Probability and Stochastic Processes

Wireless Information Transmission System Lab.
Institute of Communications Engineering
National Sun Yat-sen University

Table of Contents

- Probability
  - Random Variables, Probability Distributions, and Probability Densities
  - Statistical Averages of Random Variables
  - Some Useful Probability Distributions
  - Upper Bounds on the Tail Probability
  - Sums of Random Variables and the Central Limit Theorem
- Stochastic Processes
  - Statistical Averages
  - Power Density Spectrum
  - Response of a Linear Time-Invariant System to a Random Input Signal
  - Discrete-Time Stochastic Signals and Systems
  - Cyclostationary Processes

Probability

- *Sample space* or *certain event* of a die experiment:

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

- The six outcomes are the *sample points* of the experiment.
- An *event* is a *subset* of *S*, and may consist of any number of sample points. For example:

$$A = \{2,4\}$$

- The *complement* of the event *A*, denoted by $\overline{A}$, consists of all the sample points in *S* that are not in *A*:

$$\overline{A} = \{1,3,5,6\}$$

Probability

- Two events are said to be *mutually exclusive* if they have no sample points in common – that is, if the occurrence of one event excludes the occurrence of the other. For example:

$$A = \{2,4\}; \qquad B = \{1,3,6\}$$

A and $\overline{A}$ are mutually exclusive events.

- The *union* (*sum*) of two events in an event that consists of all the sample points in the two events. For example:

$$C = \{1,2,3\}$$

$$D = B \cup C = \{1,2,3,6\}$$

$$A \cup \overline{A} = S$$

Probability

- The *intersection* of two events is an event that consists of the points that are common to the two events. For example:

$$E = B \cap C = \{1,3\}$$

- When the events are mutually exclusive, the intersection is the *null event*, denoted as φ. For example:

$$A \cap \overline{A} = \phi$$

Probability

- Associated with each event A contained in S is its probability $P(A)$.
- Three postulations:
  - $P(A) \geq 0$.
  - The probability of the sample space is $P(S)=1$.
  - Suppose that A_i , i =1, 2, …, are a (possibly infinite) number of events in the sample space S such that

    $$A_i \bigcap A_j = \phi; \qquad i \neq j = 1,2,\ldots$$

    Then the probability of the union of these mutually exclusive events satisfies the condition:

    $$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

Probability

- *Joint events* and *joint probabilities* (two experiments)
  - If one experiment has the possible outcomes A_i , i =1,2,…,n, and the second experiment has the possible outcomes B_j , j =1,2,…,m, then the combined experiment has the possible *joint outcomes* (A_i ,B_j), i =1,2,…,n, j =1,2,…,m.
  - Associated with each joint outcome (A_i ,B_j) is the *joint probability* $P(A_i ,B_j)$ which satisfies the condition:

$$0 \le P(A_i, B_j) \le 1$$

  - Assuming that the outcomes B_j , j =1,2,…,m, are mutually exclusive, it follows that:

$$\sum_{j=1}^{m} P(A_i, B_j) = P(A_i)$$

  - If all the outcomes of the two experiments are mutually exclusive, then:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} P\left(A_i, B_j\right) = \sum_{i=1}^{n} P\left(A_i\right) = 1$$

Probability

- *Conditional probabilities*
  - The conditional probability of the event A given the occurrence of the event B is defined as:

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

    provided $P(B)>0$.
  - $P(A,B)=P(A|B)P(B)=P(B|A)P(A)$
  - $P(A,B)$ is interpreted as the probability of $A\cap B$. That is, $P(A,B)$ denotes the simultaneous occurrence of A and B.
  - If two events A and B are mutually exclusive, $A\cap B=\phi$, then $P(A|B)=0$.
  - If B is a subset of A, we have $A\cap B=B$ and $P(A|B)=1$.

Probability

- *Bayes' theorem*:
  - If $A_i,\ i=1,2,\ldots,n,$ are mutually exclusive events such that

$$\bigcup_{\mathrm{i}=1}^{n} A_i = S$$

and B is an arbitrary event with nonzero probability, then

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

$$P(B) = \sum_{j=1}^{n} P\left(B, A_j\right) = \sum_{j=1}^{n} P\left(B \mid A_j\right) P\left(A_j\right)$$

  - $P(A_i)$ represents their *a priori probabilities* and $P(A_i|B)$ is the *a posteriori probability* of A_i conditioned on having observed the received signal B.

Probability

- *Statistical independence*
  - If the occurrence of A does not depend on the occurrence of B, then $P(A|B)=P(A)$.
  - $P(A,B)=P(A|B)P(B)=P(A)P(B)$
  - When the events A and B satisfy the relation $P(A,B)=P(A)P(B)$, they are said to be *statistically independent*.
  - Three statistically independent events A_1, A_2, and A_3 must satisfy the following conditions:

$$P(A_1,A_2)=P(A_1)P(A_2)$$

$$P(A_1,A_3)=P(A_1)P(A_3)$$

$$P(A_2,A_3)=P(A_2)P(A_3)$$

$$P(A_1,A_2,A_3)=P(A_1)P(A_2)P(A_3)$$

Random Variables, Probability Distributions, and Probability Densities

- Given an experiment having a sample space S and elements $s \in S$, we define a funciton $X(s)$ whose domain is S and whose range is a set of numbers on the real line.
- The function $X(s)$ is called a *random variable*.
  - Example 1: If we flip a coin, the possible outcomes are head (H) and tail (T), so S contains two points labeled H and T. Suppose we define a function $X(s)$ such that:
    $$X(s) = \begin{cases} +1 & (\mathrm{s} = \mathrm{H}) \\ -1 & (\mathrm{s} = \mathrm{T}) \end{cases}$$
    Thus we have mapped the two possible outcomes of the coin-flipping experiment into the two points (+1,-1) on the real line.
  - Example 2: Tossing a die with possible outcomes S={1,2,3,4,5,6}. A random variable defined on this sample space may be $X(s)=s$, in which case the outcomes of the experiment are mapped into the integers 1,…,6, or, perhaps, $X(s)=s^2$, in which case the possible outcomes are mapped into the integers {1,4,9,16,25,36}.

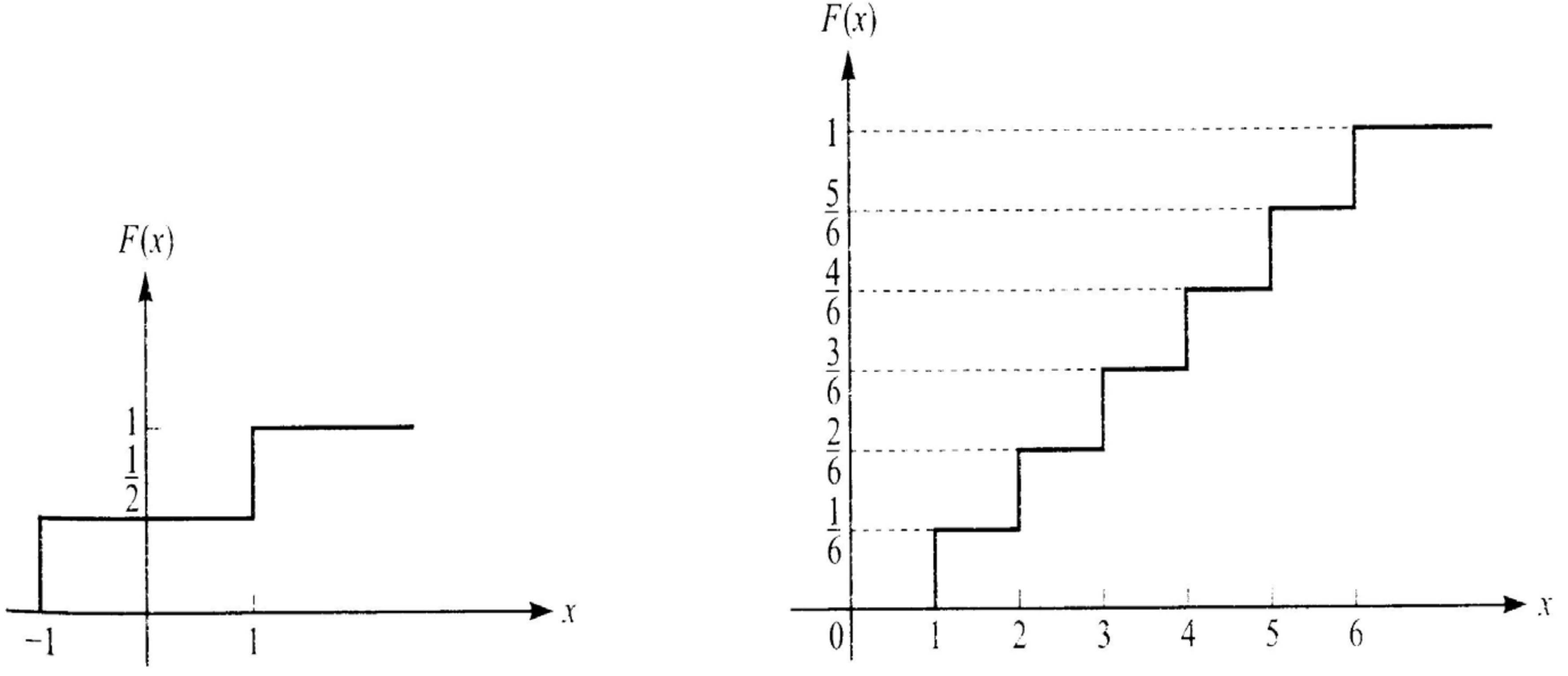
Random Variables, Probability Distributions, and Probability Densities

- Give a random variable X, let us consider the event $\{X \leq x\}$ where x is any real number in the interval $(-\infty, \infty)$. We write the probability of this event as $P(X \leq x)$ and denote it simply by $F(x)$, i.e.,

$$F(x) = P(X \leq x), \qquad -\infty < x < \infty$$

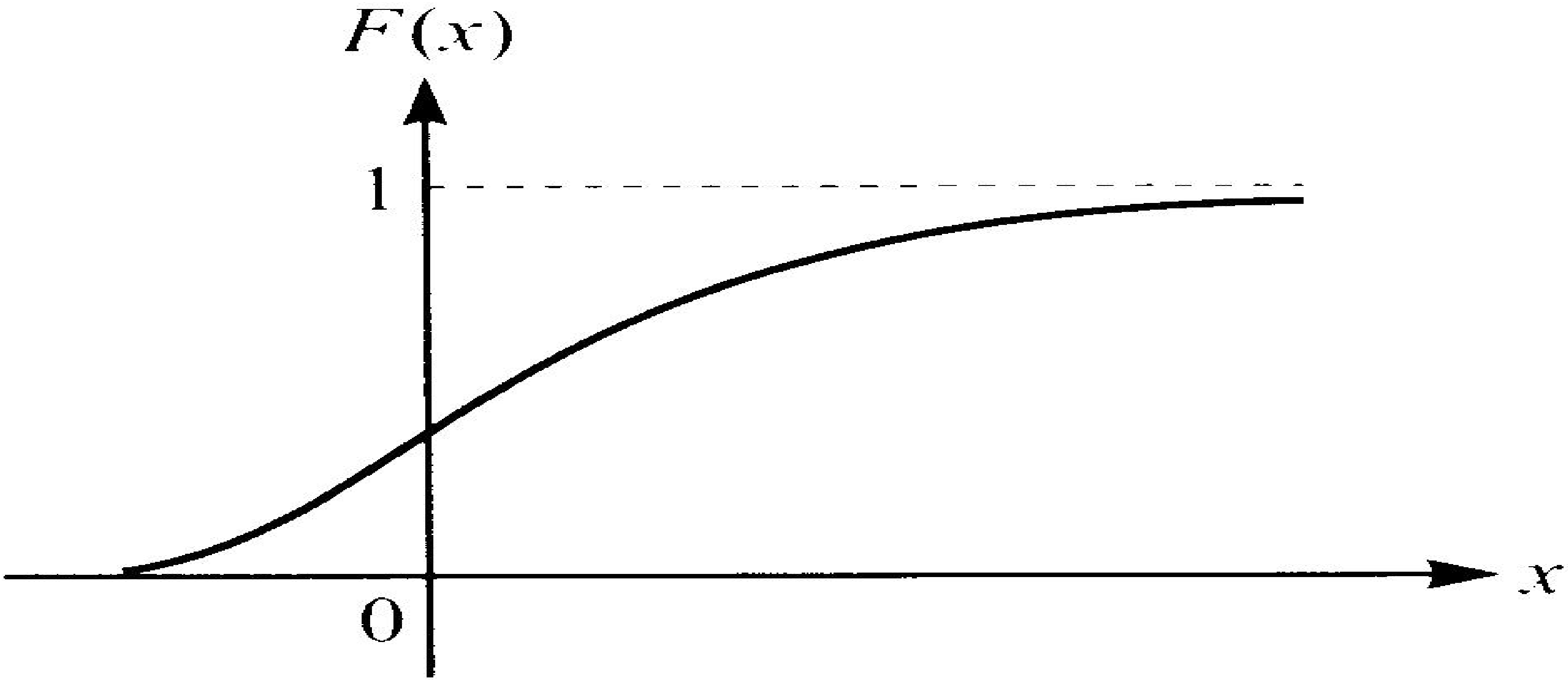
- The function $F(x)$ is called the *probability distribution function* of the random variable X.
- It is also called the *cumulative distribution function (CDF)*.
- $0 \leq F(x) \leq 1$
- $F(-\infty) = 0$ and $F(\infty) = 1$.

