ii) again and using De Morgan's theorem we get

$$\overline{\overline{A} \cup \overline{B}} = A \cap B \in F \ ,$$

Thus if

$$A \in F, B \in F,$$

then

$$F = \left\{ \Omega, A, B, \overline{A}, \overline{B}, A \cup B, A \cap B, \overline{A} \cup B, \cdots \right\}.$$

From now on, we will use the term 'event' only to members of *F*.

Assuming that the probability

$$p_i = P(\xi_i)$$

of elementary outcomes ξ_i of Ω are apriori defined, how does one assign probabilities to more 'complicated' events such as *A*, *B*, *AB*, etc.?

The three axioms of probability defined below can be used to assign probabilities to more 'complicated' events.

<u>Axioms of Probability</u>

For any event *A,* assign a number *P*(*A*), called the probability of the event *A*.

This number satisfies the following three conditions

$$\text{(i)} \quad P(A) \geq 0 \quad \text{(Probability is a nonnegative number)}$$

$$\text{(ii)} \quad P(\Omega) = 1 \quad \text{(Probability of the whole set is unity)}$$

$$\text{(iii)} \text{ If } A \cap B = \phi, \quad \text{then} \quad P(A \cup B) = P(A) + P(B).$$

Note: (iii) states that if *A* and *B* are mutually exclusive (M.E.) events, the probability of their union is the sum of their probabilities.

The above three conditions form the axioms of probability.

The following conclusions follow from these axioms:

a. Since $A \cup \overline{A} = \Omega$, we have using (ii)

$$\text{P}(A \cup \overline{A}) = P(\Omega) = 1.$$

But $A \cap \overline{A} \in \phi$, and using (iii),

$$\text{P}(A \cup \overline{A}) = P(A) + \text{P}(\overline{A}) = 1 \quad \text{or} \quad \text{P}(\overline{A}) = 1 - P(A).$$

b. Similarly, for any *A*, $A \cap \phi = \phi$.

Hence it follows that $P(A \cup \phi) = P(A) + P(\phi)$.

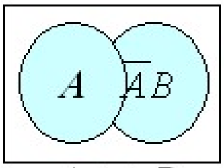
But $A \cup \phi = A$, and thus $P(\phi) = 0$.

c. Suppose *A* and *B* are *not* mutually exclusive (M.E.)?

How does one compute $P(A \cup B) = ?$

To compute the probability in (c) above, we re-express

$A \cup B$ in terms of M.E. sets so that we can make use of the probability axioms. From Fig. we have



$A \cup B$

$A \cup B = A \cup \overline{A}B,$

where A and $\overline{A}B$ are M.E. events.

Thus using axiom (iii)

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

To compute $P(\overline{A}B)$, we can express B as

$B = B \cap \Omega = B \cap (A \cup \overline{A})$

$= (B \cap A) \cup (B \cap \overline{A}) = BA \cup B\overline{A}$

Thus

$P(B) = P(BA) + P(B\overline{A}),$

since $BA = AB$ and $B\overline{A} = \overline{A}B$ are M.E. events.

$P(\overline{A}B) = P(B) - P(AB)$

and using this equation in

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

we have

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

- Question: Suppose every member of a denumerably infinite collection A_i of pair wise disjoint sets is an event, then what can we say about their union

$$A = \bigcup_{i=1}^{\infty} A_i \,?$$

i.e., suppose all $A_i \in F$, what about A? Does it belong to F?

Further, if A also belongs to F, what about $P(A)$?

For example, in a coin tossing experiment, where the same coin is tossed indefinitely, define

A = "head eventually appears".

Is A an event?

Our intuitive experience surely tells us that A is an event.

Let

$$A_n = \{\text{head appears for the 1st time on the } n\text{th toss}\}$$
$$= \{\underbrace{t, t, t, \cdots, t}_{n-1}, h\}$$

Clearly $A_i \cap A_j = \phi.$

Moreover the above A is

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_i \cup \cdots.$$

Here probability axiom (iii) can not be used to compute $P(A)$, since the axiom only deals with two (or a finite number) of Mutually Exclusive events.

To settle both questions above, extension to axioms must be done.

<u>**σ - Field**</u>:

A field F is a **σ** -field if in addition to the three conditions in the above axioms of probability, we have the following:

For every sequence

$$A_i \ , \ \ i = 1 \to \infty,$$

of pair wise disjoint events belonging to F, their union also belongs to F, i.e.,

$$A = \bigcup_{i=1}^{\infty} A_i \in F.$$

Using the above equation, another axiom can be added to the above set of 3 probability axioms, the 4th axiom as:

(iv) If A_i are pair wise mutually exclusive, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Returning back to the coin tossing experiment, from experience we know that if we keep tossing a coin, eventually, a head must show up, i.e.,

P(A) = 1

But

$$A = \bigcup_{n=1}^{\infty} A_n,$$

and using the 4th probability axiom

$$P(A) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

From,

$$\boxed{\begin{aligned} A_n &= \{\text{head appears for the 1st time on the } n\text{th toss}\} \\ &= \{\underbrace{t, t, t, \cdots, t}_{n-1}, h\} \end{aligned}}$$

for a fair coin since only one outcome in 2^n outcomes is in favor of A_n , we have

$$P(A_n) = \frac{1}{2^n} \quad \text{and} \quad \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1,$$

which agrees with P(A) = 1, thus justifying the fourth probability axiom.

In summary, the triplet (Ω, *F, P*) composed of a nonempty set Ω of elementary events, a **σ** -field *F* of subsets of Ω, and a probability measure *P* on the sets in *F*, subject to the four axioms, form a probability model.

The probability of more complicated events follow from this framework.

Conditional Probability and Independence

In *N* independent trials N_A, N_B, N_{AB} denote the number of times events *A*, *B* and *AB* occur, respectively.

According to the frequency interpretation of probability, for large *N*

$$P(A) \approx \frac{N_A}{N}, \quad P(B) \approx \frac{N_B}{N}, \quad P(AB) \approx \frac{N_{AB}}{N}.$$

Among the N_A occurrences of *A,* only N_{AB} of them are also found among the N_B occurrences of *B.* Thus the ratio

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)}$$

is a measure of “the event *A* given that *B* has already occurred”.

Denote this conditional probability by

$P(A|B)$ = Probability of “the event *A* given that *B* has occurred”.

Define

$$P(A\,|\,B) = \frac{P(AB)}{P(B)},$$

provided P(B) ≠ 0
We now show that the above definition satisfies all the 4 probability axioms.

$$P(A|B) = \frac{P(AB)}{P(B)},$$

(i) $$P(A|B) = \frac{P(AB) \geq 0}{P(B) > 0} \geq 0,$$

(ii) $$P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1,$$ since $\Omega B = B$.

(iii) Suppose $A \cap C = \phi$. Then

$$P(A \cup C|B) = \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P(AB \cup CB)}{P(B)}.$$

But $AB \cap CB = \phi$, hence $P(AB \cup CB) = P(AB) + P(CB)$.

$$P(A \cup C|B) = \frac{P(AB)}{P(B)} + \frac{P(CB)}{P(B)} = P(A|B) + P(C|B),$$

satisfying all probability axioms .

Thus $P(A|B) = \frac{P(AB)}{P(B)}$ defines a legitimate probability measure.

Properties of Conditional Probability:

a. If $B \subset A$, $AB = B$, and

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

since if $B \subset A$, then occurrence of B implies automatic occurrence of the event A.

As an example in a dice tossing experiment.

A = {outcome is even}, B={outcome is 2},

Then $B \subset A$, and $P(A|B) = 1$.

b. If $A \subset B$, $AB = A$, and

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} > P(A).$$

c. We can use the conditional probability to express the probability of a complicated event in terms of "simpler" related events.

Let $A_1, A_2, \cdots, A_n$ are pair wise disjoint and their union is Ω. Thus $A_i A_j = \phi$ and

$$\bigcup_{i=1}^{n} A_i = \Omega .$$

Thus

$$B = B(A_1 \cup A_2 \cup \cdots \cup A_n) = BA_1 \cup BA_2 \cup \cdots \cup BA_n.$$

But $A_i \cap A_j = \phi \Rightarrow BA_i \cap BA_j = \phi$, so

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

Using the concept of conditional probability, the concept of "independence of events" can be introduced.

Independence: A and B are said to be independent events, if

$$P(AB) = P(A) \cdot P(B).$$

Notice that the above definition is a probabilistic statement, *not* a set theoretic notion such as mutually exclusiveness.

Suppose A and B are independent, then

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

Thus if *A* and *B* are independent, the event that *B* has occurred does not give any clue about the event *A*.

It makes no difference to *A* whether *B* has occurred or not.

<u>Example:</u>

A box contains 6 white and 4 black balls.
Remove two balls at random without replacement.
What is the probability that the first one is white and the second one is black?

Let W_1 = "first ball removed is white"
B_2 = "second ball removed is black"

We need

$$P(W_1 \cap B_2) = ?$$

We have

$$W_1 \cap B_2 = W_1B_2 = B_2W_1.$$

$$P(W_1B_2) = P(B_2W_1) = P(B_2 \mid W_1)P(W_1).$$

Using the conditional probability rule,

$$P(W_1) = \frac{6}{6+4} = \frac{6}{10} = \frac{3}{5},$$

$$P(B_2 \mid W_1) = \frac{4}{5+4} = \frac{4}{9},$$

$$P(W_1B_2) = \frac{4}{9} \cdot \frac{3}{5} = \frac{12}{45}$$

Are the events W_1 and B_2 independent?
The first ball has two options:
W_1 = "first ball is white" or
B_1= "first ball is black".

Note that $W_1 \cap B_1 = \phi$ and $W_1 \cup B_1 = \Omega$.

Thus $P(B_2) = P(B_2 \mid W_1)P(W_1) + P(B_2 \mid B_1)P(B_1)$

$$= \frac{4}{5+4} \cdot \frac{3}{5} + \frac{3}{6+3} \cdot \frac{4}{10} = \frac{4}{9} \cdot \frac{3}{5} + \frac{1}{3} \cdot \frac{2}{5} = \frac{4+2}{15} = \frac{2}{5},$$

and

$$P(B_2)P(W_1) = \frac{2}{5} \cdot \frac{3}{5} \neq P(B_2 W_1) = \frac{12}{45}.$$

As expected, the events W_1 and B_2 are dependent.

BAYES' THEOREM

$$P(AB) = P(A \mid B)P(B). \qquad \text{Eq.1}$$

Similarly,

$$P(B \mid A) = \frac{P(BA)}{P(A)} = \frac{P(AB)}{P(A)},$$

or

$$P(AB) = P(B \mid A)P(A). \qquad \text{Eq.2}$$

From Eqs. 1 and 2 we get

$$P(A \mid B)P(B) = P(B \mid A)P(A).$$

or

$$\boxed{P(A \mid B) = \frac{P(B \mid A)}{P(B)} \cdot P(A)}$$

known as Bayes' theorem.

Interpretation of Bayes' theorem

$P(A)$ represents the a-priori probability of the event A.

Suppose B has occurred, and assume that A and B are not independent.

How can this new information ("B has occurred") be used to update our knowledge about A?

Bayes' rule takes into account the new information ("B has occurred") and gives out the a-posteriori probability of A given B.

A more general version of Bayes' theorem involves partition of Ω.

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

where

$$A_i, \; i = 1 \to n,$$

represent a set of mutually exclusive events with associated a-priori probabilities

$$P(A_i), \; i = 1 \to n.$$

With the new information "B has occurred", the information about A_i can be updated by the n conditional probabilities

$$P(B \mid A_i), \; i = 1 \to n,$$

Example:

Two boxes B_1 and B_2 contain 100 and 200 light bulbs respectively.

The first box (B_1) has 15 defective bulbs and the second 5.
Suppose a box is selected at random and one bulb is picked out.

(a) What is the probability that it is defective?

Note that box B_1 has 85 good and 15 defective bulbs. Similarly box B_2 has 195 good and 5 defective bulbs.

Let D = "Defective bulb is picked out".

Then

$$P(D \mid B_1) = \frac{15}{100} = 0.15, \quad P(D \mid B_2) = \frac{5}{200} = 0.025.$$

Since a box is selected at random, they are equally likely.

$$P(B_1) = P(B_2) = \frac{1}{2}.$$

Thus B_1 and B_2 form a partition and we obtain

$$P(D) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2)$$
$$= 0.15 \times \frac{1}{2} + 0.025 \times \frac{1}{2} = 0.0875.$$

Thus, there is 8.75% probability that a bulb picked at random is defective.

(b) Suppose we test the bulb and it is found to be defective.
What is the probability that it came from box 1?

$$P(B_1 \mid D) = ?$$

$$P(B_1 \mid D) = \frac{P(D \mid B_1)P(B_1)}{P(D)} = \frac{0.15 \times 1/2}{0.0875} = 0.8571.$$

Note that initially

$$P(B_1) = 0.5;$$

Then we picked out a box at random and tested a bulb that turned out to be defective.

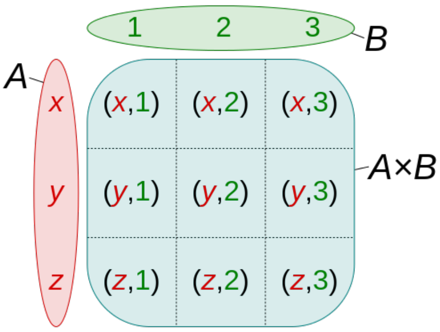
Can this information give some information about the fact that we might have picked up box 1?

$$P(B_1 \mid D) = 0.857 > 0.5,$$

Thus it is more likely that we must have chosen box 1 in favor of box 2. (Note that box 1 has six times more defective bulbs compared to box 2).

REPEATED SYSTEMS AND BERNOULLI TRIALS

Definition: For sets *A* and *B*, the **Cartesian product** *A* × *B* is the set of all ordered pairs (*a*, *b*) where $a \in A$ and $b \in B$.



