PROBABILITY THEORY

By
Prof. S. J. Soni
Assistant Professor
Computer Engg. Department
SPCE, Visnagar

Introduction

- Signals whose values at any instant t are determined by their analytical or graphical description are called **deterministic signals**, implying complete certainty about their values at any moment t. Such signals, which can be specified with certainty, cannot convey information.
- Unpredictable message signals and noise waveforms are examples of **random processes** that play key roles in communication systems and their analysis, because the higher the uncertainty about a signal to be received, the higher its information content.

Concept of Probability

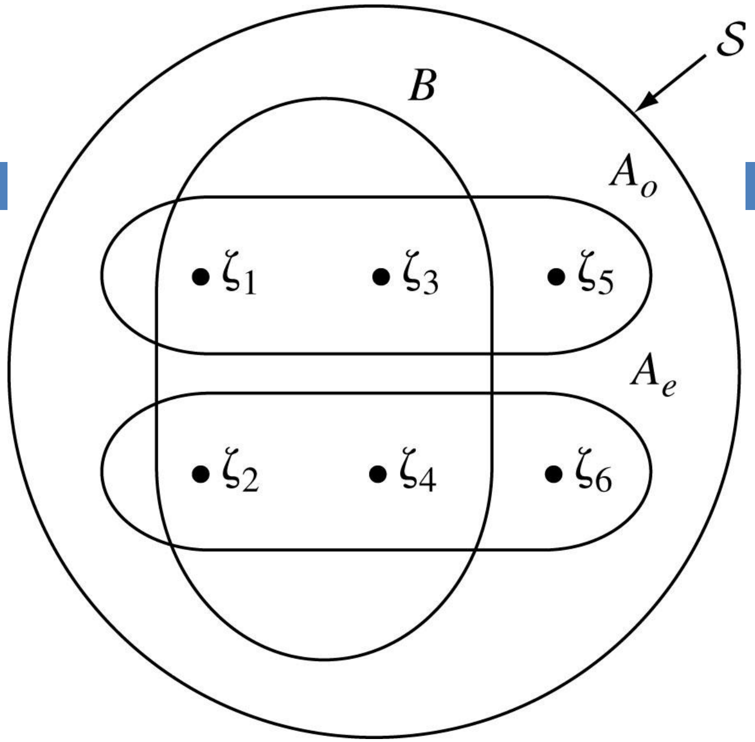
- The term <u>**experiment**</u> is used in probability theory to describe a process whose outcome cannot be fully predicted because the conditions under which it is performed cannot be predetermined with sufficient accuracy and completeness.
- Tossing a coin, rolling a die, and drawing a card from a desk are some examples of such experiment.
- An experiment may have several separately identifiable <u>**outcomes**</u>. For example, rolling a die has six possible identifiable outcomes (1,2,3,4,5, and 6).

Concept of Probability

- **<u>An event</u>** is a subset of outcomes that share some common characteristics.
- In the experiment of rolling a die, for example, the event "odd number on a throw" can result from any one of three outcomes (e.g. 1,3, or 5).
- Thus, events are grouping of outcomes into classes among which we choose to distinguish.

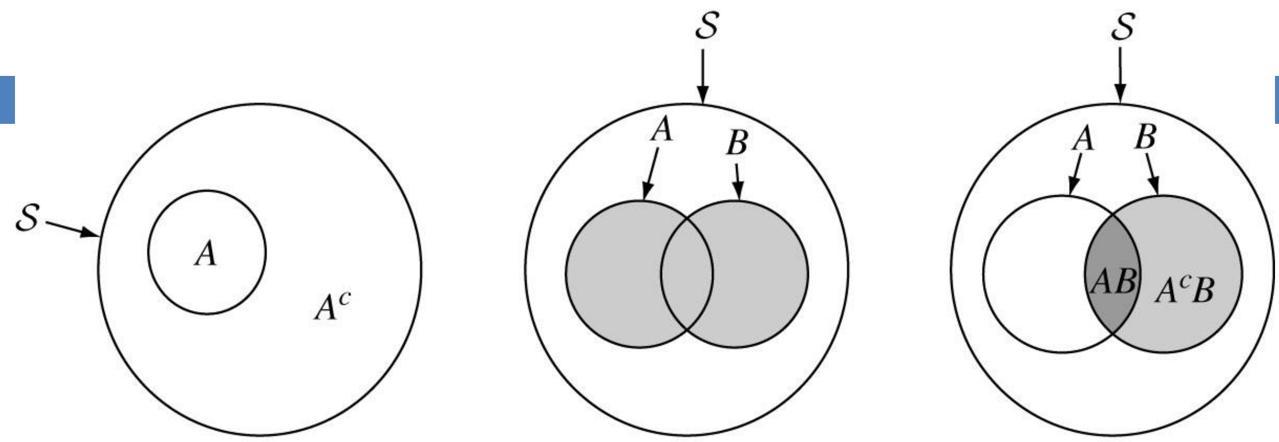
Sample Space

- The **sample space** S is a collection of all possible and separately identifiable outcomes of an experiment.



- Each outcome is an **element** or **sample point.**
- In case of rolling a die, the sample space consists of six samples points as shown in fig.
- The event "an odd number is thrown" denoted by A_0.
- The event "an even number is thrown" denoted by A_e.
- The event "a number equal to or less than 4 is thrown" as B.

Complement, Union and Intersection of Events



A^c (a) $A \cup B = B \cup A$ (b) AB and A^cB (c)

- The **complement** of any event A, denoted by A^c, is the event containing all points not in A.
- The **union** of events A & B, denoted by A U B, is the event that contains all points in A and B.
- The **intersection** of events A & B, denoted by A∩B, is the event that contains points common to A and B.