- Examples of the cumulative distribution functions of two discrete random variables.



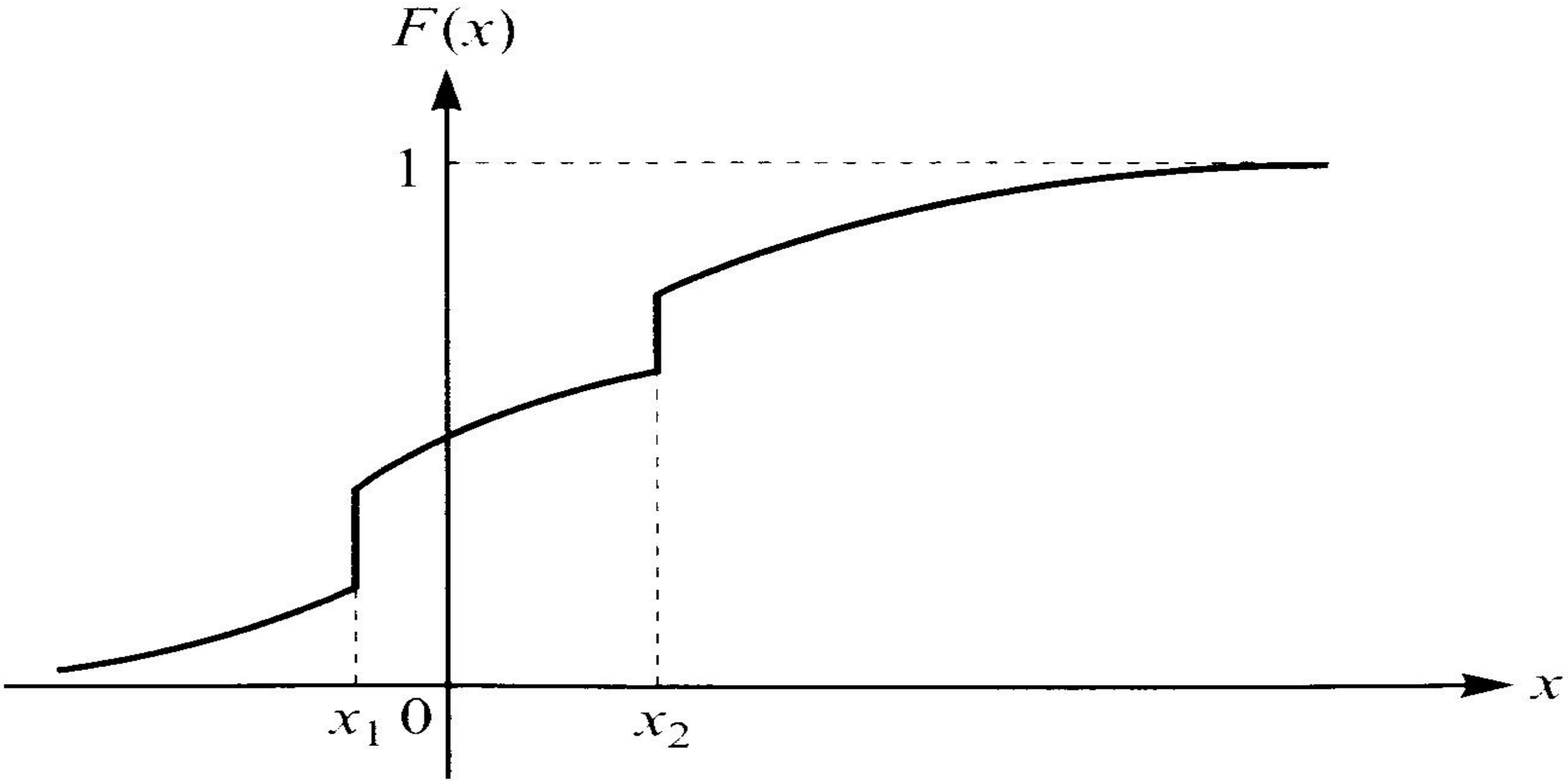
(a) (b)

- An example of the cumulative distribution function of a continuous random variable.



Random Variables, Probability Distributions, and Probability Densities

- An example of the cumulative distribution function of a random variable of a mixed type.



- The derivative of the CDF $F(x)$, denoted as $p(x)$, is called the *probability density function (PDF)* of the random variable X.

$$p(x) = \frac{dF(x)}{dx}, \qquad -\infty < x < \infty$$

$$F(x) = \int_{-\infty}^{x} p(u)du, \qquad -\infty < x < \infty$$

- When the random variable is discrete or of a mixed type, the PDF contains impulses at the points of discontinuity of $F(x)$:

$$p(x) = \sum_{i=1}^{n} P(X = x_i)\delta(x - x_i)$$

- Determining the probability that a random variable X falls in an interval (x_1, x_2) , where $x_2 > x_1$.

$$P(X \le x_2) = P(X \le x_1) + P(x_1 < X \le x_2)$$

$$F(x_2) = F(x_1) + P(x_1 < X \le x_2)$$

$$\Rightarrow \quad P(x_1 < X \le x_2) = F(x_2) - F(x_1)$$

$$= \int_{x_1}^{x_2} p(x)dx$$

The probability of the event $\{x_1 < X \le x_2\}$ is simply the area under the PDF in the range $x_1 < X \le x_2$.

- Multiple random variables, joint probability distributions, and joint probability densities: (two random variables)

Joint CDF: $F(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} p(u_1, u_2)du_1 du_2$

Joint PDF: $p(x_1, x_2) = \dfrac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2)$

$$\int_{-\infty}^{\infty} p(x_1, x_2)dx_1 = p(x_2) \qquad \int_{-\infty}^{\infty} p(x_1, x_2)dx_2 = p(x_1)$$

The PDFs $p(x_1)$ and $p(x_2)$ obtained from integrating over one of the variables are called *marginal* PDFs.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, x_2)dx_1 dx_2 = F(\infty, \infty) = 1$$

Note that: $F(-\infty, -\infty) = F(-\infty, x_2) = F(x_1, -\infty) = 0.$

- Multiple random variables, joint probability distributions, and joint probability densities: (multidimensional random variables)

Suppose that $X_i,\ i=1,2,\ldots,n,$ are random variables.

Joint CDF $F(x_1,x_2,\ldots,x_n)=P(X_1\le x_1, X_2\le x_2,\ldots,X_n\le x_n)$

$$=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n}p(u_1,u_2,\ldots,u_n)du_1du_2\ldots du_n$$

Joint PDF $p(x_1,x_2,\ldots,x_n)=\dfrac{\partial^n}{\partial x_1\partial x_2\ldots\partial x_n}F(x_1,x_2,\ldots,x_n)$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x_1,x_2,\ldots,x_n)dx_2dx_3=p(x_1,x_4,\ldots,x_n)$$

$$F\left(x_1,\infty,\infty,x_4,\ldots,x_n\right)=F(x_1,x_4,x_5,\ldots,x_n).$$

$$F\left(x_1,-\infty,-\infty,x_4,\ldots,x_n\right)=0.$$

Statistical Averages of Random Variables

- The *mean* or *expected value* of X, which characterized by its PDF $p(x)$, is defined as:

$$E(X) \equiv m_x = \int_{-\infty}^{\infty} x p(x) dx$$

This is the *first moment* of random variable X.

- The *n-th moment* is defined as:

$$E(X^n) = \int_{-\infty}^{\infty} x^n p(x) dx$$

- Define $Y=g(X)$, the expected value of Y is:

$$E(Y) = E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) p(x) dx$$

Statistical Averages of Random Variables

- The *n-th central moment* of the random variable X is:

$$E(Y) = E\left[(X - m_x)^n\right] = \int_{-\infty}^{\infty} (x - m_x)^n p(x)dx$$

- When n=2, the central moment is called the *variance* of the random variable and denoted as σ_x^2:

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 p(x)dx$$

$$\sigma_x^2 = E(X^2) - [E(X)]^2 = E(X^2) - m_x^2$$

- In the case of two random variables, X_1 and X_2, with joint PDF $p(x_1,x_2)$, we define the *joint moment* as:

$$E(X_1^k X_2^n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^k x_2^n p(x_1, x_2)dx_1dx_2$$

Statistical Averages of Random Variables

- The *joint central moment* is defined as:

$$E\left[(X_1 - m_1)^k (X_2 - m_2)^n\right]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1 - m_1)^k (x_2 - m_2)^n p(x_1, x_2) dx_1 dx_2$$

- If $k=n=1$, the *joint moment* and *joint central moment* are called the *correlation* and the *covariance* of the random variables X_1 and X_2, respectively.
- The *correlation* between X_i and X_j is given by the joint moment:

$$E(X_i X_j) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i x_j p(x_i, x_j) dx_i dx_j$$

Statistical Averages of Random Variables

- The *covariance* between X_i and X_j is given by the joint central moment:

$$\mu_{ij} \equiv E[(X_i - m_i)(X_j - m_j)]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i - m_i)(x_j - m_j)p(x_i, x_j)dx_i dx_j$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i x_j - x_j m_i - x_i m_j + m_i m_j)p(x_i, x_j)dx_i dx_j$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_i x_j p(x_i, x_j)dx_i dx_j - m_i m_j$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_i x_j p(x_i, x_j)dx_i dx_j - m_i m_j - m_i m_j + m_i m_j$$
$$= E(X_i X_j) - m_i m_j$$

- The $n \times n$ matrix with elements μ_{ij} is called the *covariance matrix* of the random variables, X_i, $i=1,2, \ldots, n$.

Statistical Averages of Random Variables

- Two random variables are said to be *uncorrelated* if $E(X_iX_j)=E(X_i)E(X_j)=m_im_j$.
- Uncorrelated → Covariance $\mu_{ij} = 0$.
- If X_i and X_j are statistically independent, they are uncorrelated.
- If X_i and X_j are uncorrelated, they are not necessary statistically independently.
- Two random variables are said to be *orthogonal* if $E(X_iX_j)=0$.
  - Two random variables are orthogonal if they are uncorrelated and either one or both of them have zero mean.

Statistical Averages of Random Variables

- Characteristic functions
  - The *characteristic function* of a random variable X is defined as the statistical average:

$$E(e^{jvX}) \equiv \psi(jv) = \int_{-\infty}^{\infty} e^{jvx} p(x)dx$$

  - $\Psi(jv)$ may be described as the *Fourier transform* of $p(x)$.
  - The inverse Fourier transform is:

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi(jv)e^{-jvx}dv$$

  - First derivative of the above equation with respect to v:

$$\frac{d\psi(jv)}{dv} = j\int_{-\infty}^{\infty} xe^{jvx} p(x)dx$$

Statistical Averages of Random Variables

- Characteristic functions (cont.)
  - First moment (mean) can be obtained by:

$$E(X) = m_x = -j \left. \frac{d\psi(jv)}{dv} \right|_{v=0}$$

  - Since the differentiation process can be repeated, *n*-th moment can be calculated by:

$$E(X^n) = (-j)^n \left. \frac{d^n\psi(jv)}{dv^n} \right|_{v=0}$$

Statistical Averages of Random Variables

- Characteristic functions (cont.)
  - Determining the PDF of a sum of statistically independent random variables:

$$Y=\sum_{i=1}^{n}X_i \quad \Rightarrow \quad \psi_Y(jv)=E(e^{jvY})=E\left[\exp\left(jv\sum_{i=1}^{n}X_i\right)\right]$$

$$=E\left[\prod_{i=1}^{n}\left(e^{jvX_i}\right)\right]=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\prod_{i=1}^{n}e^{jvx_i}\right)p(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n$$

Since the random variables are statistically independent,

$$p(x_1,x_2,\ldots,x_n)=p(x_1)p(x_2)\ldots p(x_n) \quad \Rightarrow \quad \psi_Y(jv)=\prod_{i=1}^{n}\psi_{X_i}(jv)$$

If X_i are iid (independent and identically distributed)

$$\Rightarrow \quad \psi_Y(jv)=\left[\psi_X(jv)\right]^n$$

Statistical Averages of Random Variables

- Characteristic functions (cont.)
  - The PDF of Y is determined from the inverse Fourier transform of $\Psi_Y(jv)$.
  - Since the characteristic function of the sum of n statistically independent random variables is equal to the product of the characteristic functions of the individual random variables, it follows that, in the transform domain, the PDF of Y is the n-fold convolution of the PDFs of the X_i.
  - Usually, the n-fold convolution is more difficult to perform than the characteristic function method in determining the PDF of Y.

Some Useful Probability Distributions

- Binomial distribution (discrete):
  - $P(X=0)=1-P(X=1)=p$
  - Let $Y=\sum_{i=1}^{n} X_i$ where the $X_i, i=1,2,\ldots,n$ are statistically iid, what is the probability distribution function of Y?

$$P(Y=k)=\binom{n}{k}p^k(1-p)^{n-k} \qquad \binom{n}{k}\equiv\frac{n!}{k!(n-k)!}$$

PDF of Y is: $p(y)=\sum_{k=0}^{n}P(Y=k)\delta(y-k)$

$$=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}\delta(y-k)$$

Some Useful Probability Distributions

- Binomial distribution:
  - The CDF of Y is:

$$F(y) = P(Y \le y) = \sum_{k=0}^{[y]} \binom{n}{k} p^k (1-p)^{n-k}$$

where $[y]$ denotes the largest integer m such that $m \le y$.