Cartesian Product of *A* X *B* of the sets *A*={*x*,*y*,*z*} and B={1,2,3}

Consider two independent experiments with associated probability models (Ω_1, F_1, P_1) and (Ω_2, F_2, P_2).

Let $\xi \in \Omega_1$, $\eta \in \Omega_2$ represent elementary events.
A joint performance of the two experiments produces elementary event $\omega = (\xi, \eta)$.
How to characterize an appropriate probability to this "combined event"?
Consider the Cartesian product space $\Omega = \Omega_1 \times \Omega_2$ generated from Ω_1 and Ω_2 such that if $\xi \in \Omega_1$ and $\eta \in \Omega_2$, then every ω in Ω is an ordered pair of the form $\omega = (\xi, \eta)$.

To arrive at a probability model we need to define the combined trio (Ω, F, P).

Suppose $A \in F_1$ and $B \in F_2$.

Then $A \times B$ is the set of all pairs (ξ, η),

where $\xi \in A$ and $\eta \in B$.

Any such subset of Ω appears to be a legitimate event for the combined experiment.

Let F denote the field composed of all such subsets $A \times B$ together with their unions and compliments.

In this combined experiment, the probabilities of the events $A \times \Omega_2$ and $\Omega_1 \times B$ are such that

$$P(A \times \Omega_2) = P_1(A), \quad P(\Omega_1 \times B) = P_2(B).$$

Moreover, the events $A \times \Omega_2$ and $\Omega_1 \times B$ are independent for any $A \in F_1$ and $B \in F_2$. Since

$$(A \times \Omega_2) \cap (\Omega_1 \times B) = A \times B,$$

we conclude that

$$P(A \times B) = P(A \times \Omega_2) \cdot P(\Omega_1 \times B) = P_1(A) P_2(B)$$

for all $A \in F_1$ and $B \in F_2$.

This extends to a unique probability measure

$$P(\equiv P_1 \times P_2)$$

on the sets in F and defines the combined trio (Ω, F, P).

Generalization: Given n experiments $\Omega_1, \Omega_2, \cdots, \Omega_n$, and their associated F_i and P_i, $i = 1 \to n$, let

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$$

represent their Cartesian product whose elementary events are the ordered n-tuples $\xi_1, \xi_2, \cdots, \xi_n$, where $\xi_i \in \Omega_i$.

Events in this combined space are of the form

$$A_1 \times A_2 \times \cdots \times A_n$$

where $A_i \in F_i$, and their unions an intersections.

If all these n experiments are independent, and $P_i(A_i)$ is the probability of the event A_i in F_i then

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

Example : An event A has probability p of occurring in a single trial. Find the probability that A occurs exactly k times, $k \leq n$ in n trials.

Let (Ω, F, P) be the probability model for a single trial. The outcome of n experiments is an n-tuple

$$\omega = \{ \xi_1, \xi_2, \cdots, \xi_n \} \in \Omega_0,$$

where every $\xi_i \in \Omega$ and $\Omega_0 = \Omega \times \Omega \times \cdots \times \Omega$
The event A occurs at trial # i , if $\xi_i \in A$.

Suppose A occurs exactly k times in ω.
Then k of the ξ_i belong to A, say $\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}$, and the remaining $n-k$ are contained in its compliment in $\overline{A}$.
Probability of occurrence of such an ω is

$$P_0(\omega) = P(\{\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}, \cdots, \xi_{i_n}\}) = P(\{\xi_{i_1}\})P(\{\xi_{i_2}\})\cdots P(\{\xi_{i_k}\})\cdots P(\{\xi_{i_n}\})$$
$$= \underbrace{P(A)P(A)\cdots P(A)}_{k}\ \underbrace{P(\overline{A})P(\overline{A})\cdots P(\overline{A})}_{n-k} = p^k q^{n-k}.$$

However the k occurrences of A can occur in any particular location inside ω.

Let $\omega_1, \omega_2, \cdots, \omega_N$ represent all such events in which A occurs exactly k times. Then

$$\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"} = \omega_1 \cup \omega_2 \cup \cdots \cup \omega_N.$$

But, all these ω_i s are mutually exclusive, and equiprobable.

Thus $P(\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"})$

$$= \sum_{i=1}^{N} P_0(\omega_i) = N P_0(\omega) = N p^k q^{n-k},$$

Recall that, starting with n possible choices, the first object can be chosen n different ways,
and for every such choice
the second one in $(n-1)$ ways, …
and the kth one $(n-k+1)$ ways,
and this gives the total choices for k objects out of n to be

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

But, this includes the $k!$ choices among the k objects that are indistinguishable for identical objects. As a result

$$N = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!} \triangleq \binom{n}{k}$$

represents the number of combinations, or choices of n identical objects taken k at a time.
Using the last two equations, we get

$$
\boxed{\begin{aligned} P_n(k) &= P(\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"}) \\ &= \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\cdots,n, \end{aligned}}
$$

This is known as the formula for Bernoulli trials.

Independent repeated experiments of this nature, where the outcome is either a "success" $(= A)$ or a "failure" $(= \overline{A})$

are characterized as Bernoulli trials, and the probability of *k* successes

in *n* trials is given by this Bernoulli formula, where *p* represents the probability of "success" in any one trial.

Example: Toss a coin n times. Obtain the probability of getting k heads in n trials ?
We identify "head" with "success" (A) and let $p = P(H)$.

Using Bernoulli formula

$$
p = P(H) = \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\cdots,n,
$$

<u>Example:</u> Consider rolling a fair die eight times. Find the probability that either 3 or 4 shows up five times.

We identify

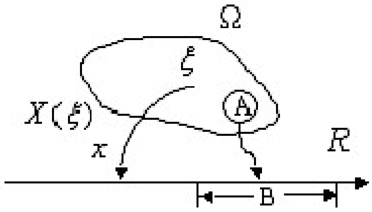
$$
\text{"success"} = A = \{ \text{ either } 3 \text{ or } 4 \} = \{f_3\} \cup \{f_4\}.
$$

$$
P(A) = P(f_3) + P(f_4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},
$$

Thus
The desired probability is found by using Bernoulli formula for n=8, k=5 and p=1/3.

3. Random Variables

Let (Ω, F, P) be a probability model for an experiment, and X a function that maps every $\xi \in \Omega$, to a unique point $x \in R$, the set of real numbers. Since the outcome is not certain, so is the value $X(\xi) = x$. Thus if B is some subset of R, we may want to determine the probability of "$X(\xi) \in B$". To determine this probability, we can look at the set $A = X^{-1}(B) \in \Omega$ that contains all $\xi \in \Omega$ that maps into B under the function X.



Obviously, if the set $A = X^{-1}(B)$ also belongs to the associated field F, then it is an event and the probability of A is well defined; in that case we can say

Probability of the event "$X(\xi) \in B$" $= P(X^{-1}(B))$.

However, $X^{-1}(B)$ may not always belong to F for all B, thus creating difficulties. The notion of random variable (r.v) makes sure that the inverse mapping always results in an event so that we are able to determine the probability for any $B \in R$.

Random Variable (r.v): A finite single valued function $X(\cdot)$ that maps the set of all experimental outcomes Ω into the set of real numbers R is said to be a r.v, if the set $\{\xi \mid X(\xi) \le x\}$ is an event $(\in F)$ for every x in R.

If X is a random variable, then

$$\{ \xi \mid X(\xi) \le x \} = \{ X \le x \}$$

is an event for every x.

Are $\{a < X \le b\}$, $\{X = a\}$ also events ?

With $b > a$ since $\{X \le a\}$ and $\{X \le b\}$ are events,

$\{X \le a\}^c = \{X > a\}$ is an event

And hence

$\{X > a\} \cap \{X \le b\} = \{a < X \le b\}$ is also an event.

Thus, $\left\{ a - \frac{1}{n} < X \le a \right\}$ is an event for every n.
Consequently

$$\bigcap_{n=1}^{\infty} \left\{ a - \frac{1}{n} < X \le a \right\} = \{ X = a \}$$

is also an event.

All events have well defined probability.

Thus the probability of the event $\{ \xi \mid X(\xi) \le x \}$ must depend on x. Denote

$$P\{ \xi \mid X(\xi) \le x \} = F_X(x) \ge 0.$$

The role of the subscript X is only to identify the actual r.v. $F_X(x)$ is the Probability Distribution Function (PDF) associated with the r.v X.

Distribution Function: Note that a distribution function $g(x)$ is nondecreasing, right-continuous and satisfies

$$g(+\infty) = 1, \quad g(-\infty) = 0,$$

i.e., if $g(x)$ is a distribution function, then

(i) $g(+\infty) = 1, \quad g(-\infty) = 0,$

(ii) if $x_1 < x_2$, then $g(x_1) \leq g(x_2)$,

and

(iii) $g(x^+) = g(x)$, for all x.

We need to show that $F_X(x)$ defined as

$$P\{\xi \mid X(\xi) \leq x\} = F_X(x) \geq 0.$$

satisfies all properties in i, ii, iii

For any r.v X,

(i) $F_X(+\infty) = P\{\xi \mid X(\xi) \leq +\infty\} = P(\Omega) = 1$

and $F_X(-\infty) = P\{\xi \mid X(\xi) \leq -\infty\} = P(\phi) = 0.$

(ii) If $x_1 < x_2$, then the subset $(-\infty, x_1) \subset (-\infty, x_2)$.
Consequently the event $\{\xi \mid X(\xi) \leq x_1\} \subset \{\xi \mid X(\xi) \leq x_2\}$, since $X(\xi) \leq x_1$ implies $X(\xi) \leq x_2$. As a result

$$F_X(x_1) \triangleq P(X(\xi) \leq x_1) \leq P(X(\xi) \leq x_2) \triangleq F_X(x_2),$$

implying that the probability distribution function is nonnegative and monotone nondecreasing.

(iii) Let $x < x_n < x_{n-1} < \cdots < x_2 < x_1,$ and consider the event

$A_k = \{ \xi \mid x < X(\xi) \leq x_k \}$

since

$\{ x < X(\xi) \leq x_k \} \cup \{ X(\xi) \leq x \} = \{ X(\xi) \leq x_k \},$

using mutually exclusive property of events we get

$$P(A_k) = P\big(x < X(\xi) \leq x_k\big) = F_X(x_k) - F_X(x).$$

But $\cdots A_{k+1} \subset A_k \subset A_{k-1} \cdots,$ and hence

$$\lim_{k\to\infty} A_k = \bigcap_{k=1}^{\infty} A_k = \phi \quad \text{and hence} \quad \lim_{k\to\infty} P(A_k) = 0.$$

Thus

$$\lim_{k\to\infty} P(A_k) = \lim_{k\to\infty} F_X(x_k) - F_X(x) = 0.$$

But $\lim_{k\to\infty} x_k = x^+,$ the right limit of x, and hence

$$F_X(x^+) = F_X(x),$$

i.e., $F_X(x)$ is right-continuous, justifying all properties of a distribution function.

Additional Properties of a PDF

(iv) If $F_X(x_0) = 0$ for some x_0, then $F_X(x) = 0, \quad x \leq x_0.$

This follows, since $F_X(x_0) = P\big(X(\xi) \leq x_0\big) = 0$ implies $\{X(\xi) \leq x_0\}$ is the null set, and for any $x \leq x_0$, $\{ X(\xi) \leq x \}$ will be a subset of the null set.

(v) $P\{ X(\xi) > x \} = 1 - F_X(x)$

We have $\{X(\xi) \le x\} \cup \{X(\xi) > x\} = \Omega$ and since the two events are mutually exclusive, $P\{ X(\xi) > x \} = 1 - F_X(x)$

(vi) $P\{ x_1 < X(\xi) \le x_2 \} = F_X(x_2) - F_X(x_1), \ \ x_2 > x_1.$

The events $\{X(\xi) \le x_1\}$ and $\{x_1 < X(\xi) \le x_2\}$ are mutually exclusive and their union represents the event $\{X(\xi) \le x_2\}$.

(vii) $P(X(\xi) = x) = F_X(x) - F_X(x^-).$

Let $x_1 = x - \varepsilon, \ \ \varepsilon > 0,$ and $x_2 = x.$

From $P\{ x_1 < X(\xi) \le x_2 \} = F_X(x_2) - F_X(x_1), \ \ x_2 > x_1.$

$$\lim_{\varepsilon \to 0} P\{ x - \varepsilon < X(\xi) \le x \} = F_X(x) - \lim_{\varepsilon \to 0} F_X(x - \varepsilon),$$

or

$$P\{ X(\xi) = x \} = F_X(x) - F_X(x^-).$$

According to $F_X(x^+) = F_X(x)$,

$F_X(x_0^+)$, the limit of $F_X(x)$ as $x \to x_0$ from the right always exists and equals $F_X(x_0)$.

However the left limit value $F_X(x_0^-)$ need not be equal to $F_X(x_0)$.

Thus

$F_X(x)$ need not be continuous from the left.

At a discontinuity point of the distribution,

the left and right limits are different and from

$$P\{ X(\xi) = x \} = F_X(x) - F_X(x^-).$$

$$P\{ X(\xi) = x_0 \} = F_X(x_0) - F_X(x_0^-) > 0.$$

Thus the only discontinuities of a distribution function $F_X(x)$ are of the jump type, and occur at points x_0 where

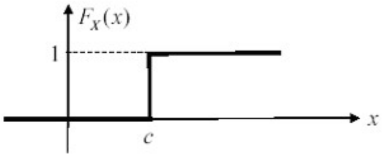
$P\{X(\xi)=x_0\}=F_X(x_0)-F_X(x_0^-)>0.$

is satisfied.