Examples

- ***Two dices are thrown. Determine the probability that the sum on the dice is seven.***

  For this experiment, the sample space contains 36 sample points because 36 possible outcomes exists. All outcomes are equally likely. Hence the probability of each outcome is 1/36.

  A sum of seven can be obtained by the six combinations: (1,6), (2,5), (3,4), (4,3), (5,2), & (6,1)

  $$P(\text{“a seven is thrown”}) = 1/36 + 1/36 + 1/36 + 1/36 + 1/36 + 1/36$$

  $$= 6/36 = 1/6$$

Examples

- ***A coin is tossed four times in succession. Determine the probability of obtaining exactly two heads.***

A total of $2^4=16$ distinct outcomes are possible. Hence the sample space consists of 16 points, each with probability 1/16. The 16 outcomes are as follows.

HHHH, HHHT, HHTH, **HHTT**, HTHH, **HTHT**, **HTTH**, HTTT

TTTT, TTTH, TTHT, **TTHH**, THTT, **THTH**, **THHT**, THHH

so, P("obtaining exactly two heads") = 6/16 = 3/8

Conditional Probability

- The **<u>conditional probability P(B|A)</u>** to denote the probability of event B when it is known that event A has occurred.
- P(B|A) is read as "probability of B given A".
- $P(A \cap B) = P(A)\,P(B|A)$ and $P(B|A) = P(A \cap B)/P(A)$
- $P(A|B) = P(A \cap B)/P(B)$

Example

- *An experiment consists of drawing two cards from a desk in succession (without replacing the first card drawn). Assign a value to the probability of obtaining two red aces in two cards.*
- Let A and B be the events “red ace in first draw” and “red ace in second draw” respectively.
- We wish to determine $P(A \cap B)$,

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

Example

- The relative frequency of A is 2/52 = 1/26.
- Hence, P(A) = 1/26
- Also for P(B|A) = 1/51

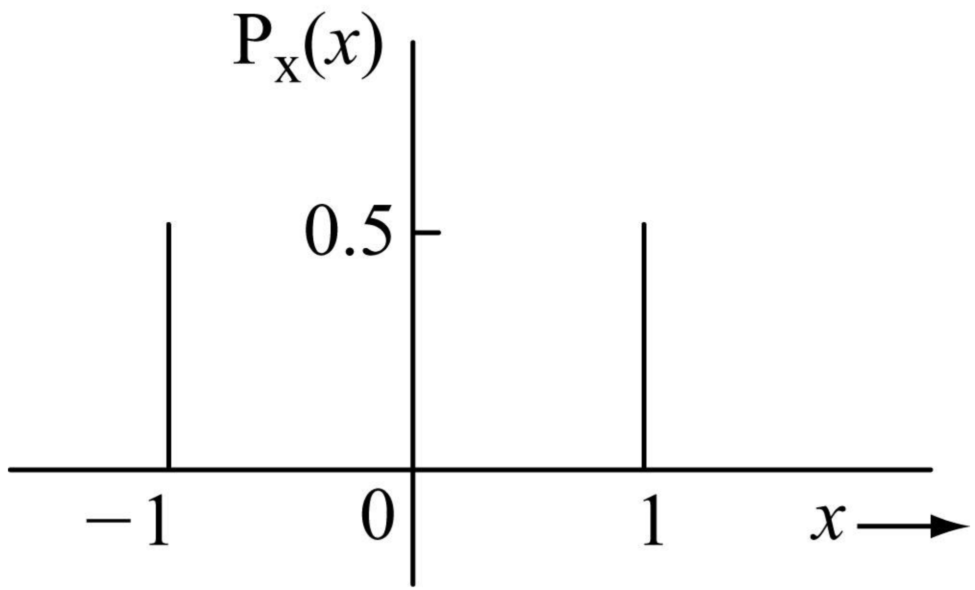
- Hence,

$P(A \cap B) = (1/26)\,(1/51) = 1/1326$

Discrete Random Variable

- The outcome of an experiment may be a real number (as in the case of rolling a die), or it may be nonnumerical and describable by a phrase (such as “heads” or “tail” in tossing a coin).
- From a mathematical point of view, it is simpler to have numerical values for all outcomes.
- For this reason, we assign a real number to each sample point according to some rule.

Probabilities in a coin-tossing experiment



- Here, we may assign the number 1 for the outcome heads and the number -1 for the outcome tails.

Random Variable

- We have a random variable x that takes on values x_1, x_2, …, x_n. We shall use roman type (x) to denote a random variable (RV) and italic type (e.g. x_1, x_2, …, x_n) to denote the value it takes.
- The probability of an RV x taking a value x_i is

$$\mathbf{P_X(x_i) = Probability\ of\ "X=x_i"}$$

- **Random Variable (RV)**: A finite single valued function that maps the set of all experimental outcomes into the set of real numbers *R* is said to be a RV, if the set is an event for every *x* in *R*.

Example

- ***Two dices are thrown. The sum of the points appearing on the two dices is an RV x. Find the values taken by x, and the corresponding probabilities.***

- Here, x can take on all integral values from 2 through 12.
- There are 36 sample points in all, each with probability 1/36.
- Note in the table that although there are 36 sample points, they all map into 11 values of x.