  - The first two moments of Y are:

$$E(Y) = np$$

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

$$\sigma^2 = np(1-p)$$

  - The characteristic function is:

$$\psi(jv) = (1 - p + pe^{jv})^n$$

Some Useful Probability Distributions

- Uniform Distribution
  - The first two moments of X are:

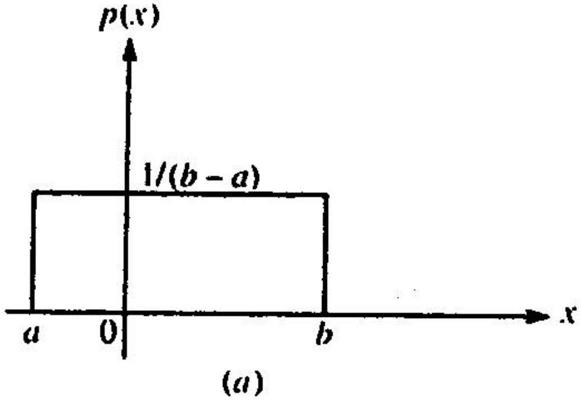
$$E(X) = \frac{1}{2}(a+b)$$

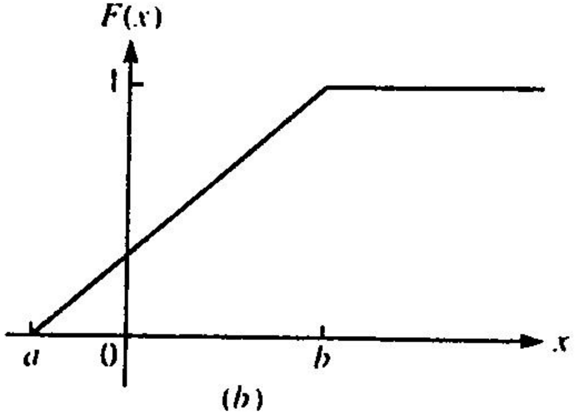
$$E(X^2) = \frac{1}{3}(a^2 + b^2 + ab)$$

$$\sigma^2 = \frac{1}{12}(a-b)^2$$

  - The characteristic function is:

$$\psi(jv) = \frac{e^{jvb} - e^{jva}}{jv(b-a)}$$





Some Useful Probability Distributions

- Gaussian (Normal) Distribution
  - The PDF of a Gaussian or normal distributed random variable is:

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

where m_x is the mean and σ^2 is the variance of the random variable.

$$t = \frac{(u-m_x)}{\sqrt{2}\sigma}, \quad dt = \frac{du}{\sqrt{2}\sigma}$$

  - The CDF is:

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-(u-m_x)^2/2\sigma^2} du = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{(x-m_x)/\sqrt{2}\sigma} e^{-t^2} dt$$

$$= \frac{1}{2} + \frac{1}{2} erf\left(\frac{x-m_x}{\sqrt{2}\sigma}\right) = 1 - \frac{1}{2} erfc\left(\frac{x-m_x}{\sqrt{2}\sigma}\right)$$

Some Useful Probability Distributions

- Gaussian (Normal) Distribution
  - *erf*( ) and *erfc*( ) denote the *error function* and *complementary error function*, respectively, and are defined as:

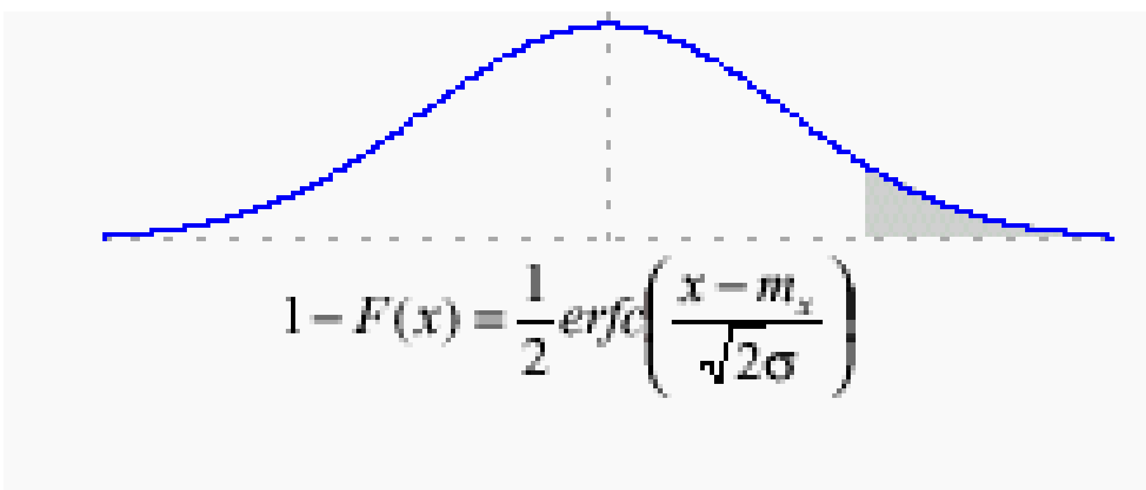
$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad \text{and} \quad erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt = 1 - erf(x)$$

  - $erf(-x) = -erf(x)$, $erfc(-x) = 2 - erfc(x)$, $erf(0) = erfc(\infty) = 0$, and $erf(\infty) = erfc(0) = 1$.
  - For $x > m_x$, the complementary error functions is proportional to the area under the tail of the Gaussian PDF.

Some Useful Probability Distributions

- Gaussian (Normal) Distribution
  - The function that is frequently used for the area under the tail of the Gaussian PDF is denoted by $Q(x)$ and is defined as:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt = \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right) \qquad x \geq 0$$



Some Useful Probability Distributions

- Gaussian (Normal) Distribution
  - The characteristic function of a Gaussian random variable with mean m_x and variance σ^2 is:

$$\psi(jv) = \int_{-\infty}^{\infty} e^{jvx}\left[\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m_x)^2/2\sigma^2}\right]dx = e^{jvm_x-(1/2)v^2\sigma^2}$$

  - The central moments of a Gaussian random variable are:

$$E\left[(X-m_x)^k\right] = \mu_k = \begin{cases} 1\cdot 3\cdots(k-1)\sigma^k & (\text{even } k) \\ 0 & (\text{odd } k) \end{cases}$$

  - The ordinary moments may be expressed in terms of the central moments as:

$$E\left[X^k\right] = \sum_{i=0}^{k} \binom{k}{i} m_x^i \mu_{k-i}$$

Some Useful Probability Distributions

- Gaussian (Normal) Distribution
  - The sum of n statistically independent Gaussian random variables is also a Gaussian random variable.

$$Y = \sum_{i=1}^{n} X_i$$

$$\psi_Y(jv) = \prod_{i=1}^{n} \psi_{X_i}(jv) = \prod_{i=1}^{n} e^{jvm_i - v^2\sigma_i^2/2} = e^{jvm_y - v^2\sigma_y^2/2}$$

where $m_y = \sum_{i=1}^{n} m_i$ and $\sigma_y^2 = \sum_{i=1}^{n} \sigma_i^2$

Therefore, Y is Gaussian-distributed with mean m_y and variance σ_y^2.

Some Useful Probability Distributions

- Chi-square distribution
  - If $Y=X^2$, where X is a Gaussian random variable, Y has a chi-square distribution. Y is a transformation of X.
  - There are two type of chi-square distribution:
    - *Central chi-square distribution*: X has zero mean.
    - *Non-central chi-square distribution*: X has non-zero mean.
  - Assuming X be Gaussian distributed with zero mean and variance σ^2, we can apply (2.1-47) to obtain the PDF of Y with a=1 and b=0;

$$p_Y(y) = \frac{p_X\left[x_1 = \sqrt{(y-b)/a}\right]}{\left|g'\left[x_1 = \sqrt{(y-b)/a}\right]\right|} + \frac{p_X\left[x_1 = -\sqrt{(y-b)/a}\right]}{\left|g'\left[x_1 = -\sqrt{(y-b)/a}\right]\right|}$$

- Central chi-square distribution
  - The PDF of Y is:

$$p_Y(y) = \frac{1}{\sqrt{2\pi y}\sigma} e^{-y/2\sigma^2}, \quad y \geq 0$$

  - The CDF of Y is:

$$F_Y(y) = \int_0^y p_Y(u)du = \frac{1}{\sqrt{2\pi}\sigma} \int_0^y \frac{1}{\sqrt{u}} e^{-u/2\sigma^2} du$$

  - The characteristic function of Y is:

$$\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{1/2}}$$

Some Useful Probability Distributions

- Chi-square (Gamma) distribution with n degrees of freedom.
  - $Y=\sum_{i=1}^{n} X_i^2,\quad X_i, i=1,2,\ldots,n,$ are statistically independent and identically distributed (iid) Gaussian random variables with zero mean and variance σ^2.
  - The characteristic function is:

$$\psi_Y(jv)=\frac{1}{(1-j2v\sigma^2)^{n/2}}$$

  - The inverse transform of this characteristic function yields the PDF:

$$p_Y(y)=\frac{1}{\sigma^n 2^{n/2}\Gamma\left(\frac{1}{2}n\right)} y^{n/2-1}e^{-y/2\sigma^2},\quad y\geq 0$$

Some Useful Probability Distributions

- Chi-square (Gamma) distribution with *n* degrees of freedom (cont.).
  - $\Gamma(p)$ is the gamma function, defined as:

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \quad p > 0$$

$$\Gamma(p) = (\mathrm{p}-1)! \qquad p \text{ an integer} > 0$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \qquad \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\sqrt{\pi}$$

  - When *n*=2, the distribution yields the exponential distribution.

Some Useful Probability Distributions

- Chi-square (Gamma) distribution with *n* degrees of freedom (cont.).
  - The PDF of a chi-square distributed random variable for several degrees of freedom.

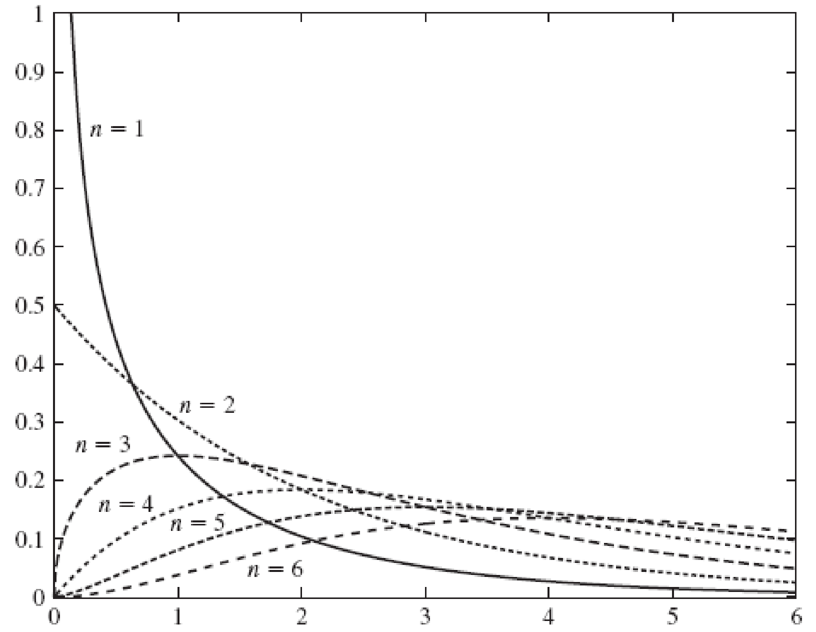


FIGURE 2.3–3
The PDF of the χ^2 random variable for different values of n. All plots are shown for $\sigma = 1$.