Example :

X is a r.v such that $X(\xi)=c,\ \xi\in\Omega.$ Find $F_X(x).$

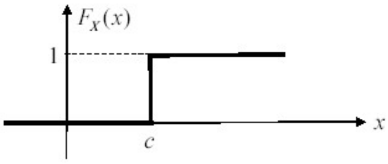
For $x<c,\ \{X(\xi)\le x\}=\{\phi\},$ so that $F_X(x)=0,$ and for $x>c,\ \{X(\xi)\le x\}=\Omega,$ so that $F_X(x)=1.$



- X is said to be a continuous-type r.v if its distribution function $F_X(x)$ is continuous. In that case $F_X(x^-)=F_X(x)$ for all x, and from $P\{X(\xi)=x_0\}=F_X(x_0)-F_X(x_0^-)>0$ we get $P\{X=x\}=0.$

- If $F_X(x)$ is constant except for a finite number of jump discontinuities (piece-wise constant; step-type), then X is said to be a discrete-type r.v. If x_i is such a discontinuity point, then from $P\{X(\xi)=x_0\}=F_X(x_0)-F_X(x_0^-)>0$ we get $p_i=P\{X=x_i\}=F_X(x_i)-F_X(x_i^-).$

From Fig.



at a point of discontinuity we get

$P\{X=c\}=F_X(c)-F_X(c^-)=1-0=1.$

Example:

A fair coin is tossed twice, and let the r.v X represents the number of heads. Find $F_X(x).$

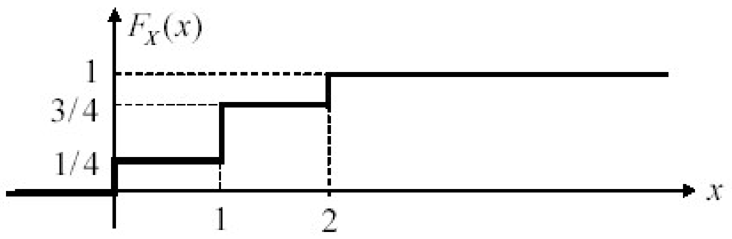
In this case $\Omega=\{HH,HT,TH,TT\},$ and $X(HH)=2, X(HT)=1, X(TH)=1, X(TT)=0.$

$$x<0,\quad \{X(\xi)\le x\}=\phi\Rightarrow F_X(x)=0,$$

$$0\le x<1, \{X(\xi)\le x\}=\{TT\}\Rightarrow F_X(x)=P\{TT\}=P(T)P(T)=\frac{1}{4},$$

$$1\le x<2, \{X(\xi)\le x\}=\{TT,HT,TH\}\Rightarrow F_X(x)=P\{TT,HT,TH\}=\frac{3}{4},$$

$$x\ge 2,\quad \{X(\xi)\le x\}=\Omega\Rightarrow F_X(x)=1$$



From Fig., $P\{X=1\} = F_X(1) - F_X(1^-) = 3/4 - 1/4 = 1/2.$

Probability density function (p.d.f)

The derivative of the distribution function $F_X(x)$ is called the probability density function $f_X(x)$ of the r.v X.
Thus

$$f_X(x) \stackrel{\Delta}{=} \frac{dF_X(x)}{dx}.$$

Since

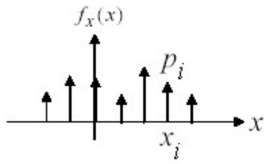
$$\frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \geq 0,$$

from the monotone-nondecreasing nature of $F_X(x)$, it follows that $f_X(x) \geq 0$ for all x.

$f_X(x)$ will be a continuous function, if X is a continuous type r.v.

However, if X is a discrete type r.v , then its p.d.f has the general form

$$f_X(x) = \sum_i p_i \delta(x - x_i),$$



where x_i represent the jump-discontinuity points in $F_X(x)$.

As Fig. shows $f_X(x)$ represents a collection of positive discrete masses, and it is known as the probability mass function (p.m.f) in the discrete case.

From $f_X(x) \stackrel{\Delta}{=} \dfrac{dF_X(x)}{dx}$ we also obtain by integration

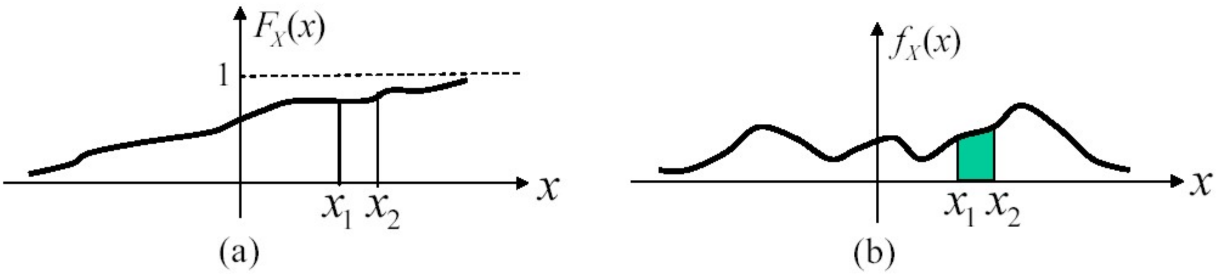
$$F_X(x) = \int_{-\infty}^{x} f_x(u)\,du \,.$$

Since $F_X(+\infty) = 1$, this yields $\int_{-\infty}^{+\infty} f_x(x)\,dx = 1$

Further, from $F_X(x) = \int_{-\infty}^{x} f_x(u)\,du$ we also get

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

Thus the area under $f_X(x)$ in the interval (x_1, x_2) represents the probability in the above equation



Continuous-type random variables

1. Normal (Gaussian): X is said to be normal or Gaussian r.v, if

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

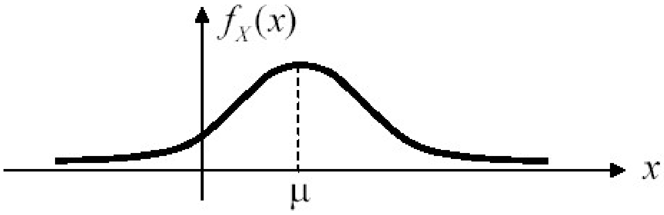
This is a bell shaped curve, symmetric around the parameter μ, and its distribution function is given by

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy \stackrel{\Delta}{=} G\left(\frac{x-\mu}{\sigma}\right),$$

where $G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is often tabulated.

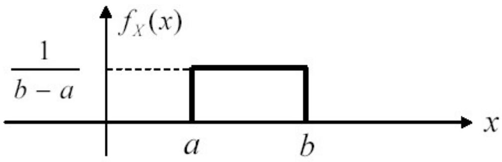
Since $f_X(x)$ depends on two parameters μ and σ^2, the notation $X \sim N(\mu, \sigma^2)$ will be used to represent

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



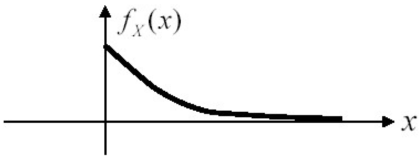
2. Uniform: $X \sim U(a,b),\ a<b,$ if

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



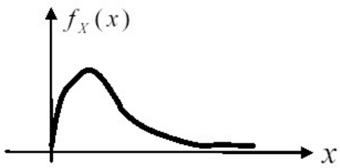
3. Exponential: $X \sim \varepsilon(\lambda)$ if

$$f_X(x) = \begin{cases} \dfrac{1}{\lambda} e^{-x/\lambda}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$



4. Gamma: $X \sim G(\alpha,\beta)$ if $(\alpha > 0,\ \beta > 0)$

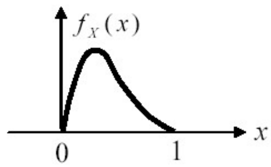
$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-x/\beta}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$



If $\alpha = n$ an integer $\Gamma(n) = (n-1)!$

5. Beta: $X \sim \beta(a,b)$ if $(a>0,\ b>0)$

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

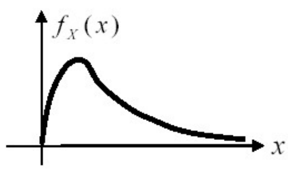


where the Beta function $\beta(a,b)$ is defined as

$$\beta(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\,du.$$

6. Chi-Square: $X \sim \chi^2(n),$ if

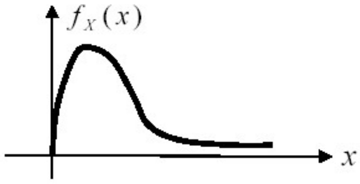
$$f_X(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$



Note that $\chi^2(n)$ is the same as Gamma $(n/2, 2)$.

7. Rayleigh: $X \sim R(\sigma^2)$, if

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

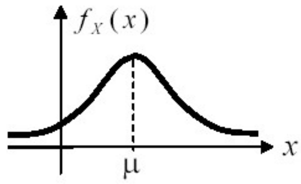


8. Nakagami – m distribution:

$$f_X(x) = \begin{cases} \dfrac{2}{\Gamma(m)} \left(\dfrac{m}{\Omega} \right)^m x^{2m-1} e^{-mx^2/\Omega}, & x \geq 0 \\ 0 \quad , & \text{otherwise} \end{cases}$$

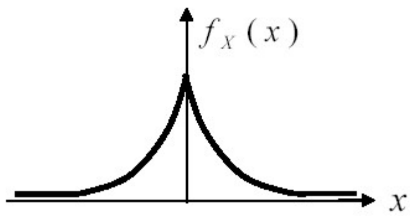
9. Cauchy: $X \sim C(\alpha, \mu)$, if

$$f_X(x) = \frac{\alpha / \pi}{\alpha^2 + (x - \mu)^2}, \quad -\infty < x < +\infty .$$



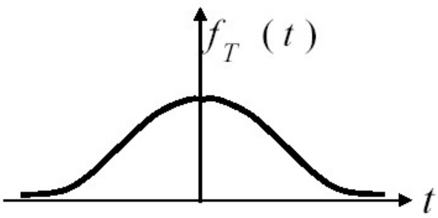
10. Laplace:

$$f_X(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}, \quad -\infty < x < +\infty .$$



11. Student's t-distribution with n degrees of freedom

$$f_T(t) = \frac{\Gamma\left((n+1)/2\right)}{\sqrt{\pi n}\, \Gamma(n/2)} \left(1 + \frac{t^2}{n} \right)^{-(n+1)/2}, \quad -\infty < t < +\infty .$$



12. Fisher's F-distribution

$$f_z(z) = \begin{cases} \dfrac{\Gamma\{(m+n)/2\}\, m^{m/2} n^{n/2}}{\Gamma(m/2)\,\Gamma(n/2)} \dfrac{z^{m/2-1}}{(n+mz)^{(m+n)/2}}, & z \geq 0 \\ 0 \quad , & \text{otherwise} \end{cases}$$

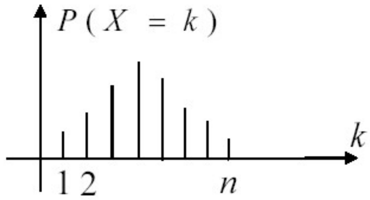
Discrete-type random variables

1. Bernoulli: X takes the values (0,1), and

$$P(X=0)=q, \quad P(X=1)=p.$$

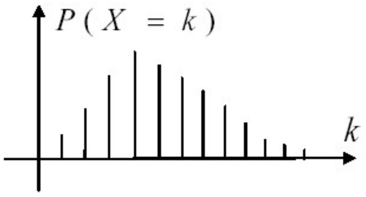
2. Binomial: $X \sim B(n,p)$, if

$$P(X=k)=\binom{n}{k} p^k q^{n-k}, \quad k=0,1,2,\cdots,n.$$



3. Poisson: $X \sim P(\lambda)$, if

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}, \quad k=0,1,2,\cdots,\infty.$$



4. Hypergeometric:

$$P(X=k)=\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, \qquad \max(0,m+n-N) \le k \le \min(m,n)$$

5. Geometric: $X \sim g(p)$ if

$$P(X=k)=pq^k, \quad k=0,1,2,\cdots,\infty, \quad q=1-p.$$

6. Negative Binomial: $X \sim NB(r,p)$, if

$$P(X=k)=\binom{k-1}{r-1} p^r q^{k-r}, \quad k=r,r+1,\cdots.$$

7. Discrete-Uniform:

$$P(X=k)=\frac{1}{N}, \quad k=1,2,\cdots,N.$$

Conditional Probability Density Function

For any two events A and B, we have defined the conditional probability of A given B as

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

Noting that the probability distribution function $F_X(x)$ is given by

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

we may define the conditional distribution of the r.v X given the event B as

$$F_X(x\,|\,B)=P\{X(\xi)\leq x\,|\,B\}=\frac{P\{(X(\xi)\leq x)\cap B\}}{P(B)}$$

Thus the definition of the conditional distribution depends on conditional probability, and since it obeys all probability axioms, it follows that the conditional distribution has the same properties as any distribution function. In particular

$$F_X(+\infty\,|\,B)=\frac{P\{(X(\xi)\leq +\infty)\cap B\}}{P(B)}=\frac{P(B)}{P(B)}=1,$$

$$F_X(-\infty\,|\,B)=\frac{P\{(X(\xi)\leq -\infty)\cap B\}}{P(B)}=\frac{P(\phi)}{P(B)}=0.$$

Further

$$P(x_1<X(\xi)\leq x_2\,|\,B)=\frac{P\{(x_1<X(\xi)\leq x_2)\cap B\}}{P(B)}$$

$$=F_X(x_2\,|\,B)-F_X(x_1\,|\,B),$$

Since for $x_2\geq x_1$,

$$\left(X(\xi)\leq x_2\right)=\left(X(\xi)\leq x_1\right)\cup\left(x_1<X(\xi)\leq x_2\right)$$

The conditional density function is the derivative of the conditional distribution function. Thus

$$f_X(x \mid B) = \frac{dF_X(x \mid B)}{dx}, \quad \text{and}$$

$$F_X(x \mid B) = \int_{-\infty}^{x} f_X(u \mid B)\,du. \text{ Also}$$

$$P\big(x_1 < X(\xi) \le x_2 \mid B\big) = \int_{x_1}^{x_2} f_X(x \mid B)\,dx.$$

<u>Example</u>: Given $F_X(x)$, suppose $B = \{X(\xi) \le a\}$. Find $f_X(x \mid B)$.