Example

Values of x_i	Dice Outcomes	$P_X(x_i)$
2	(1,1)	1/36
3	(1,2), (2,1)	2/36 = 1/18
4	(1,3), (2,2), (3,1)	3/36 = 1/12
5	(1,4), (2,3), (3,2), (4,1)	4/36 = 1/9
6	(1,5), (2,4), (3,3), (4,2), (5,1)	5/36
7	(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)	6/36 = 1/6
8	(2,6), (3,5), (4,4), (5,3), (6,2)	5/36
9	(3,6), (4,5), (5,4), (6,3)	4/36 = 1/9
10	(4,6), (5,5), (6,4)	3/36 = 1/12
11	(5,6), (6,5)	2/36 = 1/18
12	(6,6)	1/36

Cumulative Distribution Function (CDF)

- The cumulative distribution function (CDF) $\mathbf{F_x(x)}$ of an RV x is the probability that x takes a value less than or equal to x; that is,

$$\mathbf{F_x(x) = P\ (x <= x)}$$

- **CDF $\mathbf{F_x(x)}$** has the following four properties:
  1. $F_x(x) >= 0$
  2. $F_x(\infty) = 1$
  3. $F_x(-\infty) = 0$
  4. $F_x(x)$ is a nondecreasing function, that is,

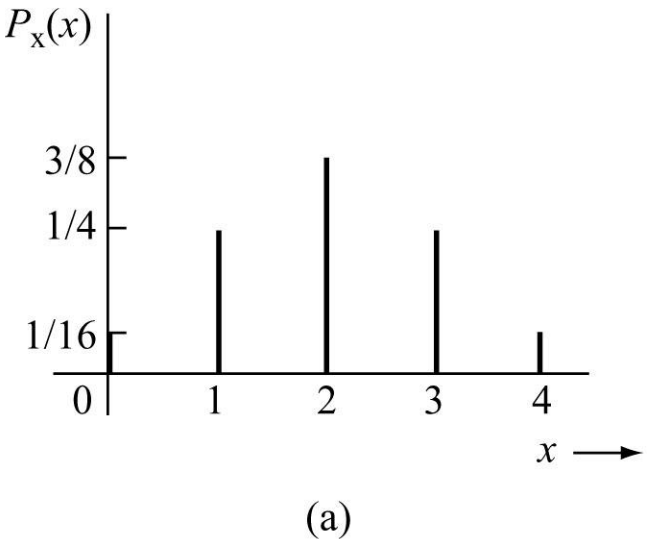
     $F_x(x_1) <= F_x(x_2)$ for $x_1 <= x_2$

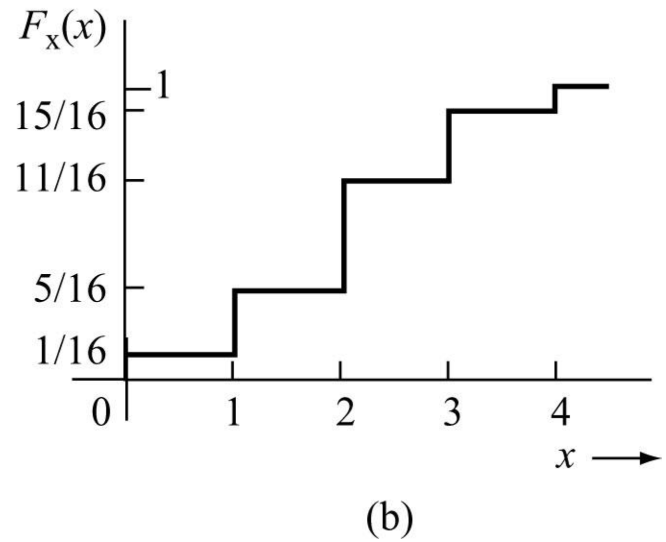
CDF Example

- In an experiment, a trial consists of four successive tosses of a coin. If we define an RV x as the number of heads appearing in a trial, determine **$P_x(x)$ and $F_x(x)$.**
- A total of 16 distinct equiprobable outcomes are listed in earlier example. (slide no. 8)
- A table can be formulated to find **$P_x(x)$.**

CDF Example

Values of x_i	Dice Outcomes	$P_X(x_i)$	$F_X(x_i)$
0	TTTT	1/16	1/16 + 0 = 1/16
1	HTTT, TTTH, TTHT, THTT	4/16 = 1/4	1/16 + ¼ = 5/16
2	HHTT, HTHT, HTTH, TTHH, THTH, THHT	6/16 = 3/8	5/16+3/8=11/16
3	HHHT, HHTH, HTHH, THHH	4/16 = 1/4	11/16+1/4=15/16
4	HHHH	1/16	15/16+1/16 =1





(a) Probabilities $P_x(x_i)$ and (b) the cumulative distribution function (CDF).