Some Useful Probability Distributions

- Chi-square (Gamma) distribution with *n* degrees of freedom (cont.).
  - The first two moments of *Y* are:

$$E(Y) = n\sigma^2$$

$$E(Y^2) = 2n\sigma^4 + n^2\sigma^4$$

$$\sigma_y^2 = 2n\sigma^4$$

  - The CDF of *Y* is:

$$F_Y(y) = \int_0^y \frac{1}{\sigma^n 2^{n/2}\Gamma\left(\frac{1}{2}n\right)} u^{n/2-1} e^{-u/2\sigma^2} du, \qquad y \geq 0$$

Some Useful Probability Distributions

- Chi-square (Gamma) distribution with *n* degrees of freedom (cont.).
  - The integral in CDF of *Y* can be easily manipulated into the form of the incomplete gamma function, which is tabulated by Pearson (1965).
  - When *n* is even, the integral can be expressed in closed form. Let $m=n/2$, where m is an integer, we can obtain:

$$F_Y(y) = 1 - e^{-y/2\sigma^2} \sum_{k=0}^{m-1} \frac{1}{k!} \left(\frac{y}{2\sigma^2}\right)^k, \qquad y \geq 0$$

Some Useful Probability Distributions

- Non-central chi-square distribution
  - If X is Gaussian with mean m_x and variance σ^2, the random variable $Y=X^2$ has the PDF:

$$p_Y(y) = \frac{1}{\sqrt{2\pi y}\sigma} e^{-(y+m_x^2)/2\sigma^2} \cosh\left(\frac{\sqrt{y}m_x}{\sigma^2}\right), \qquad y \geq 0$$

  - The characteristic function corresponding to this PDF is:

$$\psi_Y(jv) = \frac{1}{(1-j2v\sigma^2)^{1/2}} e^{jm_x^2 v/(1-j2v\sigma^2)}$$

Some Useful Probability Distributions

- Non-central chi-square distribution with *n* degrees of freedom
  - $Y = \sum_{i=1}^{n} X_i^2, \quad X_i, i = 1,2,\ldots,n,$ are statistically independent and identically distributed (iid) Gaussian random variables with mean $m_i, \; i = 1,2,\ldots,n,$ and identical variance equal to σ^2.
  - The characteristic function is:

$$\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{n/2}} \exp\left(\frac{jv\sum_{i=1}^{n} m_i^2}{1 - j2v\sigma^2} \right)$$

Some Useful Probability Distributions

- Non-central chi-square distribution with n degrees of freedom
  - The characteristic function can be inverse Fourier transformed to yield the PDF:

$$p_Y(y) = \frac{1}{2\sigma^2}\left(\frac{y}{s^2}\right)^{(n-2)/4} e^{-(s^2+y)/2\sigma^2} I_{n/2-1}\left(\sqrt{y}\frac{s}{\sigma^2}\right), \qquad y \geq 0$$

where, s^2 is called the non-centrality parameter:

$$s^2 = \sum_{i=1}^{n} m_i^2$$

and $I_\alpha(x)$ is the αth-order modified Bessel function of the first kind, which may be represented by the infinite series:

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{\alpha+2k}}{k!\Gamma(\alpha+k+1)}, \qquad x \geq 0$$

Some Useful Probability Distributions

- Non-central chi-square distribution with *n* degrees of freedom
  - The CDF is:

$$F_Y(y) = \int_0^y \frac{1}{2\sigma^2}\left(\frac{u}{s^2}\right)^{(n-2)/4} e^{-(s^2+u)/2\sigma^2} I_{n/2-1}\left(\sqrt{u}\frac{s}{\sigma^2}\right)du$$

  - The first two moments of a non-central chi-square-distributed random variable are:

$$E(Y) = n\sigma^2 + s^2$$

$$E(Y^2) = 2n\sigma^4 + 4\sigma^2 s^2 + (n\sigma^2 + s^2)^2$$

$$\sigma_y^2 = 2n\sigma^4 + 4\sigma^2 s^2$$

Some Useful Probability Distributions

- Non-central chi-square distribution with *n* degrees of freedom
  - When *m*=*n*/2 is an integer, the CDF can be expressed in terms of the generalized Marcum's *Q* function:

$$Q_m(a,b) = \int_b^\infty x\left(\frac{x}{a}\right)^{m-1} e^{-(x^2+a^2)/2} I_{m-1}(ax)dx$$

$$= Q_1(a,b) + e^{-\left(a^2+b^2\right)/2}\sum_{k=0}^{m-1}\left(\frac{b}{a}\right)^k I_k(ab)$$

where $Q_1(a,b) = e^{-\left(a^2+b^2\right)/2}\sum_{k=0}^{\infty}\left(\frac{b}{a}\right)^k I_k(ab), \qquad b > a > 0$

By using $x^2 = \frac{u}{\sigma^2}$ and let $a^2 = s^2 / \sigma^2$, it is easily shown:

$$F_Y(y) = 1 - Q_m\left(\frac{s}{\sigma}, \frac{\sqrt{y}}{\sigma}\right)$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Rayleigh distribution is frequently used to model the statistics of signals transmitted through radio channels such as cellular radio.
  - Consider a carrier signal *s* at a frequency ω_0 and with an amplitude *a*:

$$s = a \cdot \exp(j\omega_0 t)$$

  - The received signal s_r is the sum of *n* waves:

$$s_r = \sum_{i=1}^{n} a_i \exp\left[j(\omega_0 t + \theta_i)\right] \equiv r \exp\left[j(\omega_0 t + \theta)\right]$$

$$\text{where} \quad r \exp(j\theta) = \sum_{i=1}^{n} a_i \exp(j\theta_i)$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Define: $r\exp(j\theta) = \sum_{i=1}^{n} a_i \cos\theta_i + j\sum_{i=1}^{n} a_i \sin\theta_i \equiv x + jy$

    We have: $x \equiv \sum_{i=1}^{n} a_i \cos\theta_i$ and $y \equiv \sum_{i=1}^{n} a_i \sin\theta_i$

    where: $r^2 = x^2 + y^2 \quad x = r\cos\theta \quad y = r\sin\theta$
  - Because (1) *n* is usually very large, (2) the individual amplitudes a_i are random, and (3) the phases θ_i have a uniform distribution, it can be assumed that (from the central limit theorem) *x* and *y* are both Gaussian variables with means equal to zero and variance:

$$\sigma_x^2 = \sigma_y^2 \equiv \sigma^2$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Because x and y are independent random variables, the joint distribution $p(x,y)$ is

$$p(x,y) = p(x)p(y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

  - The distribution $p(r,\theta)$ can be written as a function of $p(x,y)$:

$$p(r,\theta) = |J|\,p(x,y)$$

$$J \equiv \begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

$$p(r,\theta) = \frac{r}{2\pi\sigma}\exp\left(-\frac{r^2}{2\sigma^2}\right)$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Thus, the Rayleigh distribution has a PDF given by:

$$p_R(r) = \int_0^{2\pi} p(r,\theta)d\theta = \begin{cases} \dfrac{r}{\sigma^2} e^{-r^2/2\sigma^2} & r \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  - The probability that the envelope of the received signal does not exceed a specified value *R* is given by the corresponding cumulative distribution function (CDF):

$$F_R(r) = \int_0^r \frac{u}{\sigma^2} e^{-u^2/2\sigma^2} du = 1 - \exp^{-r^2/2\sigma^2}, \qquad r \geq 0$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Mean: $r_{mean} = E[R] = \int_0^{\infty} r p(r) dr = \sigma\sqrt{\frac{\pi}{2}} = 1.2533\sigma$
  - Variance: $\sigma_r^2 = E[R^2] - E^2[R] = \int_0^{\infty} r^2 p(r) dr - \frac{\sigma^2 \pi}{2}$

$$= \sigma^2\left(2 - \frac{\pi}{2}\right) = 0.4292\sigma^2$$

  - Median value of r is found by solving: $\frac{1}{2} = \int_0^{r_{median}} p(r) dr$

$$r_{median} = 1.177\sigma$$

  - Monents of R are: $E[R^k] = \left(2\sigma^2\right)^{k/2}\Gamma\left(1+\frac{k}{2}\right)$
  - Most likely value:= max { $p_R(r)$ } = σ.

Some Useful Probability Distributions

- Rayleigh distribution

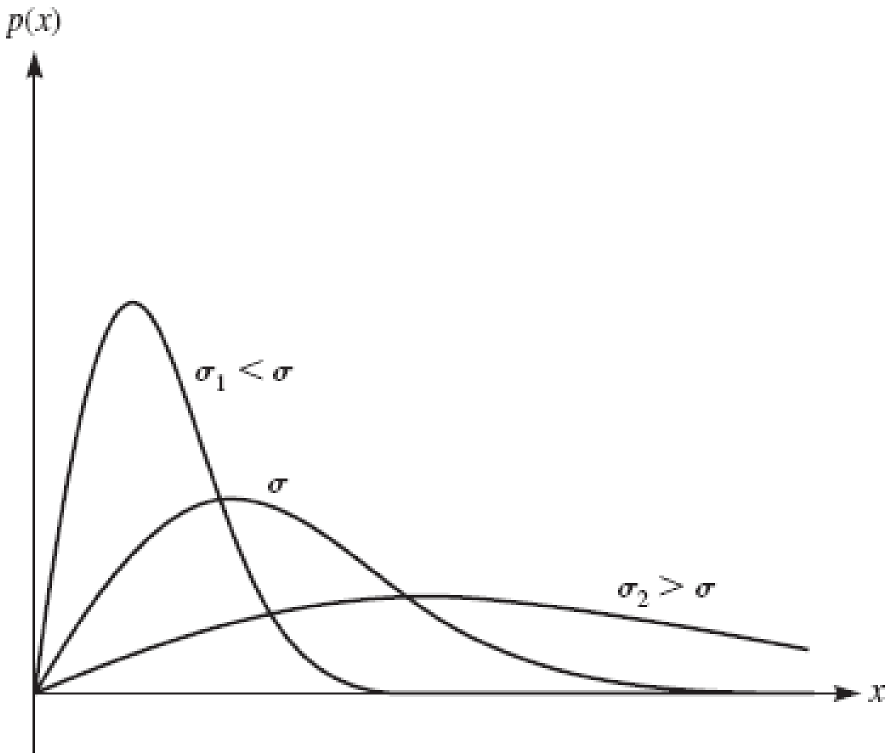


FIGURE 2.3–4
The PDF of the Rayleigh random variable for three different values of σ.

Some Useful Probability Distributions

- Rayleigh distribution
  - Probability That Received Signal Doesn't Exceed A Certain Level (R)

$$F_R(r) = \int_0^r p(u)du$$

$$= \int_0^r \frac{u}{\sigma^2} \exp\left(-\frac{u^2}{2\sigma^2}\right) du$$

$$= -\exp\left(-\frac{u^2}{2\sigma^2}\right)\Bigg|_0^r$$

$$= 1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Mean value:

$$r_{mean} = E[R] = \int_0^\infty r p(r) dr$$

$$= \int_0^\infty \frac{r^2}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \;\; = \int_0^\infty -r d \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

$$= -r \cdot \exp\left(-\frac{r^2}{2\sigma^2}\right)\Bigg|_0^\infty + \int_0^\infty \exp\left(-\frac{r^2}{2\sigma^2}\right) dr$$

$$= \sqrt{2\pi}\sigma \cdot \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr$$

$$= \sqrt{2\pi}\sigma \cdot \frac{1}{2} = \sigma\sqrt{\frac{\pi}{2}} = 1.2533\sigma$$

0-0=0

Some Useful Probability Distributions

- Rayleigh distribution:
  - Mean square value:

$$E[R^2] = \int_0^\infty r^2 p(r) dr$$

$$= \int_0^\infty \frac{r^3}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr = \int_0^\infty -r^2 d \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

0-0=0

$$= -r^2 \cdot \exp\left(-\frac{r^2}{2\sigma^2}\right)\Bigg|_0^\infty + \int_0^\infty 2r \cdot \exp\left(-\frac{r^2}{2\sigma^2}\right) dr$$

$$= -2\sigma^2 \cdot \exp\left(-\frac{r^2}{2\sigma^2}\right)\Bigg|_0^\infty = 2\sigma^2$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Variance:

$$
\begin{aligned}
\sigma_r^2 &= E[R^2] - E^2[R] \\
&= (2\sigma^2) - (\sigma \cdot \sqrt{\frac{\pi}{2}})^2 \\
&= \sigma^2 \cdot \left(2 - \frac{\pi}{2}\right) = 0.4292\sigma^2
\end{aligned}
$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Most likely value
    - Most Likely Value happens when: $dp(r)/dr = 0$

$$\frac{dp(r)}{dr} = \frac{1}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right) - \frac{2r^2}{2\cdot\sigma^4}\cdot\exp\left(-\frac{r^2}{2\sigma^2}\right) = 0$$

$$\Rightarrow \quad r = \sigma$$

$$\Rightarrow \quad p(r)\Big|_{r=\sigma} = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)\Bigg|_{r=\sigma} = \frac{\exp\left(-\frac{1}{2}\right)}{\sigma} = \frac{0.6065}{\sigma}$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Characteristic function

$$\psi_R(jv) = \int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} e^{jvr} dr$$

$$= \int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \cos(vr) dr + j\int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \sin(vr) dr$$

$$= {}_1F_1\left(1, \frac{1}{2}; -\frac{1}{2} v^2\sigma^2\right) + j\sqrt{\frac{1}{2}\pi} v\sigma^2 e^{-v^2\sigma^2/2}$$

where ${}_1F_1\left(1, \frac{1}{2}; -a\right)$ is the confluent hyupergeometric function :

$${}_1F_1(\alpha; \beta; x) = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha + k)\Gamma(\beta) x^k}{\Gamma(\alpha)\Gamma(\beta + k) k!}, \qquad \beta \neq 0, -1, -2, \ldots$$

Some Useful Probability Distributions

- Rayleigh distribution
  - Characteristic function (cont.)