We will first determine $F_X(x \mid B)$. From

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap B\}}{P(B)} \quad \text{and}$$

B as given above, we have

$$F_X(x \mid B) = \frac{P\{(X \le x) \cap (X \le a)\}}{P(X \le a)}.$$

For $x < a$, $(X \le x) \cap (X \le a) = (X \le x)$ so that

$$F_X(x \mid B) = \frac{P(X \le x)}{P(X \le a)} = \frac{F_X(x)}{F_X(a)}.$$

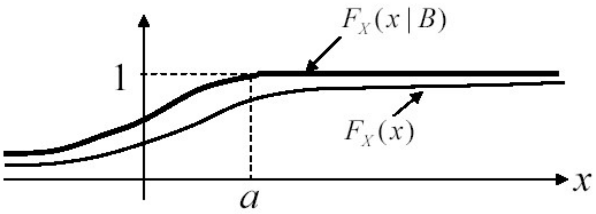
For $x \ge a$, $(X \le x) \cap (X \le a) = (X \le a)$ so that $F_X(x \mid B) = 1$.

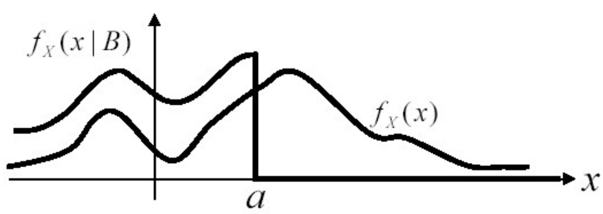
Thus

$$F_X(x \mid B) = \begin{cases} \dfrac{F_X(x)}{F_X(a)}, & x < a, \\ 1, & x \ge a, \end{cases}$$

and hence

$$F_X(x \mid B) = \begin{cases} \dfrac{F_X(x)}{F_X(a)}, & x < a, \\ 1, & x \ge a, \end{cases}$$





Example: Let B represent the event $\{a < X(\xi) \le b\}$ with $b > a$. For a given $F_X(x)$, determine $F_X(x \mid B)$ and $f_X(x \mid B)$.

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap (a < X(\xi) \le b)\}}{P(a < X(\xi) \le b)}$$

$$= \frac{P\{(X(\xi) \le x) \cap (a < X(\xi) \le b)\}}{F_X(b) - F_X(a)}.$$

For $x < a$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \phi$, hence $F_X(x \mid B) = 0.$

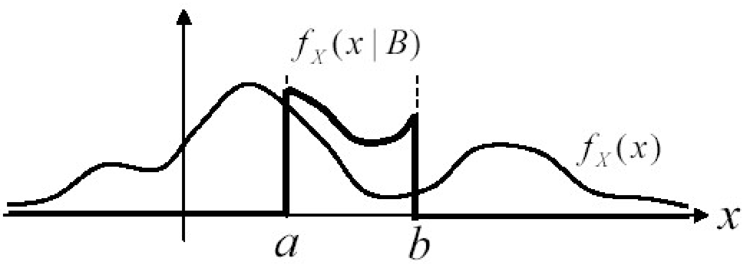
For $a \le x < b$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \{a < X(\xi) \le x\}$ and hence

$$F_X(x \mid B) = \frac{P(a < X(\xi) \le x)}{F_X(b) - F_X(a)} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}.$$

For $x \ge b$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \{a < X(\xi) \le b\}$ so that $F_X(x \mid B) = 1$.

Using $\quad f_X(x \mid B) = \dfrac{dF_X(x \mid B)}{dx}$

$$f_X(x \mid B) = \begin{cases} \dfrac{f_X(x)}{F_X(b) - F_X(a)}, & a < x \le b, \\ 0, & \text{otherwise.} \end{cases}$$



Functions of a Random Variable

Let X be a r.v defined on the model (Ω, F, P), and suppose $g(x)$ is a function of the variable x. Define

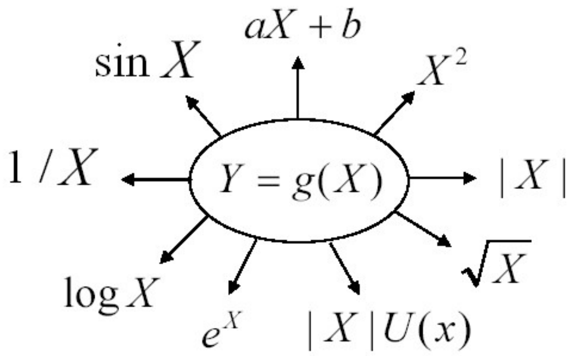
$Y = g(X).$

Is Y necessarily a r.v? If so what is its PDF $F_Y(y)$, pdf $f_Y(y)$?

If X is a r.v, so is Y,

$$F_Y(y) = P(Y(\xi) \le y) = P\big(g(X(\xi)) \le y\big) = P\big(X(\xi) \le g^{-1}(-\infty, y]\big).$$

Thus the distribution function as well of the density function of Y can be determined in terms of that of X. To obtain the distribution function of Y, we must determine the set on the x-axis such that $X(\xi) \le g^{-1}(y)$ for every given y, and the probability of that set.



Example: $Y = aX + b$

Suppose $a > 0$.

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(aX(\xi) + b \le y\big) = P\left(X(\xi) \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right) \text{ and}$$

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right). \qquad \text{Eq. 1}$$

On the other hand if $a < 0$, then

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(aX(\xi) + b \le y\big) = P\left(X(\xi) > \frac{y-b}{a}\right)$$

$$= 1 - F_X\left(\frac{y-b}{a}\right),$$

and hence

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right). \qquad \text{Eq. 2}$$

From Eqs. 1 and 2, we obtain (for all a)

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Example: $Y = X^2.$

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(X^2(\xi) \le y\big).$$

If $y < 0,$ then the event $\{X^2(\xi) \le y\} = \phi$, and hence

$$F_Y(y) = 0, \quad y < 0.$$

For $y > 0,$ from Fig.

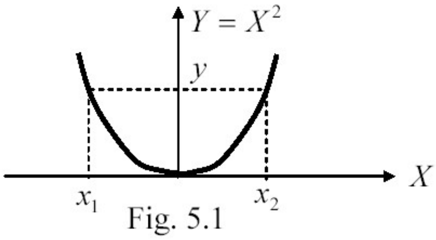


Fig. 5.1

the event $\{Y(\xi) \le y\} = \{X^2(\xi) \le y\}$ is equivalent to $\{x_1 < X(\xi) \le x_2\}.$
Hence

$$F_Y(y) = P\big(x_1 < X(\xi) \le x_2\big) = F_X(x_2) - F_X(x_1)$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \quad y > 0.$$

By direct differentiation, we get

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\big(f_X(\sqrt{y}) + f_X(-\sqrt{y})\big), & y > 0, \\ 0, & \text{otherwise} \end{cases} .$$

If $f_X(x)$ represents an even function, then $f_Y(y)$ reduces to

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X\big(\sqrt{y}\big)\, U(y) \quad \text{where } U \text{ is the unit function}$$

In particular if $X \sim N(0,1),$ so that

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

we obtain the p.d.f of $Y = X^2$ to be

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} U(y).$$

Notice that the above $f_Y(y)$ represents a Chi-square r.v with $n = 1$ since $\Gamma(1/2) = \sqrt{\pi}.$

Thus, if X is a Gaussian r.v with $\mu = 0,$ then $Y = X^2$ represents a Chi-square r.v with one degree of freedom $(n = 1).$

Example: Let

$$Y = g(X) = \begin{cases} X - c, & X > c, \\ 0, & -c < X \le c, \\ X + c, & X \le -c. \end{cases}$$

In this case

$$P(Y = 0) = P(-c < X(\xi) \le c) = F_X(c) - F_X(-c).$$

For $y > 0$, we have $x > c$, and $Y(\xi) = X(\xi) - c$ so that

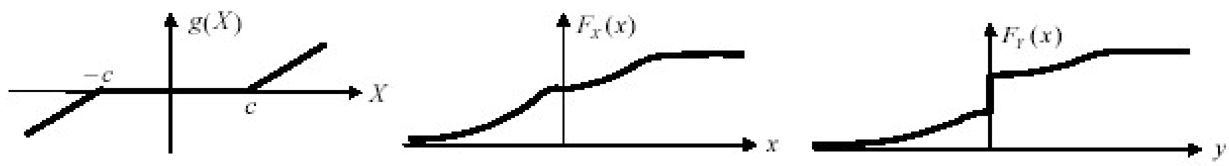
$$F_Y(y) = P\big(Y(\xi) \le y\big) = P(X(\xi) - c \le y)$$
$$= P\big(X(\xi) \le y + c\big) = F_X(y + c), \quad y > 0.$$

Similarly $y < 0$, if $x \le -c$, and $Y(\xi) = X(\xi) + c$ so that

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P(X(\xi) + c \le y)$$
$$= P\big(X(\xi) \le y - c\big) = F_X(y - c), \quad y < 0.$$

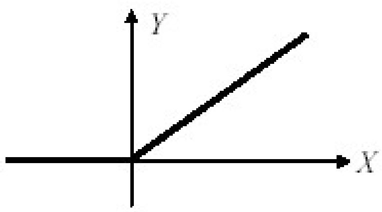
Thus

$$f_Y(y) = \begin{cases} f_X(y + c), & y \ge 0, \\ f_X(y - c), & y < 0. \end{cases}$$



Example: Half-wave rectifier

$$Y = g(X); \quad g(x) = \begin{cases} x, & x > 0, \\ 0, & x \le 0. \end{cases}$$



In this case

$$P(Y = 0) = P(X(\xi) \le 0) = F_X(0).$$

and for $y > 0$, since $Y = X$,

$$F_Y(y) = P(Y(\xi) \le y) = P(X(\xi) \le y) = F_X(y).$$

Thus

$$f_Y(y) = \begin{cases} f_X(y), & y > 0, \\ 0, & y \le 0, \end{cases} = f_X(y)U(y).$$

Note: As a general approach, given $Y = g(X)$,

first sketch the graph $y = g(x)$, and determine the range space of y.

Suppose $a < y < b$ is the range space of $y = g(x)$.

Then clearly for $y < a$, $F_Y(y) = 0$, and for $y > b$, $F_Y(y) = 1$, so that $F_Y(y)$ can be nonzero only in $a < y < b$.

Next, determine whether there are discontinuities in the range space of y.

If so evaluate $P(Y(\xi) = y_i)$ at these discontinuities.

In the continuous region of y, use the basic approach

$$F_Y(y) = P(g(X(\xi)) \le y)$$

and determine appropriate events in terms of the r.v X for every y.

Finally, we must have $F_Y(y)$ for $-\infty \le y \le +\infty$, and obtain

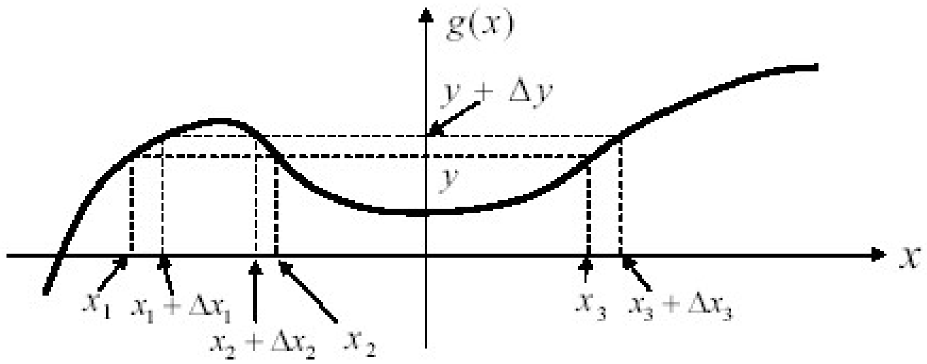
$$f_Y(y) = \frac{dF_Y(y)}{dy} \quad \text{in} \quad -a < y < b.$$

However, if $Y = g(X)$ is a continuous function, it is easy to establish a direct procedure to obtain $f_Y(y)$.

A continuos function $g(x)$ with $g'(x)$ nonzero at all but a finite number of points, has only a finite number of maxima and minima, and

it eventually becomes monotonic as $|x| \to \infty$.

Consider a specific y on the y-axis, and a positive increment Δy as shown in the below Fig.



$f_Y(y)$ for $Y = g(X)$, where $g(\cdot)$ is of continuous type.

Using

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

we can write

$$P\{y < Y(\xi) \le y + \Delta y\} = \int_{y}^{y+\Delta y} f_Y(u)\,du \approx f_Y(y) \cdot \Delta y.$$

But the event $\{y < Y(\xi) \le y + \Delta y\}$ can be expressed in terms of $X(\xi)$ as well.

To see this, referring back to the above Fig. , we notice that the equation $y = g(x)$ has three solutions x_1, x_2, x_3 (for the specific y chosen there).

As a result when $\{y < Y(\xi) \le y + \Delta y\}$, the r.v X could be in any one of the three mutually exclusive intervals

$\{x_1 < X(\xi) \le x_1 + \Delta x_1\}$, $\{x_2 + \Delta x_2 < X(\xi) \le x_2\}$ or $\{x_3 < X(\xi) \le x_3 + \Delta x_3\}$.

Hence the probability of the event in

$$P\{y < Y(\xi) \le y + \Delta y\} = \int_y^{y+\Delta y} f_Y(u)\,du \approx f_Y(y) \cdot \Delta y.$$

is the sum of the probability of the above three events, i.e.,

$$P\{y < Y(\xi) \le y + \Delta y\} = P\{x_1 < X(\xi) \le x_1 + \Delta x_1\}$$
$$+ P\{x_2 + \Delta x_2 < X(\xi) \le x_2\} + P\{x_3 < X(\xi) \le x_3 + \Delta x_3\}.$$

For small $\Delta y, \Delta x_i$, making use of the approximation in

$$P\{y < Y(\xi) \le y + \Delta y\} = \int_y^{y+\Delta y} f_Y(u)\,du \approx f_Y(y) \cdot \Delta y.$$

we get

$$f_Y(y)\Delta y = f_X(x_1)\Delta x_1 + f_X(x_2)(-\Delta x_2) + f_X(x_3)\Delta x_3.$$

In this case, $\Delta x_1 > 0$, $\Delta x_2 < 0$ and $\Delta x_3 > 0$, so that the above equation can be rewritten as

$$f_Y(y) = \sum_i f_X(x_i)\frac{|\Delta x_i|}{\Delta y} = \sum_i \frac{1}{|\Delta y / \Delta x_i|} f_X(x_i)$$

and as $\Delta y \to 0$, the above equation can be expressed as

$$f_Y(y) = \sum_i \frac{1}{|dy/dx|_{x_i}} f_X(x_i) = \sum_i \frac{1}{|g'(x_i)|} f_X(x_i).$$

The summation index i in the above equation depends on y, and for every y the equation $y = g(x_i)$ must be solved to obtain the total number of solutions at every y, and the actual solutions $x_1, x_2, \cdots$ all in terms of y.