Continuous Random Variables

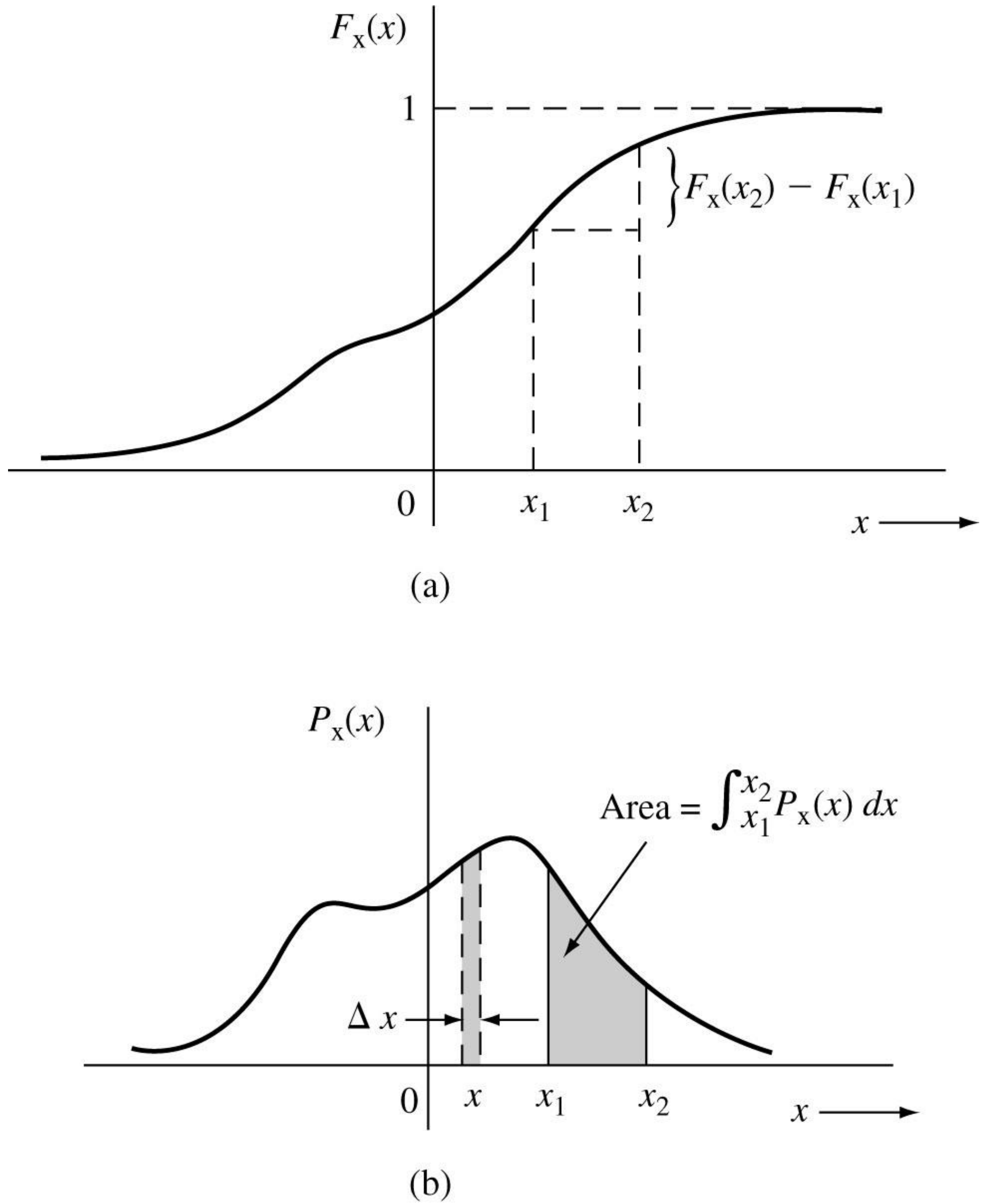
- A continuous RV x can assume any value in a certain interval.
- In a continuum of any range, an uncountably infinite number of possible values exist, and $P_x(x_i)$, the probabilities that $x = x_i$, as one of the uncountably infinite values, is generally zero.
- Properties of the CDF derived earlier are general and are valid for continuous as well as discrete RVs.

Probability Density Function (PDF)

- Probability density function can be describe as follows:

$$p_X(x) = \frac{dF_X(x)}{dx}.$$

- The function $p_x(x)$ is called the probability density function (PDF) of the RV x.



(a) Cumulative distribution function (CDF). (b) Probability density function (PDF).

Gaussian distribution for continuous random variables

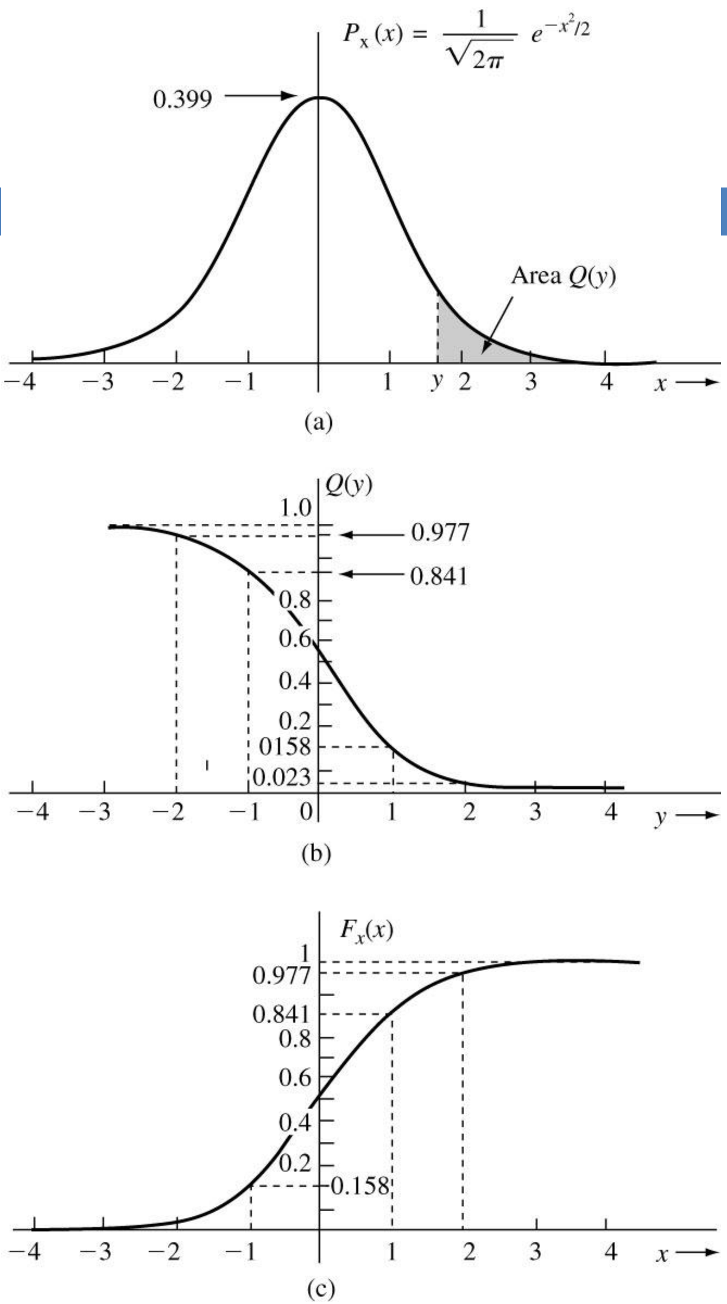
- A r.v. X is called a *normal (or gaussian) r.v. if its pdf is given by*