Beaulieu (1990) has shown that ${}_1F_1\left(1,\frac{1}{2};-a\right)$ may be expressed as:

$$ {}_1F_1\left(1,\frac{1}{2};-a\right) = -e^{-a}\sum_{k=0}^{\infty}\frac{a^k}{(2k-1)k!} $$

Some Useful Probability Distributions

- Rice distribution
  - When there is a dominant stationary (non-fading) signal component present, such as a line-of-sight (LOS) propagation path, the small-scale fading envelope distribution is Rice.

$$s_r = \overbrace{r'\exp[j(\omega_0 t + \theta)]}^{\text{scattered waves}} + \overbrace{A\exp(j\omega_0 t)}^{\text{direct waves}}$$

$$\equiv [(x + A) + jy]\exp(j\omega_0 t) \equiv r\exp[j(\omega_0 t + \theta)]$$

$$r^2 = (x + A)^2 + y^2$$

$$x + A = r\cos\theta$$

$$y = r\sin\theta$$

Some Useful Probability Distributions

- Rice distribution
  - By following similar steps described in Rayleigh distribution, we obtain:

$$p(r)=\begin{cases}\dfrac{r}{\sigma_r^2}\exp\left(-\dfrac{r^2+A^2}{2\sigma_r^2}\right)I_0\left(\dfrac{Ar}{\sigma_r^2}\right) & \text{for } \left(\mathrm{A}\geq 0,\mathrm{r}\geq 0\right)\\ 0 & \text{for } (r<0)\end{cases}$$

where

$$I_0\left(\frac{Ar}{\sigma_r^2}\right)=\frac{1}{2\pi}\int_0^{2\pi}\exp\left(\frac{Ar\cos\theta}{\sigma_r^2}\right)d\theta$$

is the modified zeroth-order Bessel function.

$$I_0(x)=\sum_{i=0}^{\infty}\left(\frac{x^i}{i!\cdot 2^i}\right)$$

Some Useful Probability Distributions

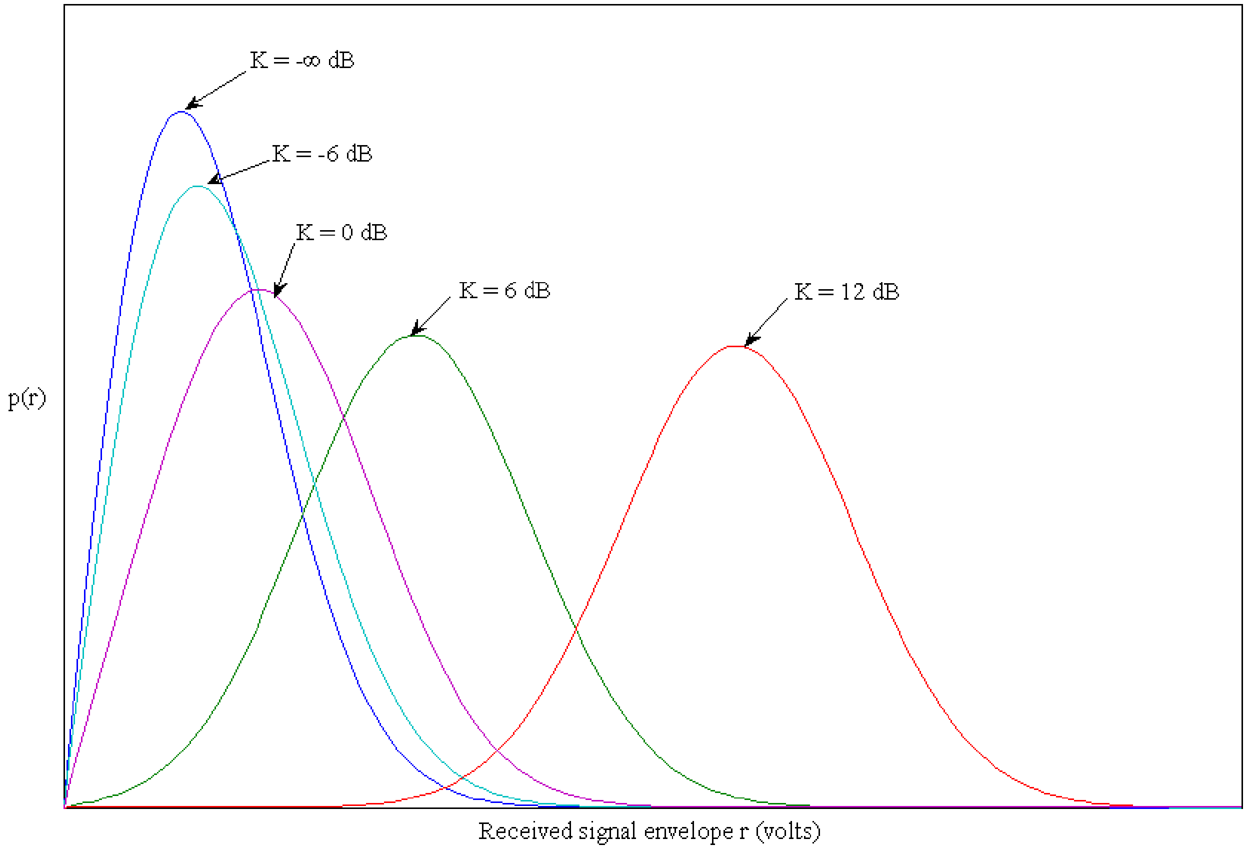
- Rice distribution
  - The Rice distribution is often described in terms of a parameter K which is defined as the ratio between the deterministic signal power and the variance of the multi-path. It is given by $K = A^2/(2\sigma^2)$ or in terms of dB:

$$K(\text{dB}) = 10 \cdot \log \frac{A^2}{2\sigma^2} \quad [\text{dB}]$$

  - The parameter K is known as the Rice factor and completely specifies the Rice distribution.
  - As $A \rightarrow 0$, $K \rightarrow -\infty$ dB, and as the dominant path decreases in amplitude, the Rice distribution degenerates to a Rayleigh distribution.

Some Useful Probability Distributions

- Rice distribution



Some Useful Probability Distributions

- Rice distribution

$$s_r = \sum_{i=1}^{n} a_i \exp\left[j(\omega_0 t + \theta_i)\right] + A\exp(j\omega_0 t)$$

$$= \left[\sum_{i=1}^{n} a_i \exp(j\theta_i)\right] j(\omega_0 t) + A\exp(j\omega_0 t)$$

$$\equiv r'\exp(j\theta)\exp(j\omega_0 t) + A\exp(j\omega_0 t)$$

$$= r'\exp\left[j(\omega_0 t + \theta)\right] + A\exp(j\omega_0 t)$$

$$\equiv \left[(x + A) + jy\right]\exp(j\omega_0 t) \equiv r\exp\left[j(\omega_0 t + \theta)\right]$$

where $r'\exp(j\theta) = \sum_{i=1}^{n} a_i \exp(j\theta_i)$

Some Useful Probability Distributions

- Rice distribution
  - Define: $r'\exp(j\theta) = \sum_{i=1}^{n} a_i \cos\theta_i + j\sum_{i=1}^{n} a_i \sin\theta_i \equiv x + jy$

    We have: $x \equiv \sum_{i=1}^{n} a_i \cos\theta_i$ and $y \equiv \sum_{i=1}^{n} a_i \sin\theta_i$

    and $r^2 = (x+A)^2 + y^2 \quad x + A = r\cos\theta \quad y = r\sin\theta$
  - Because (1) n is usually very large, (2) the individual amplitudes a_i are random, and (3) the phases θ_i have a uniform distribution, it can be assumed that (from the central limit theorem) x and y are both Gaussian variables with means equal to zero and variance:

$$\sigma_x^2 = \sigma_y^2 \equiv \sigma^2$$

Some Useful Probability Distributions

- Rice distribution
  - Because x and y are independent random variables, the joint distribution $p(x,y)$ is

$$p(x,y) = p(x)p(y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

  - The distribution $p(r,\theta)$ can be written as a function of $p(x,y)$:

$$p(r,\theta) = |J|\,p(x,y)$$

$$J \equiv \begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

Some Useful Probability Distributions

- Rice distribution

$$p(r,\theta)=\frac{r}{2\pi\sigma}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

$$=\frac{r}{2\pi\sigma}\exp\left(-\frac{(r\cos\theta-A)^2+(r\sin\theta)^2}{2\sigma^2}\right)$$

$$=\frac{r}{2\pi\sigma}\exp\left(-\frac{r^2+A^2-2Ar\cos\theta}{2\sigma^2}\right)$$

$$=\frac{r}{\sigma}\exp\left(-\frac{r^2+A^2}{2\sigma^2}\right)\frac{1}{2\pi}\exp\left(\frac{Ar\cos\theta}{\sigma^2}\right)$$

Some Useful Probability Distributions

◇ Rice distribution

The Rice distribution has a probability density function (pdf) given by :

$$p(r) = \int_0^{2\pi} p(r,\theta)d\theta$$

$$= \begin{cases} \dfrac{r}{\sigma^2}\exp\left(-\dfrac{r^2+A^2}{2\sigma^2}\right)\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}\exp\left(\dfrac{Ar\cos\theta}{\sigma^2}\right)d\theta & r \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Some Useful Probability Distributions

- Nakagami *m*-distribution
  - Frequently used to characterize the statistics of signals transmitted through multi-path fading channels.
  - PDF is given by Nakagami (1960) as:

$$p_R(r) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m r^{2m-1} e^{-mr^2/\Omega}$$

$$\Omega = E\left(R^2\right)$$

$$m = \frac{\Omega^2}{E[\left(R^2 - \Omega\right)^2]}, \qquad m \geq \frac{1}{2}$$

The parameter m is defined as the ratio of moments, called the fading figure.

Some Useful Probability Distributions

- Nakagami *m*-distribution
  - The *n*-th moment of *R* is:

$$E\left(R^n\right)=\frac{\Gamma\left(m+n/2\right)}{\Gamma\left(m\right)}\left(\frac{\Omega}{m}\right)^{n/2}$$

  - By setting *m*=1, the PDF reduces to a Rayleigh PDF.

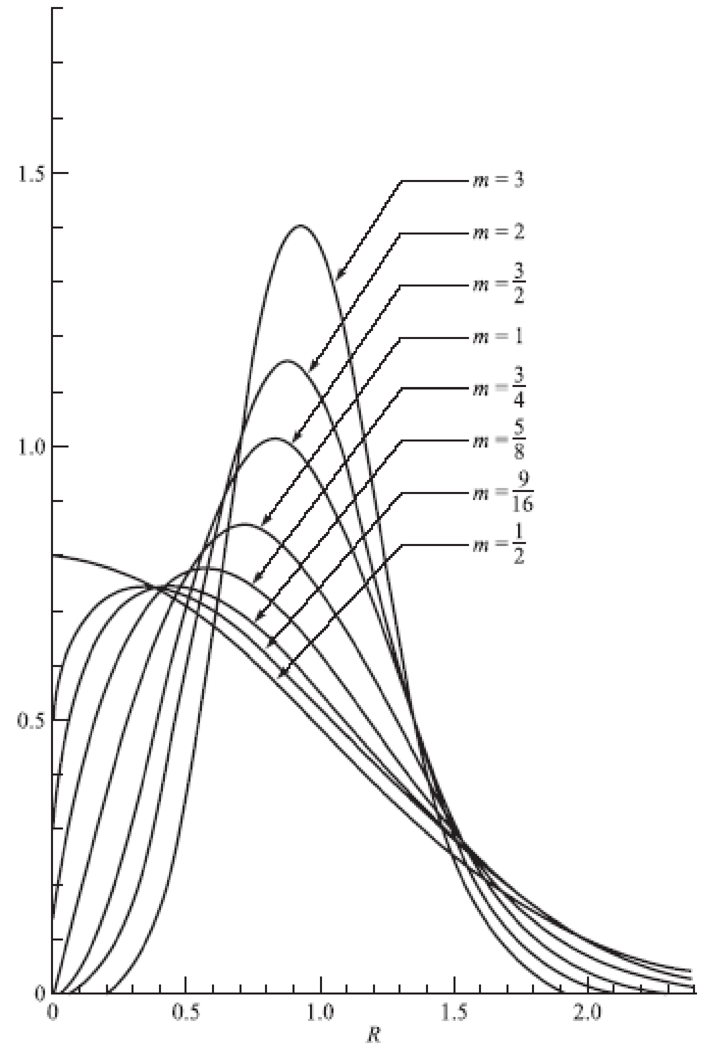


FIGURE 2.3–6
The PDF for the Nakagami *m* distribution, shown with $\Omega = 1$. *m* is the fading figure.

Some Useful Probability Distributions

- Lognormal distribution:
  - Let $X = \ln R$, where X is normally distributed with mean m and variance σ^2.
  - The PDF of R is given by:

$$p(r) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma r} e^{-(\ln r - m)^2 / 2\sigma^2} & (r \geq 0) \\ 0 & (r < 0) \end{cases}$$

  - The lognormal distribution is suitable for modeling the effect of *shadowing* of the signal due to large obstructions, such as tall buildings, in mobile radio communications.

Some Useful Probability Distributions

- Multivariate Gaussian distribution
  - Assume that X_i, i=1,2,…,n, are Gaussian random variables with means m_i, i=1,2,…,n; variances σ_i^2, i=1,2,…,n; and covariances μ_{ij}, i,j=1,2,…,n. The joint PDF of the Gaussian random variables X_i, i=1,2,…,n, is defined as

$$p(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} (\det \mathbf{M})^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{m}_x)' \mathbf{M}^{-1} (\mathbf{x} - \mathbf{m}_x) \right]$$

    - **M** denotes the $n \times n$ covariance matrix with elements $\{\mu_{ij}\}$;
    - **x** denotes the $n \times 1$ column vector of the random variables;
    - $\mathbf{m}_x$ denote the $n \times 1$ column vector of mean values m_i, i=1,2,…,n.
    - $\mathbf{M}^{-1}$ denotes the inverse of **M**.
    - **x’** denotes the transpose of **x**.