Example: $Y = \dfrac{1}{X}$. Find $f_Y(y)$.

Here for every y, $x_1 = 1/y$ is the only solution, and

$$\frac{dy}{dx} = -\frac{1}{x^2} \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x=x_1} = \frac{1}{1/y^2} = y^2,$$

and substituting this into

$$f_Y(y) = \sum_i \frac{1}{|dy/dx|_{x_i}} f_X(x_i) = \sum_i \frac{1}{|g'(x_i)|} f_X(x_i).$$

we obtain

$$f_Y(y) = \frac{1}{y^2} f_X\left(\frac{1}{y}\right).$$

In particular, suppose X is a Cauchy r.v $X \sim C(\alpha, \mu = 0)$, i.e.

$$f_X(x) = \frac{\alpha/\pi}{\alpha^2 + (x-\mu)^2} = \frac{\alpha/\pi}{\alpha^2 + x^2}, \quad -\infty < x < +\infty.$$

In that case from

$$f_Y(y) = \frac{1}{y^2} f_X\left(\frac{1}{y}\right) \;,$$

$Y = 1/X$ has the p.d.f

$$f_Y(y) = \frac{1}{y^2}\frac{\alpha/\pi}{\alpha^2 + (1/y)^2} = \frac{(1/\alpha)/\pi}{(1/\alpha)^2 + y^2}, \quad -\infty < y < +\infty.$$

which represents the p.d.f of a Cauchy r.v with parameter $1/\alpha$.

Thus if $X \sim C(\alpha, \mu = 0)$ then $1/X \sim C(1/\alpha, \mu = 0)$

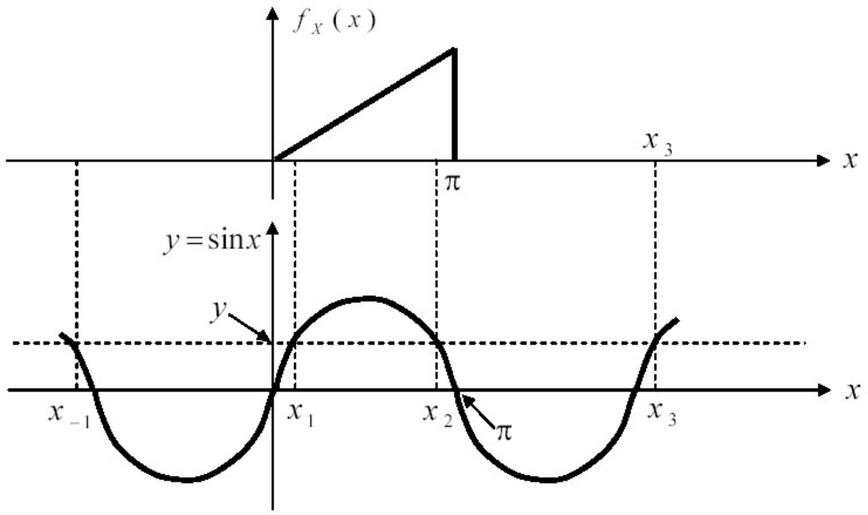
Example: $f_X(x) = 2x/\pi^2$, $0 < x < \pi$, and $Y = \sin X$. Determine $f_Y(y)$.

Since X has zero probability of falling outside the interval $(0, \pi)$,

$y = \sin x$ has zero probability of falling outside the interval $(0,1)$.

Thus $f_Y(y) = 0$ outside $(0, \pi)$

For any $0 < y < 1$, from Fig.,



the equation $y = \sin x$ has an infinite number of solutions $\cdots, x_1, x_2, x_3, \cdots,$ where $x_1 = \sin^{-1} y$ is the principal solution.

Moreover, using the symmetry we also get $x_2 = \pi - x_1$ etc.

Further,

$$\frac{dy}{dx} = \cos x = \sqrt{1 - \sin^2 x} = \sqrt{1 - y^2}$$

so that

$$\left|\frac{dy}{dx}\right|_{x=x_i} = \sqrt{1 - y^2}.$$

Using this in

$$f_Y(y) = \sum_i \frac{1}{|dy/dx|_{x_i}} f_X(x_i) = \sum_i \frac{1}{|g'(x_i)|} f_X(x_i)$$ we obtain for $0 < y < 1$,

$$f_Y(y) = \sum_{i=-\infty}^{+\infty} \frac{1}{\sqrt{1 - y^2}} f_X(x_i).$$

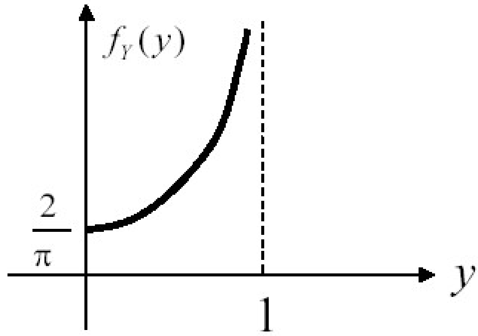
But from Fig., in this case $f_X(x_{-1}) = f_X(x_3) = f_X(x_4) = \cdots = 0$

(Except for $f_X(x_1)$ and $f_X(x_2)$ the rest are all zeros).

Thus

$$f_Y(y) = \frac{1}{\sqrt{1-y^2}}\big(f_X(x_1) + f_X(x_2)\big) = \frac{1}{\sqrt{1-y^2}}\left(\frac{2x_1}{\pi^2} + \frac{2x_2}{\pi^2}\right)$$

$$= \frac{2(x_1 + \pi - x_1)}{\pi^2\sqrt{1-y^2}} = \begin{cases} \dfrac{2}{\pi\sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$



Functions of a discrete-type r.v

Suppose X is a discrete-type r.v with

$P(X = x_i) = p_i, \quad x = x_1, x_2, \cdots, x_i, \cdots$ and

$Y = g(X).$

Clearly Y is also of discrete-type, and when $x = x_i, \ \ y_i = g(x_i),$ and for those y_i

$P(Y = y_i) = P(X = x_i) = p_i, \quad y = y_1, y_2, \cdots, y_i, \cdots$

Example: $X \sim P(\lambda)$, so that

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots$$

Define $Y = X^2 + 1$. Find the p.m.f of Y.

X takes the values $0, 1, 2, \cdots, k, \cdots$ so that Y only takes the value $1, 3, \cdots, k^2 + 1, \cdots$ and

$$P(Y = k^2 + 1) = P(X = k)$$

so that for $j = k^2 + 1$

$$P(Y = j) = P\left(X = \sqrt{j-1}\right) = e^{-\lambda} \frac{\lambda^{\sqrt{j-1}}}{(\sqrt{j-1})!}, \quad j = 1, 3, \cdots, k^2 + 1, \cdots.$$

Mean, Variance

Mean or the **Expected Value** of a r.v X is defined as

$$\eta_X = \overline{X} = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x) dx.$$

If X is a discrete-type r.v, then using

$f_X(x) = \sum_i p_i \delta(x - x_i),$ we get

$$\eta_X = \overline{X} = E(X) = \int x \sum_i p_i \delta(x - x_i) dx = \sum_i x_i p_i \underbrace{\int \delta(x - x_i) dx}_{1}$$

$$= \sum_i x_i p_i = \sum_i x_i\, P(X = x_i).$$

Mean represents the average (mean) value of the r.v in a very large number of trials.

For example if $X \sim U(a, b)$, (U representing the Uniform Distribution) Then using

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

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2}\bigg|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

is the midpoint of the interval (a,b).

If X is exponential with parameter λ as

$$f_X(x) = \begin{cases} \dfrac{1}{\lambda} e^{-x/\lambda}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

then $E(X) = \int_0^\infty \frac{x}{\lambda} e^{-x/\lambda} dx = \lambda \int_0^\infty y e^{-y} dy = \lambda,$

implying that the parameter λ represents the mean value of the exponential r.v.

Similarly if X is Poisson with parameter λ as

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0,1,2,\cdots,\infty.$$

Using

$$\eta_X = \sum_i x_i p_i = \sum_i x_i \, P(X = x_i).$$

we get

$$E(X) = \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Thus the parameter λ also represents the mean of the Poisson r.v

For the normal r.v

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

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (y+\mu) e^{-y^2/2\sigma^2} dy$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{\int_{-\infty}^{+\infty} y e^{-y^2/2\sigma^2} dy}_{0} + \mu \cdot \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-y^2/2\sigma^2} dy}_{1} = \mu.$$

Thus the first parameter in $X \sim N(\mu, \sigma^2)$ is the mean of the Gaussian r.v X.

Given $X \sim f_X(x)$, let $Y = g(X)$ defines a new r.v with p.d.f $f_Y(y)$. Then the new r.v Y has a mean μ_Y given by

$$\mu_Y = E(Y) = \int_{-\infty}^{+\infty} y \, f_Y(y) dy.$$

It appears that to determine $E(Y)$, we need to determine $f_Y(y)$.

However this is not the case if only $E(Y)$ is the quantity of interest.

The following formula can be derived

$$\boxed{E(Y) = E\big(g(X)\big) = \int_{-\infty}^{+\infty} y \, f_Y(y) dy = \int_{-\infty}^{+\infty} g(x) f_X(x) dx.}$$

In the discrete case, the above formula reduces to

$$\boxed{E(Y) = \sum_i g(x_i) P(X = x_i).}$$

Thus, $f_Y(y)$ is not required to evaluate $E(Y)$ for $Y = g(X)$.

Example: Determine the mean of $Y = X^2$, where X is Poisson r.v. Using

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0,1,2,\cdots,\infty.$$

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

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

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

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

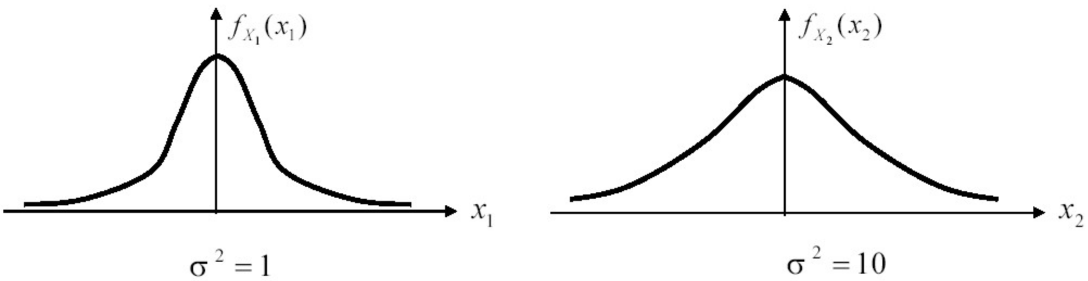
$$= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda} \right) = \lambda^2 + \lambda.$$

In general, $E(X^k)$ is known as the *k*th moment of r.v X. Thus if $X \sim P(\lambda)$, its second moment is given by

$$E\left(X^2\right) = \lambda e^{-\lambda}\left(\lambda e^{\lambda} + e^{\lambda}\right) = \lambda^2 + \lambda .$$

Mean alone will not be able to truly represent the p.d.f of any r.v. To illustrate this, consider the following scenario: Consider two Gaussian r.vs $X_1 \sim N(0,1)$ and $X_2 \sim N(0,10)$. Both of them have the same mean $\mu = 0$.

However, as the below Fig. shows, their p.d.fs are different.



One is more concentrated around the mean, whereas the other one (X_2) has a wider spread.

Clearly, we need atleast an additional parameter to measure this spread around the mean.

For a r.v X with mean μ, $X - \mu$ represents the deviation of the r.v from its mean.

Since this deviation can be either positive or negative, consider the quantity

$(X - \mu)^2$

and its average value

$E[(X - \mu)^2]$

represents the average mean square deviation of X around its mean.

Define

$$\sigma_X^2 \stackrel{\Delta}{=} E[(X-\mu)^2] > 0.$$

With $g(X) = (X-\mu)^2$ and using

$$E(Y) = E(g(X)) = \int_{-\infty}^{+\infty} y\, f_Y(y)dy = \int_{-\infty}^{+\infty} g(x) f_X(x)dx.$$

we get

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx > 0.$$

σ_X^2 is known as the variance of the r.v X, and its square root $\sigma_X = \sqrt{E(X-\mu)^2}$ is known as the standard deviation of X. Note that the standard deviation represents the root mean square spread of the r.v X around its mean μ.

Expanding

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx > 0.$$

and using the linearity of the integrals, we get

$$\begin{aligned} Var(X) = \sigma_X^2 &= \int_{-\infty}^{+\infty} \left(x^2 - 2x\mu + \mu^2\right) f_X(x)dx \\ &= \int_{-\infty}^{+\infty} x^2 f_X(x)dx - 2\mu \int_{-\infty}^{+\infty} x\, f_X(x)dx + \mu^2 \\ &= E\left(X^2\right) - 2\mu\mu + \mu^2 = E\left(X^2\right) - \mu^2 \\ &= E\left(X^2\right) - [E(X)]^2 = \overline{X^2} - \overline{X}^2. \end{aligned}$$

Alternatively, we can use

$$\sigma_X^2 = E\left(X^2\right) - [E(X)]^2 = \overline{X^2} - \overline{X}^2.$$

to compute σ_X^2.