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

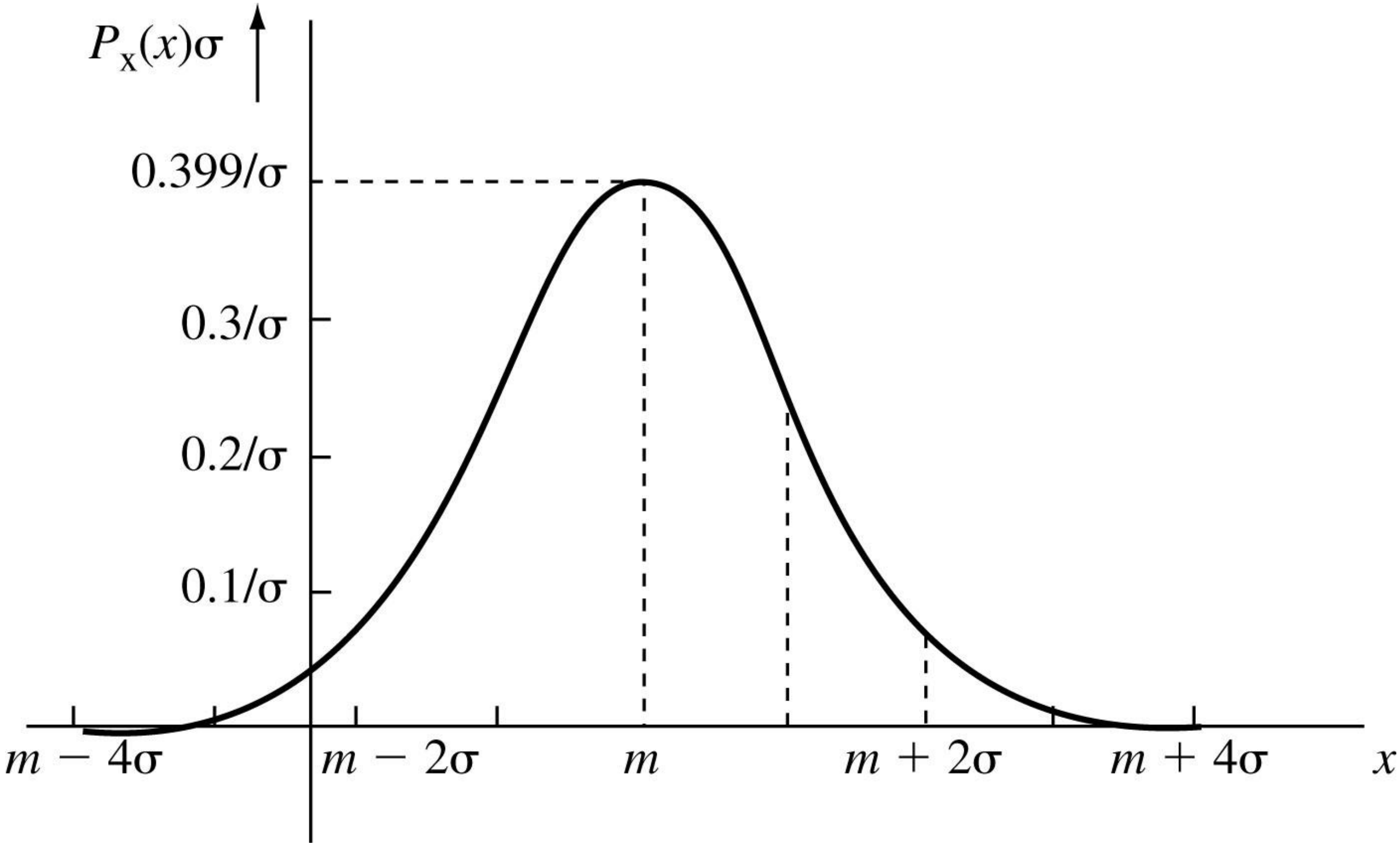
$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\xi^2/2}\, d\xi$$

Gaussian distribution

Figure (a) Gaussian PDF.
(b) Function Q(*y*).
(c) CDF of the Gaussian PDF.



Gaussian Density Function with two parameters



Gaussian PDF with mean *m* and variance σ^2.