Some Useful Probability Distributions

- Multivariate Gaussian distribution (cont.)
  - Given **v** the *n*-dimensional vector with elements υ_i, $i=1,2,\ldots,n$, the characteristic function corresponding to the *n*-dimentional joint PDF is:

$$\psi(j\mathbf{v}) = E\left(e^{j\mathbf{v}'\mathbf{x}}\right) = \exp\left(j\mathbf{m}'_X\mathbf{v} - \frac{1}{2}\mathbf{v}'\mathbf{M}\mathbf{v}\right)$$

Some Useful Probability Distributions

- Bi-variate or two-dimensional Gaussian
  - The bivariate Gaussian PDF is given by:

$$p(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\times\exp\left[-\frac{\sigma_2^2(x_1-m_1)^2-2\rho\sigma_1\sigma_2(x_1-m_1)(x_2-m_2)+\sigma_1^2(x_2-m_2)^2}{2\pi\sigma_1^2\sigma_2^2(1-\rho^2)}\right]$$

$$\mathbf{m}_X=\begin{bmatrix}m_1\\m_2\end{bmatrix},\quad \mathbf{M}=\begin{bmatrix}\sigma_1^2 & \mu_{12}\\ \mu_{12} & \sigma_2^2\end{bmatrix},\quad \mu_{12}=E\left[(x_1-m_1)(x_2-m_2)\right]$$

$$\rho_{ij}=\frac{\mu_{ij}}{\sigma_i\sigma_j},\quad i\neq j,\quad 0\leq\left|\rho_{ij}\right|\leq 1$$

$$\mathbf{M}=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix},\quad \mathbf{M}^{-1}=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{bmatrix}$$

Some Useful Probability Distributions

- Bi-variate or two-dimensional Gaussian
  - ρ is a measure of the correlation between X_1 and X_2.
  - When $\rho=0$, the joint PDF $p(x_1,x_2)$ factors into the product $p(x_1)p(x_2)$, where $p(x_i)$, $i=1,2$, are the marginal PDFs.
  - When the Gaussian random variables X_1 and X_2 are uncorrelated, they are also statistically independent. This property does not hold in general for other distributions.
  - This property can be extended to n-dimensional Gaussian random variables: if $\rho_{ij}=0$ for $i \neq j$, then the random variables X_i, $i=1,2,\ldots,n$, are uncorrelated and, hence, statistically independent.

Upper Bounds on the Tail Probability

- Chebyshev inequality
  - Suppose X is an arbitrary random variable with finite mean m_x and finite variance σ_x^2. For any positive number δ:

$$P\left(\left|X - m_x\right| \geq \delta\right) \leq \frac{\sigma_x^2}{\delta^2}$$

  - Proof:

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - m_x)^2 p(x)dx \geq \int_{|x-m_x|\geq\delta} (x - m_x)^2 p(x)dx$$

$$\geq \delta^2 \int_{|x-m_x|\geq\delta} p(x)dx = \delta^2 P(|X - m_x| \geq \delta)$$

Upper Bounds on the Tail Probability

- Chebyshev inequality
  - Another way to view the Chebyshev bound is working with the zero mean random variable $Y=X-m_x$.
  - Define a function $g(Y)$ as:

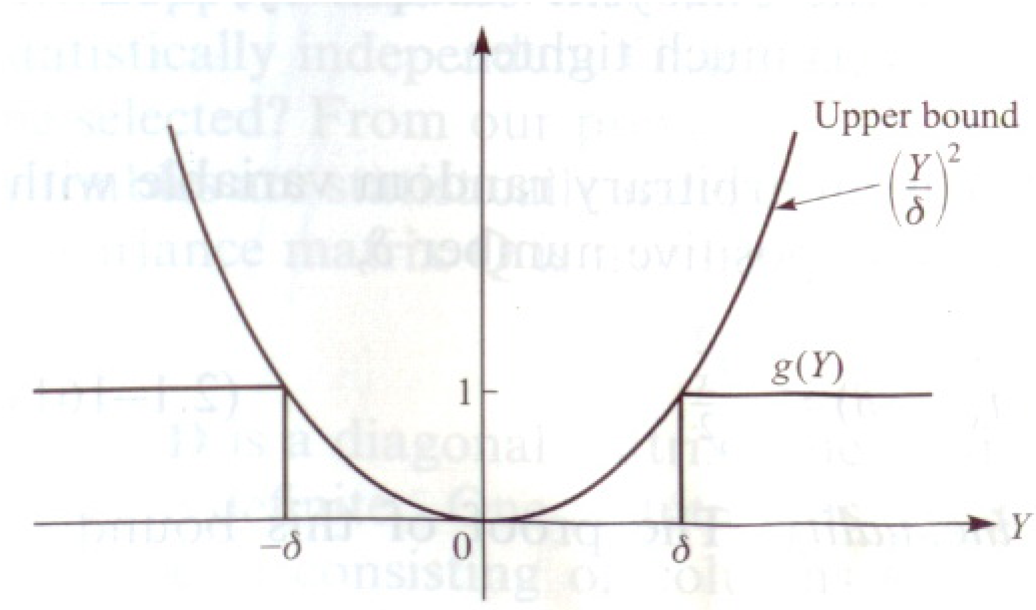
$$g(Y)=\begin{cases}1 & (|Y|\geq\delta)\\ 0 & (|Y|<\delta)\end{cases}$$

$$E\left[g(Y)\right]=\int g(y)p(y)dy$$
$$=\int_{-\infty}^{-\delta}p(y)dy+\int_{\delta}^{\infty}p(y)dy=P\left(|Y|\geq\delta\right)$$

  - Upper-bound $g(Y)$ by the quadratic $(Y/\delta)^2$, i.e. $g(Y)\leq\left(\frac{Y}{\delta}\right)^2$
  - The tail probability $E[g(Y)]\leq E\left(\frac{Y^2}{\delta^2}\right)=\frac{E\left(Y^2\right)}{\delta^2}=\frac{\sigma_y^2}{\delta^2}=\frac{\sigma_x^2}{\delta^2}$

Upper Bounds on the Tail Probability

- Chebychev inequality
  - A quadratic upper bound on $g(Y)$ used in obtaining the tail probability (Chebyshev bound)



  - For many practical applications, the Chebyshev bound is extremely loose.

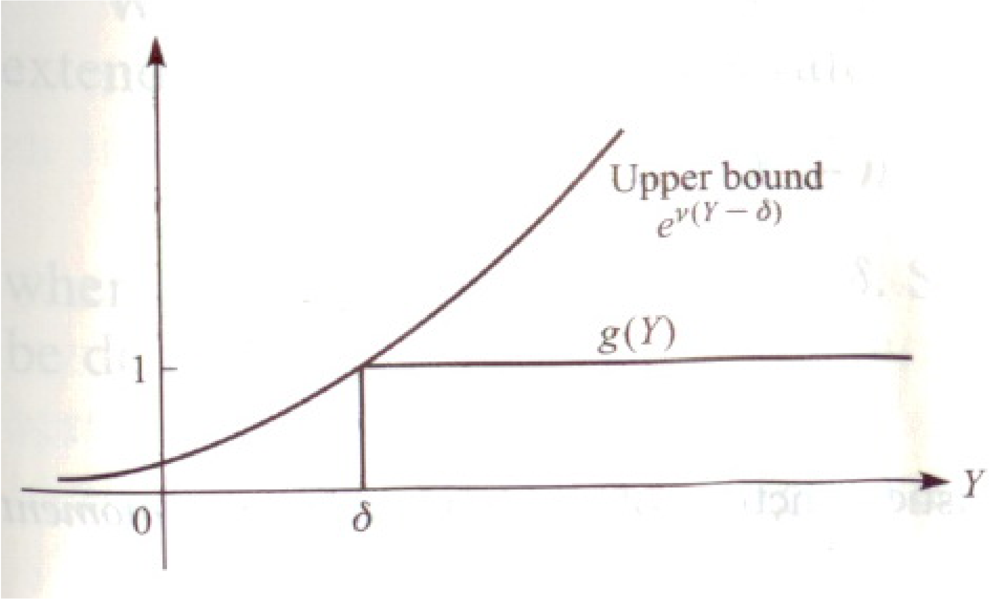
Upper Bounds on the Tail Probability

- Chernoff bound
  - The Chebyshev bound given above involves the area under the two tails of the PDF. In some applications we are interested only in the area under one tail, either in the interval (δ, ∞) or in the interval $(-\infty, \delta)$.
  - In such a case, we can obtain an extremely tight upper bound by over-bounding the function $g(Y)$ by an exponential having a parameter that can be optimized to yield as tight an upper bound as possible.
  - Consider the tail probability in the interval (δ, ∞).

    $g(Y) \le e^{v(Y-\delta)}$ and $g(Y)$ is defined as $g(Y)=\begin{cases}1 & (Y \ge \delta)\\ 0 & (Y < \delta)\end{cases}$

    where $v \ge 0$ is the parameter to be optimized.

Upper Bounds on the Tail Probability

- Chernoff bound
  - 

  - The expected value of $g(Y)$ is

$$E[g(y)] = P(Y \geq \delta) \leq E\left(e^{\nu(Y-\delta)}\right)$$

  - This bound is valid for any υ ≥0.

Upper Bounds on the Tail Probability

- Chernoff bound
  - The tightest upper bound is obtained by selecting the value that minimizes $E(e^{\nu(Y-\delta)})$.
  - A necessary condition for a minimum is:

$$\frac{d}{d\nu}E\left(e^{\nu(Y-\delta)}\right)=0$$

$$\frac{d}{d\nu}E\left(e^{\nu(Y-\delta)}\right)=E\left(\frac{d}{d\nu}e^{\nu(Y-\delta)}\right)=E\left[(Y-\delta)e^{\nu(Y-\delta)}\right]$$

$$=e^{-\nu\delta}\left[E\left(Ye^{\nu Y}\right)-\delta E\left(e^{\nu Y}\right)\right]=0$$

$$\rightarrow \boxed{E\left(Ye^{\nu Y}\right)-\delta E\left(e^{\nu Y}\right)=0} \quad \text{Find } \nu$$

Upper Bounds on the Tail Probability

- Chernoff bound
  - Let $\hat{v}$ be the solution, the upper bound on the one-sided tail probability is:

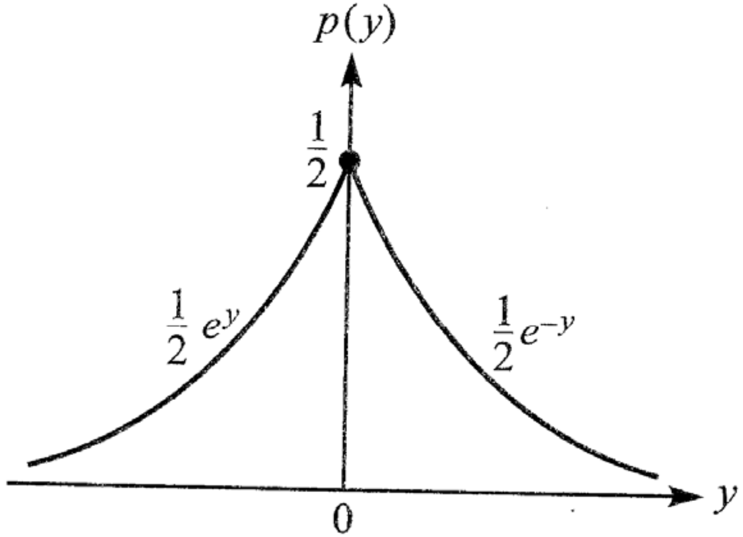
$$P(Y \geq \delta) \leq e^{-\hat{v}\delta} E\left(e^{\hat{v}Y}\right)$$

  - An upper bound on the lower tail probability can be obtained in a similar manner, with the result that

$$P(Y \leq \delta) \leq e^{-\hat{v}\delta} E\left(e^{\hat{v}Y}\right) \qquad \delta < 0$$

Upper Bounds on the Tail Probability

- Chernoff bound
  - Example: Consider the (Laplace) PDF $p(y)=e^{-|y|}/2$.



  - The true tail probability is:

$$P(Y \geq \delta) = \int_{\delta}^{\infty} \frac{1}{2} e^{-y} dy = \frac{1}{2} e^{-\delta}$$

Upper Bounds on the Tail Probability

- Chernoff bound
  - Example (cont.)

$$E\left(Ye^{vY}\right)=\frac{2v}{(v+1)^2(v-1)^2} \qquad E\left(e^{vY}\right)=\frac{1}{(1+v)(1-v)}$$

Since $E\left(Ye^{vY}\right)-\delta E\left(e^{vY}\right)=0$, we obtain $v^2\delta+2v-\delta=0$

$$\hat{v}=\frac{-1+\sqrt{1+\delta^2}}{\delta} \quad (\hat{v}\text{ must be positive})$$

$$\Rightarrow \qquad P(Y\geq\delta)\leq\frac{\delta^2}{2\left(-1+\sqrt{1+\delta^2}\right)}e^{1-\sqrt{1+\delta^2}}$$

for $\delta>>1$: $P(Y\geq\delta)\leq\frac{\delta}{2}e^{-\delta}$ $\left(\frac{1}{\delta^2}\text{ for Chebyshev bound}\right)$

Sums of Random Variables and the Central Limit Theorem

- Sum of random variables
  - Suppose that X_i, i=1,2,…,n, are statistically independent and identically distributed (iid) random variables, each having a finite mean m_x and a finite variance σ_x^2. Let Y be defined as the normalized sum, called the sample mean:

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  - The mean of Y is

$$E(Y) = m_y = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= m_x$$

- Sum of random variables
  - The variance of *Y* is:

$$\sigma_y^2 = E\left(Y^2\right) - m_y^2 = E\left(Y^2\right) - m_x^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(X_i X_j\right) - m_x^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E\left(X_i^2\right) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} E\left(X_i\right)E\left(X_j\right) - m_x^2$$

$$= \frac{1}{n^2}\left[n\left(\sigma_x^2 + m_x^2\right)\right] + \frac{1}{n^2} n(n-1) m_x^2 - m_x^2 = \frac{\sigma_x^2}{n}$$

  - An estimate of a parameter (in this case the mean m_x) that satisfies the conditions that its <u>expected value converges to the true value of the parameter</u> and the <u>variance converges to zero</u> as $n \to \infty$ is said to be a *consistent estimate*.