Example : For Poisson r.v

$$\sigma_X^2 = \overline{X^2} - \overline{X}^2 = \left(\lambda^2 + \lambda\right) - \lambda^2 = \lambda.$$

Thus for a Poisson r.v, mean and variance are both equal to its parameter λ.

To determine the variance of the normal r.v $N(\mu,\sigma^2)$, we can use

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx.$$

To simplify , we can make use of the identity

$$\int_{-\infty}^{+\infty} f_X(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = 1$$

for a normal p.d.f. This gives

$$\int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}\,\sigma.$$

Differentiating both sides with respect to σ, we get

$$\int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma^3} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}$$

or

$$\int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \sigma^2,$$

Thus for a normal r.v as in

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

$$Var(X) = \sigma^2$$

and the second parameter in $N(\mu,\sigma^2)$ represents the variance of the Gaussian r.v.

The larger the σ, the larger the spread of the p.d.f around its mean.

Thus as the variance of a r.v tends to zero, it will begin to concentrate more and more around the mean ultimately behaving like a constant.

Moments

In general

$$m_n = \overline{X^n} = E(X^n), \quad n \geq 1$$

are known as the moments of the r.v X, and

$$\mu_n = E[(X - \mu)^n]$$

are known as the central moments of X.

The mean is

$$\mu = m_1$$

and the variance is

$$\sigma^2 = \mu_2$$

Generalized moments of X about a, are

$$E[(X - a)^n]$$

Absolute moments of X, are

$$E[|X|^n]$$

For example, if $X \sim N(0, \sigma^2)$, then it can be shown that

$$E(X^n) = \begin{cases} 0, & n \text{ odd,} \\ 1 \cdot 3 \cdots (n-1)\sigma^n, & n \text{ even.} \end{cases}$$

$$E(|X|^n) = \begin{cases} 1 \cdot 3 \cdots (n-1)\sigma^n, & n \text{ even,} \\ 2^k k! \sigma^{2k+1} \sqrt{2/\pi}, & n = (2k+1), \text{ odd.} \end{cases}$$

To compute the mean and variance using the generalized moments of X and absolute moments of X is often difficult, however, using characteristic function can be helpful.

Characteristic Functions

Characteristic function of a r.v X is defined as

$$\Phi_X(\omega) \triangleq E\left(e^{jX\omega}\right) = \int_{-\infty}^{+\infty} e^{jx\omega} f_X(x)dx.$$

Thus $\Phi_X(0) = 1$, and $|\Phi_X(\omega)| \le 1$ for all ω.

For discrete r.v s the characteristic function reduces to

$$\Phi_X(\omega) = \sum_k e^{jk\omega} P(X = k).$$

Example: If $X \sim P(\lambda)$ then its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} e^{jk\omega} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{j\omega})^k}{k!} = e^{-\lambda} e^{\lambda e^{j\omega}} = e^{\lambda(e^{j\omega}-1)}.$$

It can be shown that

$$\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0} = jE(X)$$

or $$E(X) = \frac{1}{j} \left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0}.$$

Similarly,

$$E(X^2) = \frac{1}{j^2} \left.\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2}\right|_{\omega=0}$$

Repeating this procedure k times, we obtain the k^{th} moment of X as

$$E(X^k) = \frac{1}{j^k} \frac{\partial^k \Phi_X(\omega)}{\partial \omega^k}\bigg|_{\omega=0}, \quad k \geq 1.$$

Thus characteristic functions are used to compute the mean, variance and higher order moments of any random variable *X*

Example:

If $X \sim P(\lambda)$ is Poisson then

$$\Phi_X(\omega) = e^{\lambda(e^{j\omega}-1)}$$

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = e^{-\lambda} e^{\lambda e^{j\omega}} \lambda j e^{j\omega},$$

$$E(X) = \frac{1}{j} \frac{\partial \Phi_X(\omega)}{\partial \omega}\bigg|_{\omega=0}$$

$$E(X) = \lambda$$

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = e^{-\lambda}\left(e^{\lambda e^{j\omega}} (\lambda j e^{j\omega})^2 + e^{\lambda e^{j\omega}} \lambda j^2 e^{j\omega}\right)$$

$$E(X^2) = \frac{1}{j^2} \frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2}\bigg|_{\omega=0}$$

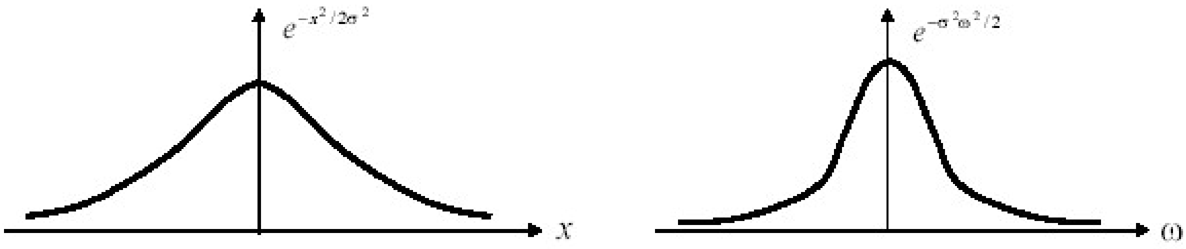
$$E(X^2) = \lambda^2 + \lambda$$

The characteristic function of a Gaussian r.v itself has the "Gaussian" bell shape. Thus if $X \sim N(0, \sigma^2)$, then

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

and

$$\Phi_X(\omega) = e^{-\sigma^2\omega^2/2}$$



Note that σ^2 in $f_X(x)$ and $\Phi_X(\omega)$ are reverse (σ^2 vs $\frac{1}{\sigma^2}$) .

Two Random Variables

In many experiments, the observations are expressible not as a single quantity, but as a family of quantities.

Example:

To record the height and weight of each person in a community, we need two numbers

or

To record the number of people and the total income in a family, we need two numbers.

Let *X* and *Y* denote two random variables (r.v) based on a probability model (Ω, *F*, *P*). Then

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

and

$$P\left(y_1 < Y(\xi) \le y_2\right) = F_Y(y_2) - F_Y(y_1) = \int_{y_1}^{y_2} f_Y(y)dy$$

What about the probability that the pair of r.vs (X,Y) belongs to an arbitrary region D? In other words, how does one estimate, for example, $P\left[(x_1 < X(\xi) \le x_2) \cap (y_1 < Y(\xi) \le y_2)\right] = ?$ For this we define the joint probability distribution function of X and Y to be

$$F_{XY}(x,y) = P\left[(X(\xi) \le x) \cap (Y(\xi) \le y)\right]$$
$$= P(X \le x, Y \le y) \ge 0,$$

where x and y are arbitrary real numbers.

Properties of the joint probability distribution function

(i) $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0,$

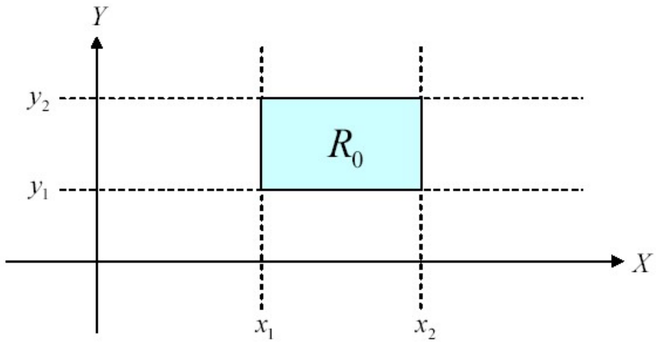
$F_{XY}(+\infty, +\infty) = 1$

(ii) $P\left(x_1 < X(\xi) \le x_2, Y(\xi) \le y\right) = F_{XY}(x_2, y) - F_{XY}(x_1, y)$

$P\left(X(\xi) \le x, y_1 < Y(\xi) \le y_2\right) = F_{XY}(x, y_2) - F_{XY}(x, y_1)$

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

This is the probability that (X,Y) belongs to the rectangle R_0 in the below Figure



Joint probability density function (Joint p.d.f)

By definition, the joint p.d.f of X and Y is given by

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

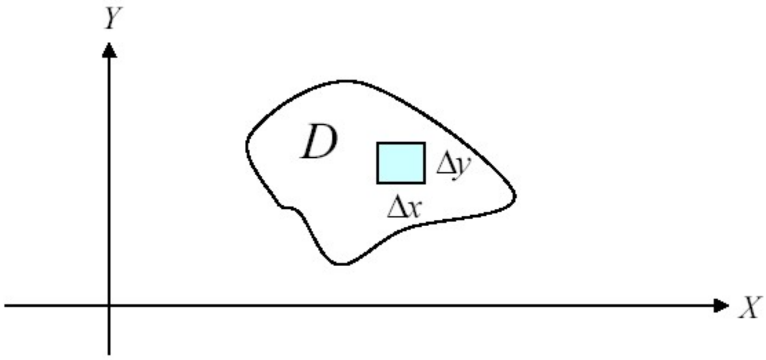
and

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

Also

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

The probability of (X,Y) Є D (where D is given below Figure) is given by



$$P\left((X, Y) \in D\right) = \int\int_{(x,y)\in D} f_{XY}(x, y)\, dxdy$$

Marginal Statistics

In case when there are several r.vs, the statistics of each individual ones are called marginal statistics.

Thus F_X (x) is the marginal probability distribution function of *X*, and f_X (x) is the marginal p.d.f of *X*.

All marginals can be obtained from the joint p.d.f.

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

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

Also

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

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dx$$

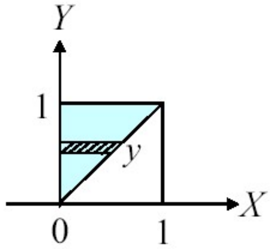
If X and Y are discrete r.vs, then $p_{ij} \triangleq P(X = x_i, Y = y_j)$ represents their joint p.d.f, and their respective marginal p.d.fs are given by

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}$$

and

$$P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij}$$

Example : Given



$$f_{XY}(x, y) = \begin{cases} \text{constant}, & 0 < x < y < 1, \\ 0, & \text{otherwise} \end{cases} .$$

Obtain the marginal p.d.fs $f_X(x)$ and $f_Y(y)$.

It is given that the joint p.d.f $f_{XY}(x, y)$ is a constant in the shaded region in Fig. To determine the constant c, we use

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

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dxdy = \int_{y=0}^{1} \left(\int_{x=0}^{y} c \cdot dx \right) dy = \int_{y=0}^{1} cy dy = \frac{cy^2}{2} \Bigg|_0^1 = \frac{c}{2} = 1 .$$

Thus $c = 2$.

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \int_{y=x}^{1} 2 dy = 2(1 - x), \quad 0 < x < 1,$$

and similarly

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \int_{x=0}^{y} 2 dx = 2y, \quad 0 < y < 1.$$

Independence of r.vs

The random variables X and Y are statistically independent if the events $\{X(\xi) \in A\}$ and $\{Y(\xi) \in B\}$ are independent events for any two sets A and B in x and y axes respectively.

$$P\big((X(\xi) \le x) \cap (Y(\xi) \le y)\big) = P(X(\xi) \le x)P(Y(\xi) \le y)$$

i.e.,

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

or equivalently, if X and Y are independent, then we must have

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

If X and Y are discrete-type r.vs then their independence implies

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \text{for all } i, j.$$

The procedure to test for independence.

Given $f_{XY}(x, y)$, obtain the marginal p.d.fs $f_X(x)$ and $f_Y(y)$ and examine whether

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

or

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \text{for all } i, j.$$

If so, the r.vs are independent, otherwise they are dependent.

Joint Moments

Given two r.vs X and Y and a function g(x,y). Define the r.v

Z = g(x,y)

Define the mean of Z as

$$\mu_Z = E(Z) = \int_{-\infty}^{+\infty} z\, f_Z(z)dz$$

it is possible to express the mean of Z = g(x,y) in terms of $f_{XY}(x,y)$ without computing $f_Z(z)$.

$$E(z) = \int_{-\infty}^{+\infty} z\, f_Z(z)dz = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y)dxdy.$$

or

$$E[g(x,y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y)dxdy$$

If X and Y are discrete-type r.vs, then

$$E[g(x,y)] = \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j).$$

If X and Y are independent r.vs, then

Z = g(x), W = h(y)

are always independent of each other. Thus <u>if X and Y are independent r.vs</u>, then

$$E[g(X)h(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y) f_X(x) f_Y(y)dxdy$$
$$= \int_{-\infty}^{+\infty} g(x) f_X(x)dx \int_{-\infty}^{+\infty} h(y) f_Y(y)dy = E[g(X)]E[h(Y)]$$

In the case of one random variable, we defined the parameters mean and variance to represent its average behavior. How can we parametrically represent similar cross-behavior between two random variables?

<u>Covariance</u>

Given <u>any</u> two r.v s X and Y, we define the covariance as

$$Cov(X,Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$
$$Cov(X,Y) = E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y)$$
$$= \overline{XY} - \overline{X}\,\overline{Y}.$$

Let ***U*** = ***X*** + ***Y***

$$Var(U) = E\left[\{(X - \mu_X) + (Y - \mu_Y)\}^2\right]$$
$$= Var(X) + 2\,Cov(X,Y) + Var(Y)$$

Defining ρ_{XY} as the correlation coefficient between *X* and *Y* where

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \rho_{XY} \le 1$$

or

$$Cov(X,Y) = \rho_{XY}\sigma_X\sigma_Y$$

<u>Uncorrelated r.v s</u>

If the correlation coefficient between *X* and *Y,* i.e., $\rho_{XY} = 0$ then X and Y are uncorrelated. When uncorrelated then

$$Cov(X,Y) = \overline{XY} - \overline{X}\,\overline{Y} = \rho_{XY}\sigma_X\sigma_Y = 0$$

i.e.,

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

<u>Orthogonality</u>

X and Y are said to be orthogonal if

$$E(XY) = 0$$

If two random variables are statistically independent, then there cannot be any correlation between them.

However, the converse is in general not true.

Random variables can be uncorrelated without being independent.