Poisson Distribution

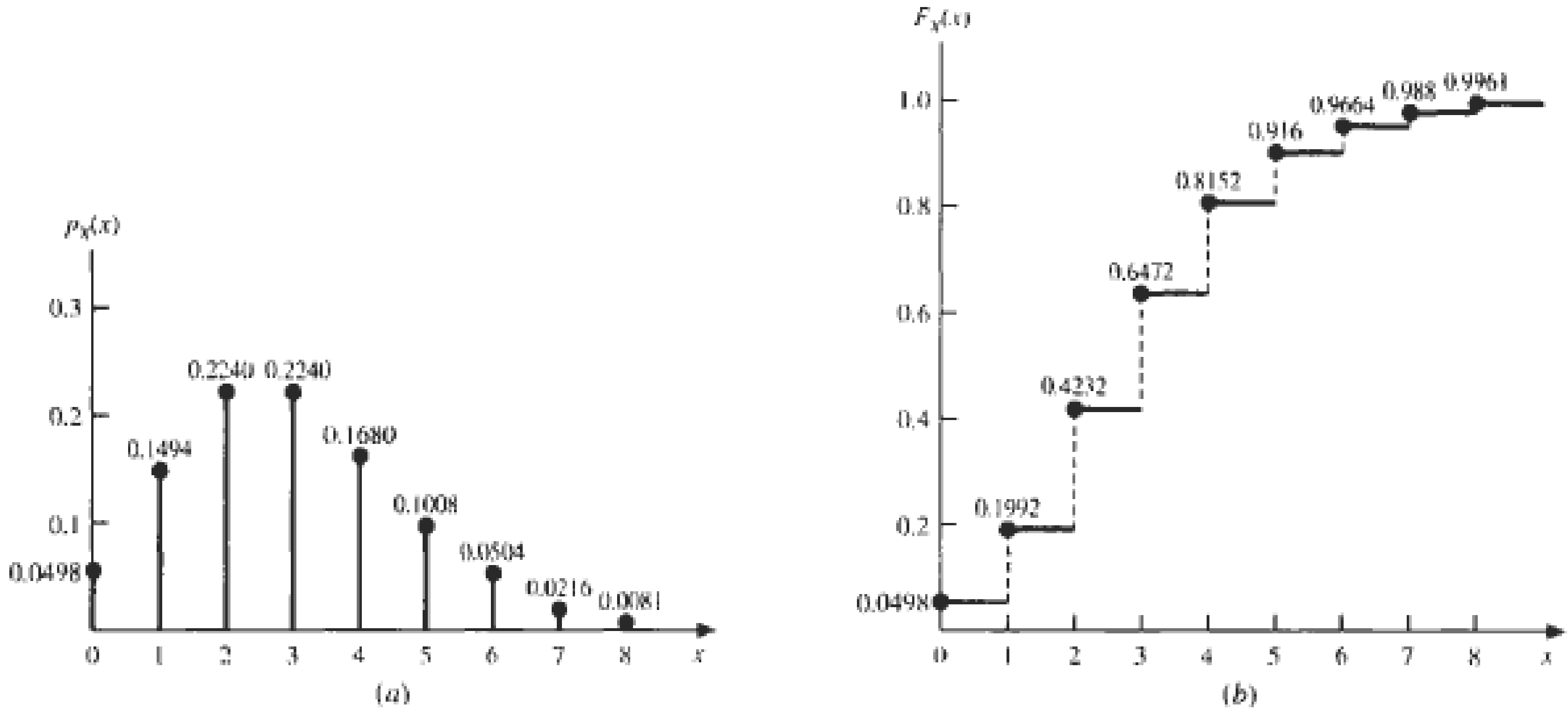
- A r.v. X is called a *Poisson r.v. with parameter A (> 0) if its pdf is given by*

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

- The corresponding cdf of X

$$F_X(x) = e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!} \qquad n \le x < n + 1$$

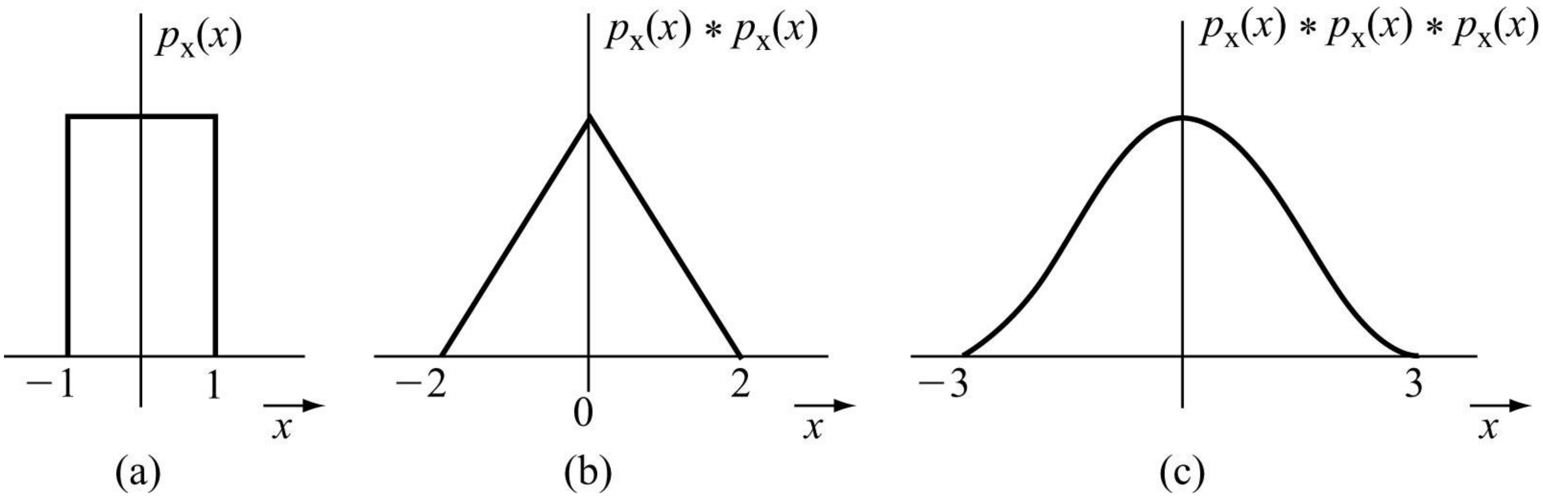
Poisson Distribution



Poisson distribution

Central Limit Theorem

- Under certain conditions, the sum of the large number of independent RVs tends to be a Gaussian random variable, independent of the probability densities of a variable added. The rigorous statement of this tendency is what is known as the **central limit theorem.**



Random/Stochastic Processes

- Here we introduce the concept of a random (or stochastic) process. The theory of random processes was first developed in connection with the study of fluctuations and noise in physical systems.
- **A random process is the mathematical model of an empirical process whose development is governed by probability laws.**
- Random processes provides useful models for the studies of such diverse fields as statistical physics, communication and control, time series analysis, population growth, and management sciences.