Stochastic Processes

- Many of random phenomena that occur in nature are functions of time.
- In digital communications, we encounter stochastic processes in:
  - The characterization and modeling of signals generated by information sources;
  - The characterization of communication channels used to transmit the information;
  - The characterization of noise generated in a receiver;
  - The design of the optimum receiver for processing the received random signal.

Stochastic Processes

- Introduction
  - At any given time instant, the value of a stochastic process is a random variable indexed by the parameter t. We denote such a process by $X(t)$.
  - In general, the parameter t is continuous, whereas X may be either continuous or discrete, depending on the characteristics of the source that generates the stochastic process.
  - The noise voltage generated by a single resistor or a single information source represents a single realization of the stochastic process. It is called a *sample function*.

Stochastic Processes

- Introduction (cont.)
  - The set of all possible sample functions constitutes an *ensemble* of sample functions or, equivalently, the *stochastic process* $X(t)$.
  - In general, the number of sample functions in the ensemble is assumed to be extremely large; often it is infinite.
  - Having defined a stochastic process $X(t)$ as an ensemble of sample functions, we may consider the values of the process at any set of time instants $t_1>t_2>t_3>\ldots>t_n$, where n is any positive integer.
  - In general, the random variables $X_{t_i} \equiv X(t_i), i=1,2,\ldots,n$, are characterized statistically by their joint PDF $p(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$.

Stochastic Processes

- Stationary stochastic processes
  - Consider another set of n random variables $X_{t_i+t} \equiv X(t_i+t)$, $i=1,2,\ldots,n$, where t is an arbitrary time shift. These random variables are characterized by the joint PDF $p\left(x_{t_1+t}, x_{t_2+t}, \ldots, x_{t_n+t}\right)$.
  - The jont PDFs of the random variables X_{t_i} and $X_{t_i+t}, i=1,2,\ldots,n$, may or may not be identical. When they are identical, i.e., when

$$p\left(x_{t_1}, x_{t_2}, \ldots, x_{t_n}\right) = p\left(x_{t_1+t}, x_{t_2+t}, \ldots, x_{t_n+t}\right)$$

    for all t and all n, it is said to be stationary in the strict sense.
  - When the joint PDFs are different, the stochastic process is *non-stationary*.

Stochastic Processes

- Averages for a stochastic process are called *ensemble averages*.
- The <u>*n*th moment</u> of the random variable X_{t_i} is defined as:

$$E\left(X_{t_i}^{n}\right)=\int_{-\infty}^{\infty} x_{t_i}^{n}\, p\left(x_{t_i}\right) dx_{t_i}$$

- In general, the value of the <u>*n*th moment</u> will depend on the time instant t_i if the PDF of X_{t_i} depends on t_i.
- When the process is stationary, $p\left(x_{t_i+t}\right)=p\left(x_{t_i}\right)$ for all t. Therefore, the PDF is independent of time, and, as a consequence, the <u>*n*th moment</u> is independent of time.

Stochastic Processes

- Two random variables: $X_{t_i} \equiv X(t_i), i = 1, 2.$
  - The correlation is measured by the *joint moment*:

$$E\left(X_{t_1} X_{t_2}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_{t_1} x_{t_2}\, p\left(x_{t_1}, x_{t_2}\right) dx_{t_1}\, dx_{t_2}$$

  - Since this joint moment depends on the time instants t_1 and t_2, it is denoted by $\varphi(t_1, t_2)$.
  - $\varphi(t_1, t_2)$ is called the *autocorrelation function* of the stochastic process.
  - For a stationary stochastic process, the joint moment is:

$$E(X_{t_1} X_{t_2}) = \phi(t_1, t_2) = \phi(t_1 - t_2) = \phi(\tau)$$

  - $\phi(-\tau) = E(X_{t_1} X_{t_1+\tau}) = E(X_{t_1+\tau} X_{t_1}) = E(X_{t_1'} X_{t_1'-\tau}) = \phi(\tau)$

(Even Function)

  - Average power in the process $X(t)$: $\varphi(0) = E(X_t^2)$.

Stochastic Processes

- Wide-sense stationary (WSS)
  - A *wide-sense stationary process* has the property that the mean value of the process is independent of time (a constant) and where the autocorrelation function satisfies the condition that $\varphi(t_1,t_2)=\varphi(t_1-t_2)$.
  - Wide-sense stationarity is a less stringent condition than strict-sense stationarity.
  - If not otherwise specified, any subsequent discussion in which correlation functions are involved, the less stringent condition (wide-sense stationarity) is implied.

Stochastic Processes

- Auto-covariance function
  - The *auto-covariance function* of a stochastic process is defined as:

$$\mu(t_1,t_2) = E\left\{\left[X_{t_1} - m(t_1)\right]\left[X_{t_2} - m(t_2)\right]\right\}$$
$$= \phi(t_1,t_2) - m(t_1)m(t_2)$$

  - When the process is stationary, the auto-covariance function simplifies to:

$$\mu(t_1,t_2) = \varphi(\tau) - m^2 = \mu(\tau) \quad \text{(function of time difference)}$$

  - For a Gaussian random process, higher-order moments can be expressed in terms of first and second moments. Consequently, a Gaussian random process is completely characterized by its first two moments.

Stochastic Processes

- Averages for a Gaussian process
  - Suppose that $X(t)$ is a Gaussian random process. At time instants $t=t_i$, $i=1,2,\ldots,n$, the random variables X_{t_i}, $i=1,2,\ldots,n$, are jointly Gaussian with mean values $m(t_i)$, $i=1,2,\ldots,n$, and auto-covariances:

$$\mu\left(t_i,t_j\right)=E\left[\left(X_{t_i}-m\left(t_i\right)\right)\left(X_{t_j}-m\left(t_j\right)\right)\right],\quad i,j=1,2,\ldots,n.$$

  - If we denote the $n\times n$ covariance matrix with elements $\mu(t_i,t_j)$ by $\mathbf{M}$ and the vector of mean values by $\mathbf{m}_x$, the joint PDF of the random variables X_{t_i}, $i=1,2,\ldots,n$, is given by:

$$p\left(x_1,x_2,\ldots,x_n\right)=\frac{1}{\left(2\pi\right)^{n/2}\left(\det\mathbf{M}\right)^{1/2}}\exp\left[-\frac{1}{2}\left(\mathbf{x}-\mathbf{m}_x\right)'\mathbf{M}^{-1}\left(\mathbf{x}-\mathbf{m}_x\right)\right]$$

  - If the Gaussian process is wide-sense stationary, it is also strict-sense stationary.

Stochastic Processes

- Averages for joint stochastic processes
  - Let $X(t)$ and $Y(t)$ denote two stochastic processes and let $X_{t_i} \equiv X(t_i)$, $i=1,2,\ldots,n$, $Y_{t'_j} \equiv Y(t'_j)$, $j=1,2,\ldots,m$, represent the random variables at times $t_1>t_2>t_3>\ldots>t_n$, and $t'_1>t'_2>t'_3>\ldots>t'_m$, respectively. The two processes are characterized statistically by their joint PDF:

$$p\left(x_{t_1}, x_{t_2}, \ldots, x_{t_n}, y_{t'_1}, y_{t'_2}, \ldots, y_{t'_m}\right)$$

  - The *cross-correlation function* of $X(t)$ and $Y(t)$, denoted by $\varphi_{xy}(t_1,t_2)$, is defined as the joint moment:

$$\phi_{xy}(t_1,t_2) = E(X_{t_1} Y_{t_2}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_{t_1} y_{t_2} p(x_{t_1}, y_{t_2}) dx_{t_1} dy_{t_2}$$

  - The *cross-covariance* is:

$$\mu_{xy}(t_1,t_2) = \phi_{xy}(t_1,t_2) - m_x(t_1) m_y(t_2)$$

Stochastic Processes

- Averages for joint stochastic processes
  - When the process are jointly and individually stationary, we have $\varphi_{xy}(t_1,t_2)=\varphi_{xy}(t_1-t_2)$, and $\mu_{xy}(t_1,t_2)=\mu_{xy}(t_1-t_2)$:

$$\phi_{xy}(-\tau) = E(X_{t_1} Y_{t_1+\tau}) = E(X_{t_1'-\tau} Y_{t_1'}) = E(Y_{t_1'} X_{t_1'-\tau}) = \phi_{yx}(\tau)$$

  - The stochastic processes $X(t)$ and $Y(t)$ are said to be *statistically independent* if and only if :

$$p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}, y_{t_1'}, y_{t_2'}, \ldots, y_{t_m'}) = p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) p(y_{t_1'}, y_{t_2'}, \ldots, y_{t_m'})$$

for all choices of t_i and t'_i and for all positive integers n and m.

  - The processes are said to be *uncorrelated* if

$$\phi_{xy}(t_1, t_2) = E(X_{t_1}) E(Y_{t_2}) \quad \Rightarrow \quad \mu_{xy}(t_1, t_2) = 0$$

Stochastic Processes

- Complex-valued stochastic process
  - A complex-valued stochastic process $Z(t)$ is defined as:

$$Z(t) = X(t) + jY(t)$$

    where $X(t)$ and $Y(t)$ are stochastic processes.
  - The joint PDF of the random variables $Z_{t_i} \equiv Z(t_i)$, $i=1,2,\ldots,n$, is given by the joint PDF of the components (X_{t_i}, Y_{t_i}), $i=1,2,\ldots,n$. Thus, the PDF that characterizes Z_{t_i}, $i=1,2,\ldots,n$, is:

$$p(x_{t_1}, x_{t_2}, \ldots, x_{t_n}, y_{t_1}, y_{t_2}, \ldots, y_{t_n})$$

  - The autocorrelation function is defined as:

$$\phi_{zz}(t_1,t_2) \equiv \frac{1}{2} E\left(Z_{t_1} Z_{t_2}^*\right) = \frac{1}{2} E\left[\left(X_{t_1} + jY_{t_1}\right)\left(X_{t_2} - jY_{t_2}\right)\right] \qquad \textbf{(**)}$$

$$= \frac{1}{2}\left\{ \phi_{xx}(t_1,t_2) + \phi_{yy}(t_1,t_2) + j\left[\phi_{yx}(t_1,t_2) - \phi_{xy}(t_1,t_2)\right] \right\}$$

Stochastic Processes

- Averages for joint stochastic processes:
  - When the processes $X(t)$ and $Y(t)$ are jointly and individually stationary, the autocorrelation function of $Z(t)$ becomes:

$$\phi_{zz}(t_1, t_2) = \phi_{zz}(t_1 - t_2) = \phi_{zz}(\tau)$$

  - $\varphi_{ZZ}(\tau) = \varphi^*_{ZZ}(-\tau)$ because from (**): $\phi_{zz}(t_1, t_2) \equiv \frac{1}{2} E(Z_{t_1} Z^*_{t_2})$

$$\phi^*_{zz}(\tau) = \frac{1}{2} E(Z^*_{t_1} Z_{t_1-\tau}) = \frac{1}{2} E(Z^*_{t'_1+\tau} Z_{t'_1}) = \frac{1}{2} E(Z_{t'_1} Z^*_{t'_1+\tau}) = \phi_{zz}(-\tau)$$

Stochastic Processes

- Averages for joint stochastic processes:
  - Suppose that $Z(t)=X(t)+jY(t)$ and $W(t)=U(t)+jV(t)$ are two complex-valued stochastic processes. The cross-correlation functions of $Z(t)$ and $W(t)$ is defined as:

$$\phi_{zw}(t_1,t_2) \equiv \frac{1}{2} E(Z_{t_1} \cdot W_{t_2}^*)$$

$$= \frac{1}{2} E\left[\left(X_{t_1} + jY_{t_1}\right)\left(U_{t_2} - jV_{t_2}\right)\right]$$

$$= \frac{1}{2}\left\{ \phi_{xu}(t_1,t_2) + \phi_{yv}(t_1,t_2) + j\left[\phi_{yu}(t_1,t_2) - \phi_{xv}(t_1,t_2)\right] \right\}$$

  - When $X(t)$, $Y(t)$, $U(t)$ and $V(t)$ are pairwise-stationary, the cross-correlation function become functions of the time difference.
  - $\phi^*_{zw}(\tau) = \frac{1}{2} E(Z^*_{t_1} W_{t_1-\tau}) = \frac{1}{2} E(Z^*_{t_1'+\tau} W_{t_1'}) = \frac{1}{2} E(W_{t_1'} Z^*_{t_1'+\tau}) = \phi_{wz}(-\tau)$

Power Density Spectrum

- A signal can be classified as having either a finite (nonzero) average power (infinite energy) or finite energy.
- The frequency content of a finite energy signal is obtained as the Fourier transform of the corresponding time function.
- If the signal is periodic, its energy is infinite and, consequently, its Fourier transform does not exist. The mechanism for dealing with periodic signals is to represent them in a Fourier series.