Example : Let $Z = aX + bY$. Determine the mean of Z in terms of μ_X and μ_Y. Also find the variance of Z in terms of σ_X, σ_Y and ρ_{XY}.

$$\mu_Z = E(z) = E(aX + bY) = a\mu_X + b\mu_Y$$

and using $Cov(X,Y) = \rho_{XY}\sigma_X\sigma_Y$

$$\sigma_Z^2 = Var(z) = E\left[(Z-\mu_Z)^2\right] = E\left[\left(a(X-\mu_X) + b(Y-\mu_Y)\right)^2\right]$$
$$= a^2E(X-\mu_X)^2 + 2abE\left((X-\mu_X)(Y-\mu_Y)\right) + b^2E(Y-\mu_Y)^2$$
$$= a^2\sigma_X^2 + 2ab\rho_{XY}\sigma_X\sigma_Y + b^2\sigma_Y^2.$$

In particular if X and Y are independent, then $\rho_{XY} = 0$, and

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2.$$

Thus the variance of the sum of independent r.vs is the sum of their variances $(a = b = 1)$.

In general

$$E[X^kY^m] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k y^m f_{XY}(x,y)\,dx\,dy,$$

represents the joint moment of order (k,m) for X and Y.

Conditional Distributions

Previously we have seen that the distribution function of X given an event B is

$$F_X(x \mid B) = P\left(X(\xi) \le x \mid B\right) = \frac{P\left((X(\xi) \le x) \cap B\right)}{P(B)}$$

Suppose, we let $B = \{y_1 < Y(\xi) \le y_2\}$

Then

$$F_X(x \mid y_1 < Y \le y_2) = \frac{P\left(X(\xi) \le x, y_1 < Y(\xi) \le y_2\right)}{P(y_1 < Y(\xi) \le y_2)}$$
$$= \frac{F_{XY}(x, y_2) - F_{XY}(x, y_1)}{F_Y(y_2) - F_Y(y_1)}$$

$$F_X(x \mid y_1 < Y \le y_2) = \frac{\int_{-\infty}^{x} \int_{y_1}^{y_2} f_{XY}(u, v)\,du\,dv}{\int_{y_1}^{y_2} f_Y(v)\,dv}$$

To determine, the limiting case $F_X(x \mid Y = y)$, we can let $y_1 = y$ and $y_2 = y + \Delta y$

This gives

$$F_X(x \mid y < Y \le y + \Delta y) = \frac{\int_{-\infty}^{x} \int_{y}^{y+\Delta y} f_{XY}(u, v)\,du\,dv}{\int_{y}^{y+\Delta y} f_Y(v)\,dv} \approx \frac{\int_{-\infty}^{x} f_{XY}(u, y)\,du\,\Delta y}{f_Y(y)\,\Delta y}$$

and hence in the limit

$$F_X(x \mid Y = y) = \lim_{\Delta y \to 0} F_X(x \mid y < Y \le y + \Delta y) = \frac{\int_{-\infty}^{x} f_{XY}(u, y)\,du}{f_Y(y)}.$$

(To remind about the conditional nature on the left hand side, we shall use the subscript $X \mid Y$ (instead of X)).

Thus

$$F_{X|Y}(x \mid Y = y) = \frac{\int_{-\infty}^{x} f_{XY}(u, y)\,du}{f_Y(y)}.$$

Differentiating with respect to x , we get

$$f_{X|Y}(x \mid Y = y) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

$f_X(x \mid Y = y)$ represents a valid probability density function so

$$f_X(x \mid Y = y) = \frac{f_{XY}(x, y)}{f_Y(y)} \ge 0$$

and

$$\int_{-\infty}^{+\infty} f_{X|Y}(x \mid Y = y)\,dx = \frac{\int_{-\infty}^{+\infty} f_{XY}(x, y)\,dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$$

we shall refer to $f_{X|Y}(x \mid Y=y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$ as the conditional p.d.f of the r.v X given $Y=y$

We may also write

$$f_{X|Y}(x \mid Y=y) = f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

similarly

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

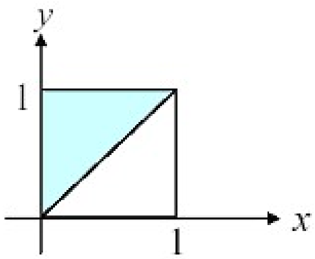
If the r.vs X and Y are independent, then $f_{XY}(x,y) = f_X(x) f_Y(y)$ and

$$f_{X|Y}(x \mid y) = f_X(x), \quad f_{Y|X}(y \mid x) = f_Y(y),$$

implying that the conditional p.d.fs coincide with their unconditional p.d.fs. This makes sense, since if X and Y are independent r.vs, information about Y shouldn't be of any help in updating our knowledge about X.

In the case of discrete-type r.vs,

$$P\left(X = x_i \mid Y = y_j\right) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$



Example: Given

$$f_{XY}(x,y) = \begin{cases} k, & 0 < x < y < 1, \\ 0, & \text{otherwise}, \end{cases}$$

determine $f_{X|Y}(x \mid y)$ and $f_{Y|X}(y \mid x)$.

The joint p.d.f is given to be a constant in the shaded region.

This gives

$$\int\int f_{XY}(x,y)dxdy = \int_0^1 \int_0^y k\ dx\ dy = \int_0^1 k\ y\ dy = \frac{k}{2} = 1 \Rightarrow k = 2.$$

Similarly

$$f_X(x) = \int f_{XY}(x,y)dy = \int_x^1 k\ dy = k\ (1-x), \quad 0 < x < 1,$$

and

$$f_Y(y) = \int f_{XY}(x,y)dx = \int_0^y k\ dx = k\ y, \quad 0 < y < 1.$$

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{y}, \quad 0 < x < y < 1,$$

and

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{1-x}, \quad 0 < x < y < 1.$$

p.d.f version of Bayes' theorem

Using $f_{X|Y}(x \mid y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$ and $f_{Y|X}(y \mid x) = \dfrac{f_{XY}(x,y)}{f_X(x)}$

$$f_{XY}(x,y) = f_{X|Y}(x \mid y) f_Y(y) = f_{Y|X}(y \mid x) f_X(x)$$

or

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{f_X(x)}$$

But

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_{-\infty}^{+\infty} f_{X|Y}(x \mid y) f_Y(y)dy$$

$$\boxed{f_{YX}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y}(x \mid y) f_Y(y)dy}}$$

This equation represents the p.d.f version of Bayes' theorem.

Mean Square Estimation

Given some information that is related to an unknown quantity of interest

The problem is to obtain a good estimate for the unknown in terms of the observed data.

Suppose X_1, X_2, ... , X_n represent a sequence of random variables about whom one set of observations are available, and

Y represents an unknown random variable.

The problem is to obtain a good estimate for Y in terms of the observations $X_1, X_2, \ldots, X_n$

Let

$$\hat{Y} = \varphi(X_1, X_2, \cdots, X_n) = \varphi(\underline{X})$$

represent such an estimate for Y.

Note that $\varphi(.)$ can be a linear or a nonlinear function of the observation $X_1, X_2, \ldots, X_n$

$$\varepsilon(\underline{X}) = Y - \hat{Y} = Y - \varphi(\underline{X})$$

represents the error in the above estimate, and

$$|\varepsilon|^2$$

is the square of the error.

Since ε is a random variable,

$$E\{|\varepsilon|^2\}$$

represents the mean square error.

One way of obtaining a good estimator is to minimize the mean square error by varying over all possible forms of the estimator $\varphi(.)$ and

This procedure gives rise to the **M**inimization of the **M**ean **S**quare **E**rror (MMSE) criterion for estimation.

Thus under MMSE criterion, the estimator φ(.) is chosen such that the mean square error

$$E\{|\varepsilon|^2\}$$

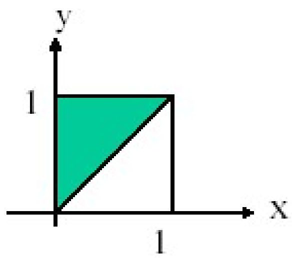
is at its minimum.

Theorem: (Without Proof): Under MMSE criterion, the best estimator for the unknown Y in terms of $X_1, X_2, \ldots, X_n$ is given by the conditional mean of Y given $\underline{X}$. Thus

$$\hat{Y} = \varphi(\underline{X}) = E\{Y \mid \underline{X}\}$$

Example : Let

$$f_{X,Y}(x,y) = \begin{cases} kxy, & 0 < x < y < 1 \\ 0 & \text{otherwise,} \end{cases}$$



where $k > 0$ is a suitable normalization constant. The best estimate for Y in terms of X is

$$\hat{Y} = \varphi(X) = E\{Y \mid \underline{X}\} = \int_x^1 y\, f_{Y|X}(y \mid x)\,dy$$

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$f_X(x) = \int_x^1 f_{X,Y}(x,y)\,dy = \int_x^1 kxy\,dy = \left.\frac{kxy^2}{2}\right|_x^1 = \frac{kx(1-x^2)}{2}, \quad 0 < \mathrm{x} < 1$$

Thus

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{kxy}{kx(1-x^2)/2} = \frac{2y}{1-x^2}; \quad 0 < x < y < 1.$$

Hence the best MMSE estimator is given by

$$\hat{Y} = \varphi(X) = E\{Y \mid \underline{X}\} = \int_x^1 y\, f_{Y|X}(y \mid x)dy$$
$$= \int_x^1 y \tfrac{2y}{1-x^2} dy = \tfrac{2}{1-x^2} \int_x^1 y^2 dy$$
$$= \left(\frac{2}{3}\right)\frac{y^3}{1-x^2}\Bigg|_x^1 = \left(\frac{2}{3}\right)\frac{1-x^3}{1-x^2} = \frac{2}{3}\frac{(1+x+x^2)}{1-x^2}.$$

Sequences of Random Variables and Central Limit Theorem

Let X_1, X_2, X_3,... be a sequence of random variables which are defined on the same probability space, share the same probability distribution D and are independent.

Assume that both the expected value μ and the standart deviation σ of D exist and are finite.

Consider the sum:

$S_n = X_1 + ... + X_n$.

Then the expected value of S_n is $n\mu$

and its standard deviation is $\sigma\, n^{1/2}$

Furthermore, the distribution of S_n approaches the normal distribution N ($n\mu$, $\sigma^2 n$) as n approaches ∞.

In order to clarify the word "approaches" in the last sentence, we standardize S_n by setting

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Then the distribution of Z_n converges towards the standard normal distribution N(0,1) as n approaches ∞.

This means, if $F_Z(z)$ is the cumulative distribution function of N(0,1), then for every real number z, we have

$$\lim_{n\to\infty} \mathrm{P}(Z_n \le z) = F_Z(z)$$

or, equivalently,

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = F_z(z)$$

where

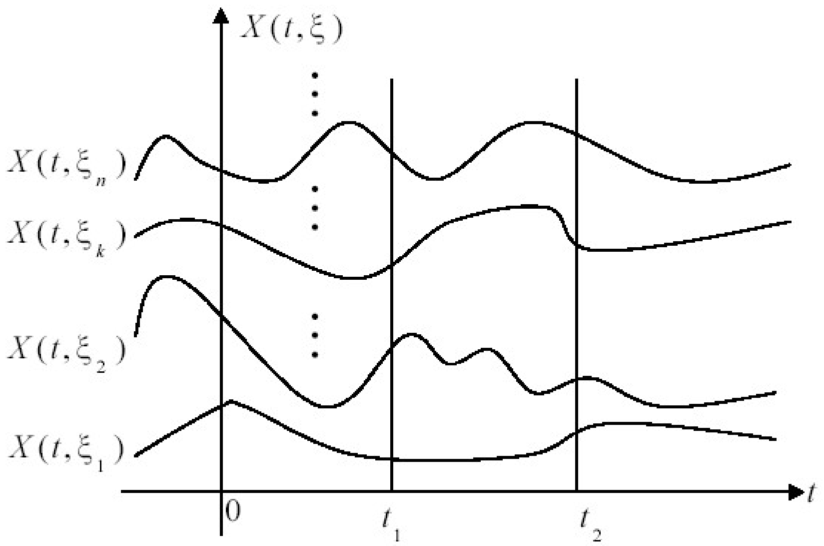
$$\overline{X}_n = S_n/n = (X_1 + \cdots + X_n)/n$$

is the "sample mean".

Stochastic Processes

Let ξ denote the random outcome of an experiment.

To every such outcome suppose a waveform $X(t, \xi)$ *is* assigned.



The collection of such waveforms form a stochastic process.

The set of $\{ \xi_k \}$ and the time index t can be continuousor discrete (countably infinite or finite) as well.

For fixed $\xi_i \in S$ (the set of all experimental outcomes), $X(t, \xi)$ is a specific time function.

For fixed t,

$X_1(t_1, \xi_i)$ is a random variable.

The ensemble of all such realizations $X(t, \xi)$ over time represents the stochastic process $X(t)$.

Example:

$$X(t) = a\cos(\omega_0 t + \varphi)$$

where φ is a uniformly distributed random variable in $(0, 2\pi)$, represents a stochastic process.

If $X(t)$ is a stochastic process, then for fixed t, $X(t)$ represents a random variable.

Its distribution function is given by

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

Notice that $F_X(x,t)$ depends on t, since for a different t, we obtain a different random variable. Further

$$f_X(x,t) \triangleq \frac{dF_X(x,t)}{dx}$$

represents the first-order probability density function of the process $X(t)$.

For $t = t_1$ and $t = t_2$, $X(t)$ represents two different random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$ respectively.

For $t = t_1$ and $t = t_2$, $X(t)$ represents two different random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$ respectively. Their joint distribution is given by

$$F_X(x_1, x_2, t_1, t_2) = P\{X(t_1) \le x_1, X(t_2) \le x_2\}$$

and

$$f_X(x_1, x_2, t_1, t_2) \triangleq \frac{\partial^2 F_X(x_1, x_2, t_1, t_2)}{\partial x_1 \, \partial x_2}$$

represents the second-order density function of the process $X(t)$.

Similarly

$f_X(x_1, x_2, \cdots x_n,\ t_1, t_2 \cdots, t_n)$

represents the n[th] order density function of the process $X(t)$.