Definition

- A random process is a family of r.v.'s $(X(t), t \in T)$ defined on a given probability space, indexed by the parameter t, where t varies over an index set T.
- Recall that a random variable is a function defined on the sample space S. Thus, a random process $(X(t), t \in T)$ is really a function of two arguments $\{X(t, c), t \in T, c \in S\}$. *For a fixed* $t(=t_k)$, $X(t_k, c) = X_k(c)$ is a r.v. denoted by $X(t_k)$, as c varies over the sample space S. On the other hand, for a fixed sample point $c_i \in S$, $X(t, c_i) = X_i(t)$ is a single function of time t, called a sample function or a realization of the process. The totality of all sample functions is called an ensemble.
- Of course if both c and t are fixed, $X(t_k, c_i)$ is simply a real number.

Classification of Random Processes

Stationary Processes:

A random process $\{X(t), t \in T\}$ is said to be *stationary* or *strict-sense stationary* if, for all n and for every set of time instants $(t_i \in T, i = 1, 2, \ldots, n)$,

$$F_X(x_1, \ldots, x_n; t_1, \ldots, t_n) = F_X(x_1, \ldots, x_n; t_1 + \tau, \ldots, t_n + \tau)$$

for any τ. Hence, the distribution of a stationary process will be unaffected by a shift in the time origin, and $X(t)$ and $X(t + \tau)$ will have the same distributions for any τ. Thus, for the first-order distribution,

$$F_X(x; t) = F_X(x; t + \tau) = F_X(x)$$

and

$$f_X(x; t) = f_X(x)$$

Then

$$\mu_X(t) = E[X(t)] = \mu$$

$$\text{Var}[X(t)] = \sigma^2$$

where μ and σ^2 are contants. Similarly, for the second-order distribution,

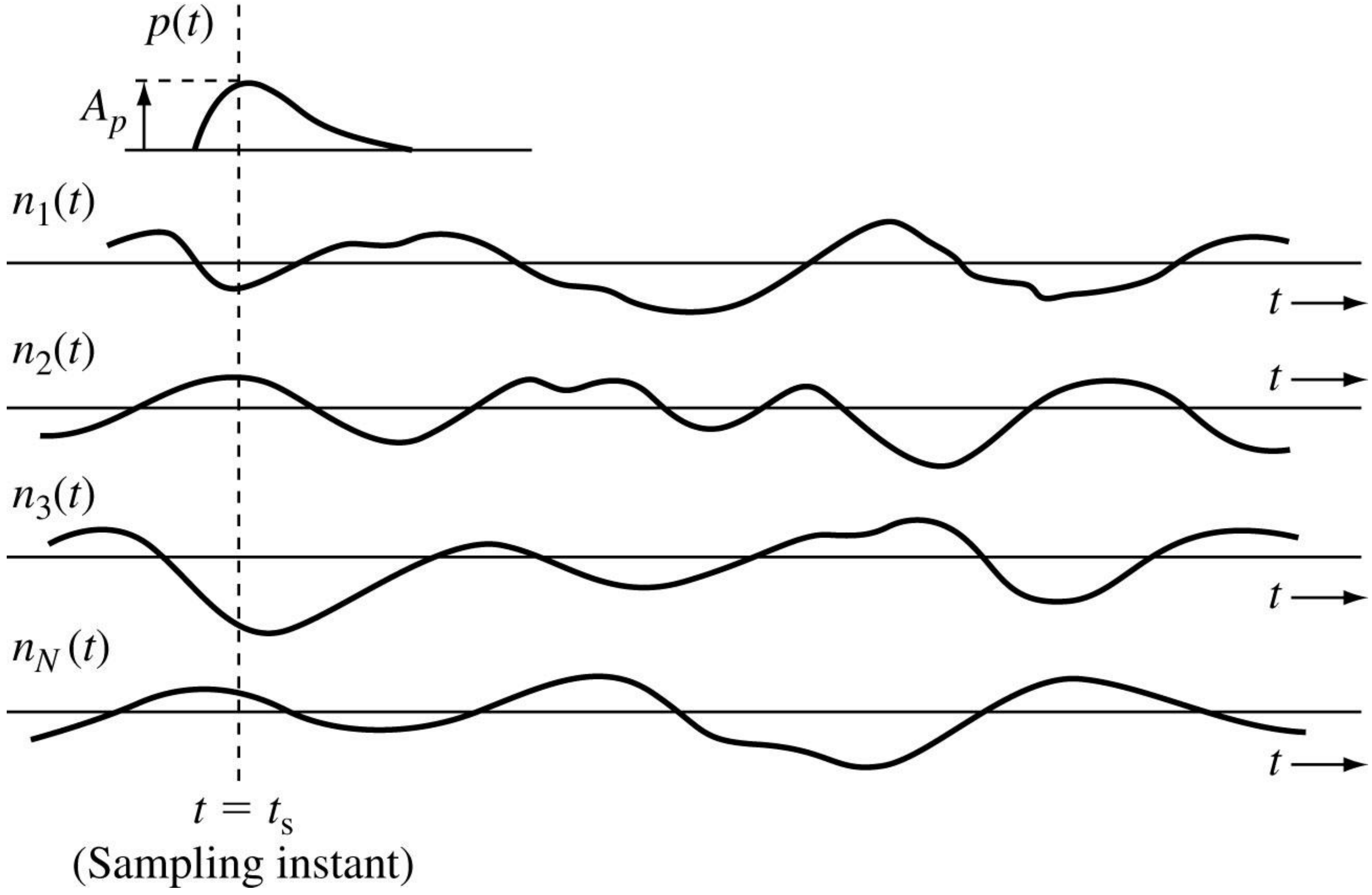
$$F_X(x_1, x_2; t_1, t_2) = F_X(x_1, x_2; t_2 - t_1)$$

and

$$f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; t_2 - t_1)$$

Nonstationary processes are characterized by distributions depending on the points $t_1, t_2, \ldots, t_n$.