Power Density Spectrum

- A stationary stochastic process is an infinite energy signal, and, hence, its Fourier transform does not exist.
- The spectral characteristic of a stochastic signal is obtained by computing the Fourier transform of the autocorrelation function.
- The distribution of power with frequency is given by the function:

$$\Phi(f)=\int_{-\infty}^{\infty}\phi(\tau)e^{-j2\pi f\tau}d\tau$$

- The inverse Fourier transform relationship is:

$$\phi(\tau)=\int_{-\infty}^{\infty}\Phi(f)e^{j2\pi f\tau}df$$

Power Density Spectrum

- $\phi(0)=\int_{-\infty}^{\infty}\Phi(f)df = E\left(\left|X_t\right|^2\right)\geq 0$
- Since φ(0) represents the average power of the stochastic signal, which is the area under Φ(*f*), Φ(*f*) is the distribution of power as a function of frequency.
- Φ(*f*) is called the *power density spectrum* of the stochastic process.
- If the stochastic process is real, φ(τ) is real and even, and, hence Φ(*f*) is real and even. (easy to prove from definition)

  (from definition) (P.94)
- If the stochastic process is complex, φ(τ)=φ*(-τ) and Φ(*f*) is real because:

  (P.101)

$$\Phi^*(f)=\int_{-\infty}^{\infty}\phi^*(\tau)e^{j2\pi f\tau}d\tau=\int_{-\infty}^{\infty}\phi^*(-\tau')e^{-j2\pi f\tau'}d\tau'$$

$$=\int_{-\infty}^{\infty}\phi(\tau)e^{-j2\pi f\tau}d\tau=\Phi(f)$$

Power Density Spectrum

- Cross-power density spectrum
  - For two jointly stationary stochastic processes $X(t)$ and $Y(t)$, which have a cross-correlation function $\varphi_{xy}(\tau)$, the Fourier transform is:

$$\Phi_{xy}(f)=\int_{-\infty}^{\infty}\phi_{xy}(\tau)e^{-j2\pi f\tau}d\tau$$

  - $\Phi_{xy}(f)$ is called the *cross-power density spectrum*.
  - $$\Phi^{*}_{xy}(f)=\int_{-\infty}^{\infty}\phi^{*}_{xy}(\tau)e^{j2\pi f\tau}d\tau=\int_{-\infty}^{\infty}\phi^{*}_{xy}(-\tau)e^{-j2\pi f\tau}d\tau$$

$$=\int_{-\infty}^{\infty}\phi_{yx}(\tau)e^{-j2\pi f\tau}d\tau=\Phi_{yx}(f)$$

  - If $X(t)$ and $Y(t)$ are real stochastic processes

$$\Phi^{*}_{xy}(f)=\int_{-\infty}^{\infty}\phi_{xy}(\tau)e^{j2\pi f\tau}d\tau=\Phi_{xy}(-f)\rightarrow\Phi_{yx}(f)=\Phi_{xy}(-f)$$

Response of a Linear Time-Invariant System to a Random Input Signal

- Consider a linear time-invariant system (filter) that is characterized by its *impulse response* $h(t)$ or equivalently, by its *frequency response* $H(f)$, where $h(t)$ and $H(f)$ are a *Fourier transform pair*. Let $x(t)$ be the input signal to the system and let $y(t)$ denote the output signal.

$$y(t) = \int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau$$

- Suppose that $x(t)$ is a sample function of a stationary stochastic process $X(t)$. Since convolution is a linear operation performed on the input signal $x(t)$, the expected value of the integral is equal to the integral of the expected value.

$$m_y = E\left[Y(t)\right] = \int_{-\infty}^{\infty} h(\tau) E\left[X(t-\tau)\right] d\tau$$

$$= m_x \int_{-\infty}^{\infty} h(\tau) d\tau = m_x H(0)$$

stationary

- The mean value of the output process is a constant.

Response of a Linear Time-Invariant System to a Random Input Signal

- The autocorrelation function of the output is:

$$\phi_{yy}(t_1,t_2)=\frac{1}{2}E\left(Y_{t_1}Y_{t_2}^*\right)$$

$$=\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\beta)h^*(\alpha)E\left[X(t_1-\beta)X^*(t_2-\alpha)\right]d\alpha d\beta$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\beta)h^*(\alpha)\phi_{xx}(t_1-t_2+\alpha-\beta)d\alpha d\beta$$

- If the input process is stationary, the output is also stationary:

$$\phi_{yy}(\tau)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h^*(\alpha)h(\beta)\phi_{xx}(\tau+\alpha-\beta)d\alpha d\beta$$

Response of a Linear Time-Invariant System to a Random Input Signal

- The power density spectrum of the output process is:

$$\Phi_{yy}(f)=\int_{-\infty}^{\infty}\phi_{yy}(\tau)e^{-j2\pi f\tau}d\tau$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h^{*}(\alpha)h(\beta)\phi_{xx}(\tau+\alpha-\beta)e^{-j2\pi f\tau}d\tau d\alpha d\beta$$

$$=\Phi_{xx}(f)\left|H(f)\right|^{2}$$

(by making $\tau_0=\tau+\alpha-\beta$)

- The power density spectrum of the output signal is the product of the power density spectrum of the input multiplied by the magnitude squared of the frequency response of the system.

Response of a Linear Time-Invariant System to a Random Input Signal

- When the autocorrelation function $\varphi_{yy}(\tau)$ is desired, it is usually easier to determine the power density spectrum $\Phi_{yy}(f)$ and then to compute the inverse transform.

$$\phi_{yy}(\tau) = \int_{-\infty}^{\infty} \Phi_{\mathrm{yy}}(f) e^{j2\pi f\tau} df$$

$$= \int_{-\infty}^{\infty} \Phi_{\mathrm{xx}}(f) \left| H(f) \right|^2 e^{j2\pi f\tau} df$$

- The average power in the output signal is:

$$\phi_{yy}(0) = \int_{-\infty}^{\infty} \Phi_{\mathrm{xx}}(f) \left| H(f) \right|^2 df$$

- Since $\varphi_{yy}(0)=E(|Y_t|^2)$, we have:

$$\int_{-\infty}^{\infty} \Phi_{\mathrm{xx}}(f) \left| H(f) \right|^2 df \geq 0 \qquad \text{valid for any } H(f).$$

Response of a Linear Time-Invariant System to a Random Input Signal

- Suppose we let $|H(f)|^2=1$ for any arbitrarily small interval $f_1 \le f \le f_2$, and $H(f)=0$ outside this interval. Then, we have:

$$\int_{f_1}^{f_2} \Phi_{xx}(f)\,df \ge 0$$

  This is possible if an only if $\Phi_{xx}(f) \ge 0$ for all f.

- Conclusion: $\Phi_{xx}(f) \ge 0$ for all f.

Response of a Linear Time-Invariant System to a Random Input Signal

- Cross-correlation function between $y(t)$ and $x(t)$
  - $$\phi_{yx}(t_1,t_2)=\frac{1}{2}E\left(Y_{t_1}X_{t_2}^{*}\right)=\frac{1}{2}\int_{-\infty}^{\infty}h(\alpha)E\left[X(t_1-\alpha)X^{*}(t_2)\right]d\alpha$$
    Function of t_1-t_2
    $$=\int_{-\infty}^{\infty}h(\alpha)\phi_{xx}(t_1-t_2-\alpha)d\alpha=\phi_{yx}(t_1-t_2)$$
    The stochastic processes $X(t)$ and $Y(t)$ are jointly stationary.
  - With t_1-t_2=τ, we have:
    $$\phi_{yx}(\tau)=\int_{-\infty}^{\infty}h(\alpha)\phi_{xx}(\tau-\alpha)d\alpha$$
  - In the frequency domain, we have:
    $$\Phi_{yx}(f)=\Phi_{xx}(f)H(f)$$

Discrete-Time Stochastic Signals and Systems

- Discrete-time stochastic process $X(n)$ consisting of an ensemble of sample sequences $\{x(n)\}$ are usually obtained by uniformly sampling a continuous-time stochastic process.
- The *mth moment* of $X(n)$ is defined as:

$$E\left[X_n^m\right]=\int_{-\infty}^{\infty} X_n^m p(X_n)dX_n$$

- The *autocorrelation sequence* is:

$$\phi(n,k)=\frac{1}{2}E\left(X_n X_k^*\right)=\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X_n X_k^* p(X_n, X_k)dX_n dX_k$$

- The *auto-covariance sequences* is:

$$\mu(n,k)=\phi(n,k)-\frac{1}{2}E(X_n)E\left(X_k^*\right)$$

- For a stationary process, we have φ(*n*,*k*)≡φ(*n*-*k*), μ(*n*,*k*)≡μ(*n*-*k*), and

$$\mu(n-k)=\phi(n-k)-\frac{1}{2}|m_x|^2$$

where $m_x=E(X_n)$ is the mean value.

- A discrete-time stationary process has infinite energy but a finite average power, which is given as:

$$E\left(|X_n|^2\right)=\phi(0)$$

- The power density spectrum for the discrete-time process is obtained by computing the Fourier transform of φ(*n*).

$$\Phi(f)=\sum_{n=-\infty}^{\infty}\phi(n)e^{-j2\pi fn}$$

- The inverse transform relationship is:

$$\phi(n) = \int_{-1/2}^{1/2} \Phi(f) e^{j2\pi fn} df$$

- The power density spectrum $\Phi(f)$ is periodic with a period $f_p=1$. In other words, $\Phi(f+k)=\Phi(f)$ for $k=0,\pm1,\pm2,\ldots$.
- The periodic property is a characteristic of the Fourier transform of any discrete-time sequence.

- Response of a discrete-time, linear time-invariant system to a stationary stochastic input signal.
  - The system is characterized in the time domain by its unit sample response $h(n)$ and in the frequency domain by the frequency response $H(f)$.

$$H(f)=\sum_{n=-\infty}^{\infty}h(n)e^{-j2\pi fn}$$

  - The response of the system to the stationary stochastic input signal $X(n)$ is given by the convolution sum:

$$y(n)=\sum_{k=-\infty}^{\infty}h(k)x(n-k)$$

- Response of a discrete-time, linear time-invariant system to a stationary stochastic input signal.
  - The mean value of the output of the system is:

$$m_y = E\left[y(n)\right] = \sum_{k=-\infty}^{\infty} h(k) E\left[x(n-k)\right]$$

$$= m_x \sum_{k=-\infty}^{\infty} h(k) = m_x H(0) \qquad \text{(P. 107)}$$

where $H(0)$ is the zero frequency [direct current (DC)] gain of the system.

- The autocorrelation sequence for the output process is:

$$\phi_{yy}(k) = \tfrac{1}{2} E\left[y^*(n) y(n+k) \right]$$

$$= \tfrac{1}{2} \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h^*(i) h(j) E\left[x^*(n-i) x(n+k-j) \right]$$

$$= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h^*(i) h(j) \phi_{xx}(k-j+i)$$

- By taking the Fourier transform of $\varphi_{yy}(k)$, we obtain the corresponding frequency domain relationship:

$$\Phi_{yy}(f) = \Phi_{xx}(f) \left| H(f) \right|^2 \qquad \text{(P. 109)}$$

- $\Phi_{yy}(f)$, $\Phi_{xx}(f)$, and $H(f)$ are periodic functions of frequency with period $f_p=1$.

Cyclostationary Processes

- For signals that carry digital information, we encounter stochastic processes with statistical averages that are periodic.
- Consider a stochastic process of the form:

$$X(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT)$$

where $\{a_n\}$ is a discrete-time sequence of random variables with mean m_a=E(a_n) for all n and autocorrelation sequence $\varphi_{aa}(k)=E(a^*_n a_{n+k})/2$.
- The signal $g(t)$ is deterministic.
- The sequence $\{a_n\}$ represents the digital information sequence that is transmitted over the communication channel and $1/T$ represents the rate of transmission of the information symbols.

Cyclostationary Processes

- The mean value is:

$$E\left[X(t)\right]=\sum_{n=-\infty}^{\infty}E(a_n)g(t-nT)$$

$$=m_a\sum_{n=-\infty}^{\infty}g(t-nT)$$

The mean is time-varying and it is periodic with period T.

- The autocorrelation function of $X(t)$ is:

$$\phi_{xx}(t+\tau,t)=\tfrac{1}{2}E\left[X(t+\tau)X^*(t)\right]$$

$$=\tfrac{1}{2}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}E\left(a_n^*a_m\right)g^*(t-nT)g(t+\tau-mT)$$

$$=\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\phi_{aa}(m-n)g^*(t-nT)g(t+\tau-mT)$$

Cyclostationary Processes

- We observe that

$$\phi_{xx}\left(t+\tau+kT, t+kT\right)=\phi_{xx}\left(t+\tau, t\right)$$

for $k=\pm1,\pm2,\ldots$. Hence, the autocorrelation function of $X(t)$ is also periodic with period T.

- Such a stochastic process is called *cyclostationary* or *periodically stationary*.
- Since the autocorrelation function depends on both the variables t and τ, its frequency domain representation requires the use of a two-dimensional Fourier transform.
- The *time-average autocorrelation function* over a single period is defined as:

$$\bar{\phi}_{xx}\left(\tau\right)=\frac{1}{T}\int_{-T/2}^{T/2}\phi_{xx}\left(t+\tau, t\right)dt$$

Cyclostationary Processes

- Thus, we eliminate the tie dependence by dealing with the average autocorrelation function.
- The Fourier transform of $\varphi_{xx}(\tau)$ yields the *average power density spectrum* of the cyclostationary stochastic process.
- This approach allows us to simply characterize cyclostationary process in the frequency domain in terms of the power spectrum.
- The power density spectrum is:

$$\Phi_{xx}(f) = \int_{-\infty}^{\infty} \bar{\phi}_{xx}(\tau) e^{-j2\pi f\tau} d\tau$$