Complete specification of the stochastic process $X(t)$ requires the knowledge of $f_X(x_1, x_2, \cdots x_n,\ t_1, t_2 \cdots, t_n)$ for all t_i, $\quad i = 1, 2, \cdots, n$ and for all n. (an almost impossible task in reality).

Mean of a Stochastic Process:

$$\mu(t) \stackrel{\Delta}{=} E\{X(t)\} = \int_{-\infty}^{+\infty} x f_X(x,t)dx$$

represents the mean value of a process $X(t)$.

In general, the mean of a process can depend on the time index t.

Autocorrelation function of a process $X(t)$ is defined as

$$R_{XX}(t_1, t_2) \stackrel{\Delta}{=} E\{X(t_1)X^*(t_2)\} = \int\int x_1 x_2^* f_X(x_1, x_2, t_1, t_2)dx_1 dx_2$$

and it represents the interrelationship between the random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$ generated from the process $X(t)$.

Properties of Autocorrelation function

1. $R_{XX}(t_1, t_2) = R_{XX}^*(t_2, t_1) = [E\{X(t_2)X^*(t_1)\}]^*$

2. $R_{XX}(t,t) = E\{|X(t)|^2\} > 0.$ (Average instantaneous power)

3. $R_{XX}(t_1, t_2)$ represents a nonnegative definite function, i.e., for *any* set of constants $\{a_i\}_{i=1}^n$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* R_{XX}(t_i, t_j) \geq 0.$$

Autocovariance function

$$C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X^*(t_2) = R_{XX}(t_1, t_2) - E\{X(t_1)\}\, E\{X(t_2)\}^*$$

represents the autocovariance of the process $X(t)$.

Example

$X(t) = a\cos(\omega_0 t + \varphi), \qquad \varphi \sim U(0, 2\pi).$

This gives

$$\mu_X(t) = E\{X(t)\} = aE\{\cos(\omega_0 t + \varphi)\}$$
$$= a\cos\omega_0 t\, E\{\cos\varphi\} - a\sin\omega_0 t\, E\{\sin\varphi\}$$

since $E\{\cos\varphi\} = \frac{1}{2\pi}\int_0^{2\pi}\cos\varphi\, d\varphi = 0 = E\{\sin\varphi\}.$

$\mu_X(t) = 0,$

Similarly

$$R_{XX}(t_1, t_2) = a^2 E\{\cos(\omega_0 t_1 + \varphi)\cos(\omega_0 t_2 + \varphi)\}$$
$$= \frac{a^2}{2} E\{\cos\omega_0(t_1 - t_2) + \cos(\omega_0(t_1 + t_2) + 2\varphi)\}$$
$$= \frac{a^2}{2}\cos\omega_0(t_1 - t_2).$$

Stationary Stochastic Processes

Stationary processes exhibit statistical properties that are invariant to shift in the time index.

For a **second-order strict-sense stationary process** we have

$$f_X(x_1, x_2, t_1, t_2) \equiv f_X(x_1, x_2, t_1 + c, t_2 + c)$$

for any c. For $c = -t_2$ we get

$$f_X(x_1, x_2, t_1, t_2) \equiv f_X(x_1, x_2, t_1 - t_2)$$

i.e., the second order density function of a strict sense stationary process depends only on the difference of the time indices $t_1 - t_2 = \tau$.

In that case the autocorrelation function is given by

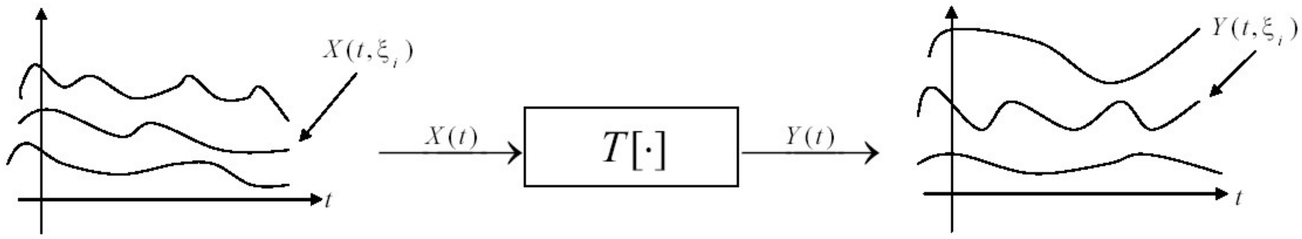
$$R_{XX}(t_1,t_2) \triangleq E\{X(t_1)X^*(t_2)\}$$
$$= \int\int x_1 x_2^* f_X(x_1,x_2,\tau = t_1 - t_2)dx_1 dx_2$$
$$= R_{XX}(t_1 - t_2) \triangleq R_{XX}(\tau) = R_{XX}^*(-\tau),$$

i.e., the autocorrelation function of a second order strict-sense stationary process depends only on the difference of the time indices $\tau = t_1 - t_2$.

Systems with Stochastic Inputs

A system transforms each input waveform $X(t,\xi_i)$ into an output waveform $Y(t,\xi_i) = T[X(t,\xi_i)]$ by operating only on the time variable t.

Thus a set of realizations at the input corresponding to a process $X(t)$ generates a new set of realizations $\{Y(t,\xi)\}$ at the output associated with a new process $Y(t)$.



Our goal is to study the output process statistics in terms of the input process statistics and the system function.

White Noise Process:

$W(t)$ is said to be a white noise process if

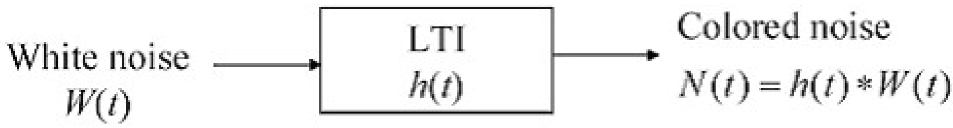
$R_{WW}(t_1,t_2) = q(t_1)\delta(t_1 - t_2)$, where $q(t_1)$ is the average noise power

i.e., $E[W(t_1)\, W^*(t_2)] = 0$ unless $t_1 = t_2$.

$W(t)$ is said to be wide-sense stationary (w.s.s) white noise if $E[W(t)]$ = constant, and

$R_{ww}(t_1, t_2) = q\delta(t_1 - t_2) = q\delta(\tau)$.

If $W(t)$ is also a Gaussian process (white Gaussian process), then all of its samples are independent random variables (why?).



LTI = Linear Time Invariant

For w.s.s. white noise input $W(t)$, we have

$E[N(t)] = \mu_W \int_{-\infty}^{+\infty} h(\tau)d\tau$, *a constant*

and

$$R_{nn}(\tau) = q\delta(\tau) * h^*(-\tau) * h(\tau)$$
$$= qh^*(-\tau) * h(\tau) = q\rho(\tau)$$

where

$$\rho(\tau) = h(\tau) * h^*(-\tau) = \int_{-\infty}^{+\infty} h(\alpha)h^*(\alpha + \tau)d\alpha.$$

Thus the output of a white noise process through an LTI system represents a (colored) noise process.

Note: White noise need not be Gaussian.
"White" and "Gaussian" are two different concepts

Power Spectrum

In signal theory, spectra are associated with Fourier transforms.

For deterministic signals, spectra are used to represent a function in terms of exponentials.

For random signals, spectrum has two interpretations:

- The first involves transforms of averages (We will work on this case)
- The second is the representation of the process as superposition of exponentials with random coefficients

Wide Sense Stationary

A stochastic process $\boldsymbol{x}(t)$ is called wide sense stationary (wss) if its mean is a constant

$$E\{ \boldsymbol{x}(t) \} = \eta$$

And its autocorrelation depends only on $\tau = t_1 - t_2$

$$E\{ \boldsymbol{x}(t+\tau)\, \boldsymbol{x}^*(t) \} = R(\tau)$$

Since τ is the distance from t to $t + \tau$, the function $R(\tau)$ can be written in the symmetrical form

$$R(\tau) = E\{ \boldsymbol{x}[t + (\tau/2)]\, \boldsymbol{x}^*[t - (\tau/2)] \}$$

In particular

$$E\{ |\boldsymbol{x}(t)|^2 \} = R(0)$$

Thus the average power of a stationary process is independent of time t and is equal to $R(0)$

Power Spectrum or Spectral Density of a WSS process $\boldsymbol{x}(t)$, real or complex, is the Fourier transform $S(\omega)$ of its autocorrelation $E\{ \boldsymbol{x}(t+\tau)\, \boldsymbol{x}^*(t) \}$. İ.e.,

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau) \exp(-j\omega\tau)\, d\tau$$

Since $R(-\tau) = R^*(\tau)$, $S(\omega)$ is a real function of ω

From the Fourier inversion formula,

$$R(\tau) = [1/(2\pi)] \int_{-\infty}^{\infty} S(\omega) \exp(j\omega\tau)\, d\omega$$

If $\boldsymbol{x}(t)$ is a real process, then $R(\tau)$ is real and even

Hence $S(\omega)$ is also real and even. In this case

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau) \cos(\omega\tau)\, d\tau = 2\int_{0}^{\infty} R(\tau) \cos(\omega\tau)\, d\tau$$

$$R(\tau) = [1/(2\pi)] \int_{-\infty}^{\infty} S(\omega) \cos(\omega\tau)\, d\omega = (1/\pi) \int_{0}^{\infty} S(\omega) \cos(\omega\tau)\, d\omega$$

The cross-power spectrum of two processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$ is the Fourier transform $S_{xy}(\omega)$ of their cross-correlation

$R_{xy}(\tau) = E\{\mathbf{x}(t+\tau)\,\mathbf{y}^*(t)\}$; i.e.,

$$S_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau) \exp(-j\omega\tau)\, d\tau$$

$$R_{xy}(\tau) = [1/(2\pi)] \int_{-\infty}^{\infty} S_{xy}(\omega) \exp(j\omega\tau)\, d\omega$$

In general, $S_{xy}(\omega)$ is complex even when both processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are real.

In all cases

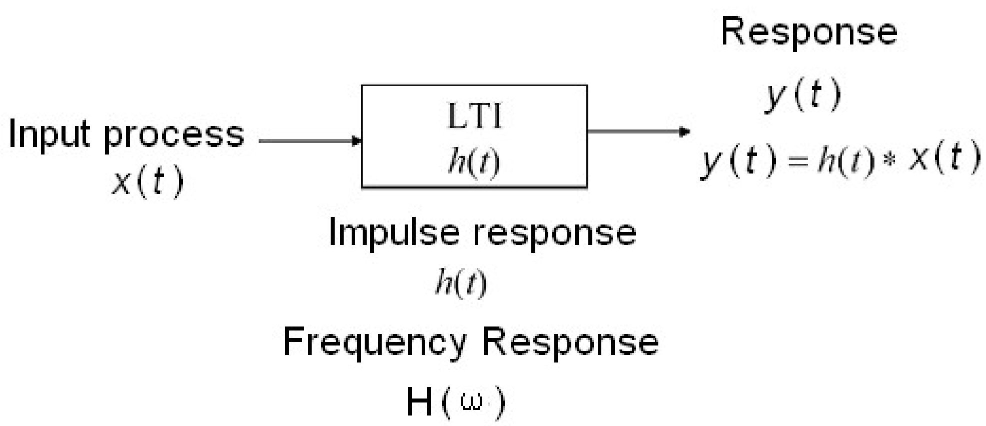
$S_{xy}(\omega) = S^*_{YX}(\omega)$

Since

$R_{xy}(-\tau) = E\{\mathbf{x}(t-\tau)\,\mathbf{y}^*(t)\} = R^*_{YX}(\tau)$

Also $S_{xy}(\omega) > 0$ for every spectrum

<u>Linear Systems</u>



Express the autocorrelation $R_{yy}(-\tau)$ and power spectrum $S_{yy}(\omega)$ of the response

$$\mathbf{y}(t) = \int_{-\infty}^{\infty} \mathbf{x}(t - \alpha)\, h(\alpha)\, d\alpha$$

of a linear system in terms of the autocorrelation $R_{xx}(\tau)$ and power spectrum $S_{xx}(\omega)$ of the input $\mathbf{x}(t)$;

It can be shown (without proof) that

$$R_{yy}(\tau) = R_{xx}(\tau) * h(\tau) * h(-\tau)$$

$$S_{yy}(\omega) = S_{xx}(\omega)\, H(\omega)\, H^{*}(\omega) = S_{xx}(\omega)\, |\, H(\omega)\, |^{2}$$

<u>Example:</u>

$\mathbf{x}(t)$ is a wide sense stationary (wss) white noise process with autocorrelation function

$$R_{xx}(\tau) = q\, \delta(\tau)$$

where q is the average power. Find the spectral density of $\mathbf{x}(t)$.

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) \exp(-j\,\omega\,\tau)\, d\tau = \int_{-\infty}^{\infty} q\, \delta(\tau) \exp(-j\,\omega\,\tau)\, d\tau = q$$

$\mathbf{x}(t)$ passes through a linear time invariant (LTI) circuit with

$$H(\omega) = \begin{cases} 1 & \text{for} \quad -1 \leq \omega \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

and the response (output) of the circuit is $\mathbf{y}(t)$. Find the spectral density of the response.

$$S_{yy}(\omega) = S_{xx}(\omega)\, |\, H(\omega)\, |^{2}$$

$$S_{yy}(\omega) = \begin{cases} q & \text{for} \quad -1 \leq \omega \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the average power of $\mathbf{y}(t)$

Average power of $\boldsymbol{y}(t) = E\{\,|\,\boldsymbol{y}(t)\,|^2\,\} = R_{yy}(0)$

$$R_{yy}(\tau) = [1/(2\pi)] \int_{-\infty}^{\infty} S_{yy}(\omega) \exp(j\,\omega\,\tau)\, d\omega$$

$$R_{yy}(0) = [1/(2\pi)] \int_{-1}^{1} q \exp(j\,\omega\,0)\, d\omega = q/\pi$$