Stationary Random Process



Random process for representing a channel noise

Wide-sense Stationary Processes

Wide-Sense Stationary Processes:

If stationary condition (*5.14*) of a random process $X(t)$ does not hold for all n but holds for $n \leq k$, then we say that the process $X(t)$ is *stationary to order k*. If $X(t)$ is stationary to order 2, then $X(t)$ is said to be *wide-sense stationary* (WSS) or *weak stationary*. If $X(t)$ is a WSS random process, then we have

1. $E[X(t)] = \mu$ (constant)
2. $R_X(t, s) = E[X(t)X(s)] = R_X(|s - t|)$

Note that a strict-sense stationary process is also a WSS process, but, in general, the converse is not true.

Ergodic Processes

Ergodic Processes:

Consider a random process $\{X(t), -\infty < t < \infty\}$ with a typical sample function $x(t)$. The time average of $x(t)$ is defined as

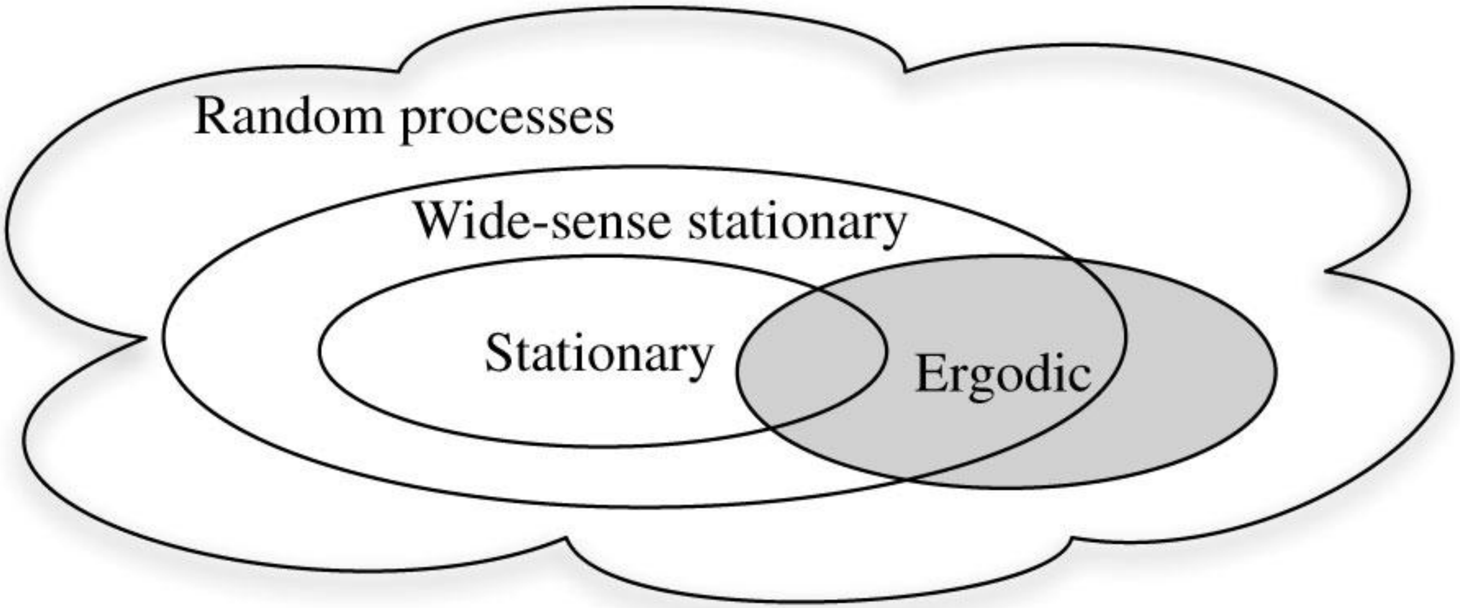
$$\langle x(t) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, dt$$

Similarly, the time autocorrelation function $\bar{R}_X(\tau)$ of $x(t)$ is defined as

$$\bar{R}_X(\tau) = \langle x(t)x(t+\tau) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)\, dt$$

A random process is said to be *ergodic* if it has the property that the time averages of sample functions of the process are equal to the corresponding statistical or ensemble averages. The subject of *ergodicity* is extremely complicated. However, in most physical applications, it is assumed that stationary processes are ergodic.

General Classification



Classification of random processes

Auto-Correlation Function

Autocorrelation Functions:

The autocorrelation function of a continuous-time random process $X(t)$ is defined as

$$R_X(\tau) = E[X(t)X(t+\tau)]$$

Properties of $R_X(\tau)$:

1. $R_X(-\tau) = R_X(\tau)$
2. $|R_X(\tau)| \leq R_X(0)$
3. $R_X(0) = E[X^2(t)] \geq 0$

Cross-Correlation Functions

Cross-Correlation Functions

The cross-correlation function of two continuous-time jointly WSS random processes $X(t)$ and $Y(t)$ is defined by

$$R_{XY}(\tau) = E[X(t)Y(t+\tau)]$$

Properties of $R_{XY}(\tau)$:

1. $R_{XY}(-\tau) = R_{YX}(\tau)$
2. $|R_{XY}(\tau)| \leq \sqrt{R_X(0)R_Y(0)}$
3. $|R_{XY}(\tau)| \leq \frac{1}{2}[R_X(0) + R_Y(0)]$

These properties are verified Two processes $X(t)$ and $Y(t)$ are called (*mutually*) *orthogonal* if

$$R_{XY}(\tau) = 0 \qquad \text{for all } \tau$$

Similarly, the cross-correlation function of two discrete-time jointly WSS random processes $X(n)$ and $Y(n)$ is defined by

$$R_{XY}(k) = E[X(n)Y(n+k)]$$

and various properties of $R_{XY}(k)$ similar to those of $R_{XY}(\tau)$ can be obtained by replacing τ by k