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

PPT ON PROBABILITY THEORY &STOCHASTIC PROCESS

II B.Tech I semester (JNTUH-R15)

Prepared by
Ms.G.Mary Swarna Latha
(Assistant professor)
Mr.G.Anil kumar reddy
(Assistant professor)

probability introduced through sets and relative frequency

- Experiment:- a ***random experiment*** is an action or process that leads to one of several possible outcomes

Experiment	Outcomes
Flip a coin	Heads, Tails
Exam Marks	Numbers: 0, 1, 2, ..., 100
Assembly Time	t > 0 seconds
Course Grades	F, D, C, B, A, A+

Sample Space

- List: “Called the Sample Space”
- Outcomes: “Called the Simple Events”

This list must be ***exhaustive***, i.e. ALL possible outcomes included.

- Die roll {1,2,3,4,5} Die roll {1,2,3,4,5,6}

- The list must be ***mutually exclusive***, i.e. no two outcomes can occur at the same time:
- Die roll {odd number or even number}
- Die roll{ number less than 4 or even number}

Sample Space

- A list of exhaustive [don't leave anything out] and mutually exclusive outcomes [impossible for 2 different events to occur in the same experiment] is called a ***sample space*** and is denoted by S.

- The outcomes are denoted by $O_1, O_2, \ldots, O_k$

- Using notation from set theory, we can represent the sample space and its outcomes as:

- $S = \{O_1, O_2, \ldots, O_k\}$

- Given a sample space $S = \{O_1, O_2, \ldots, O_k\}$, the ***probabilities*** assigned to the outcome must satisfy these requirements:

(1) The probability of any outcome is between 0 and 1

- i.e. $0 \leq P(O_i) \leq 1$ for each *i*, and

(2) The sum of the probabilities of all the outcomes equals 1

- i.e. $P(O_1) + P(O_2) + \ldots + P(O_k) = 1$

$$\sum_{i=1}^{k} P(O_i) = 1$$

Relative Frequency

Random experiment with sample space S. we shall assign non-negative number called probability to each event in the sample space.

Let A be a particular event in S. then “the probability of event A” is denoted by P(A).

Suppose that the random experiment is repeated n times, if the event A occurs n_A times, then the probability of event A is defined as “Relative frequency “

- **Relative Frequency Definition**: The probability of an
- event A is defined as

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

Axioms of Probability

For any event *A*, we assign a number *P*(*A*), called the probability of the event *A*. This number satisfies the following three conditions that act the axioms of probability.

(i) $P(A) \geq 0$ (Probability is a nonnegative number)

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

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

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

Events

- The ***probability of an event*** is the **sum** of the probabilities of the simple events that constitute the event.
- E.g. (assuming a fair die) S = {1, 2, 3, 4, 5, 6} and
- P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6
- Then:
- P(EVEN) = P(2) + P(4) + P(6) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2

Conditional Probability

- ***Conditional probability*** is used to determine how two events are related; that is, we can determine the probability of one event ***given*** the occurrence of another related event.
- Experiment: random select one student in class.
- P(randomly selected student is male) =
- P(randomly selected student is male/student is on 3rd row) =
- Conditional probabilities are written as **P(A | B)** and read as "the probability of A *given* B" and is calculated as

$$P(A \mid B) = \frac{P(A \; and \; B)}{P(B)}$$

- P(A and B) = P(A)*P(B/A) = P(B)*P(A/B) both are true
- Keep this in mind!

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

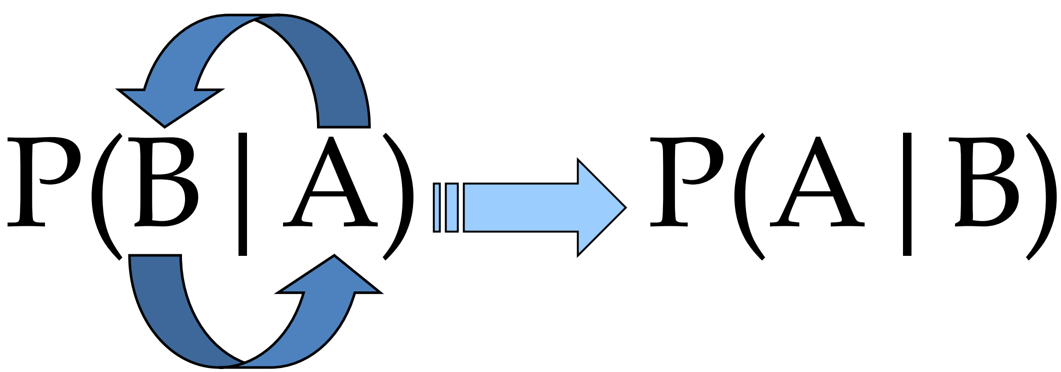
$$P(B \mid A) = \frac{P(A\ and\ B)}{P(A)}$$

Bayes' Law

- Bayes' Law is named for Thomas Bayes, an eighteenth century mathematician.

- In its most basic form, **<u>if we know</u>** P(B | A),

- we can apply Bayes' Law to determine P(A | B)



- The probabilities P(A) and $P(A^C)$ are called ***prior probabilities*** because they are determined ***prior*** to the decision about taking the preparatory course.
- The conditional probability P(A | B) is called a ***posterior probability*** (or revised probability), because the prior probability is revised ***after*** the decision about taking the preparatory course.

Total probability theorem

- Take events A_i for I = 1 to k to be:
  - Mutually exclusive: $A_i \cap A_j = 0$ for all i,j
  - Exhaustive: $A_1 \cup \cdots \cup A_k = S$

For any event B on S

$$p(B) = p(B|A_1)p(A_1) + \cdots + p(B|A_k)p(A_k)$$

$$p(B) = \sum_{i=1}^{k} p(B|A_i)p(A_i)$$

Bayes theorem follows

$$p(A_j|B) = \frac{p(A_j \cap B)}{p(B)} = \frac{p(B|A_j) \cdot p(A)}{\sum_{i=1}^{k} p(B|A_i)p(A_i)}$$

Independence

- Do A and B depend on one another?
  - Yes! B more likely to be true if A.
  - A should be more likely if B.
- If Independent

$$p(A \cap B) = p(A) \cdot p(B)$$

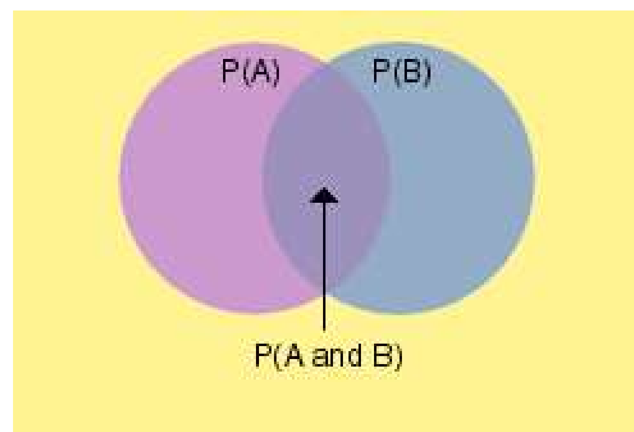
$$p(A|B) = p(A) \quad p(B|A) = p(B)$$

- If Dependent

$$p(A \cap B) \neq p(A) \cdot p(B)$$

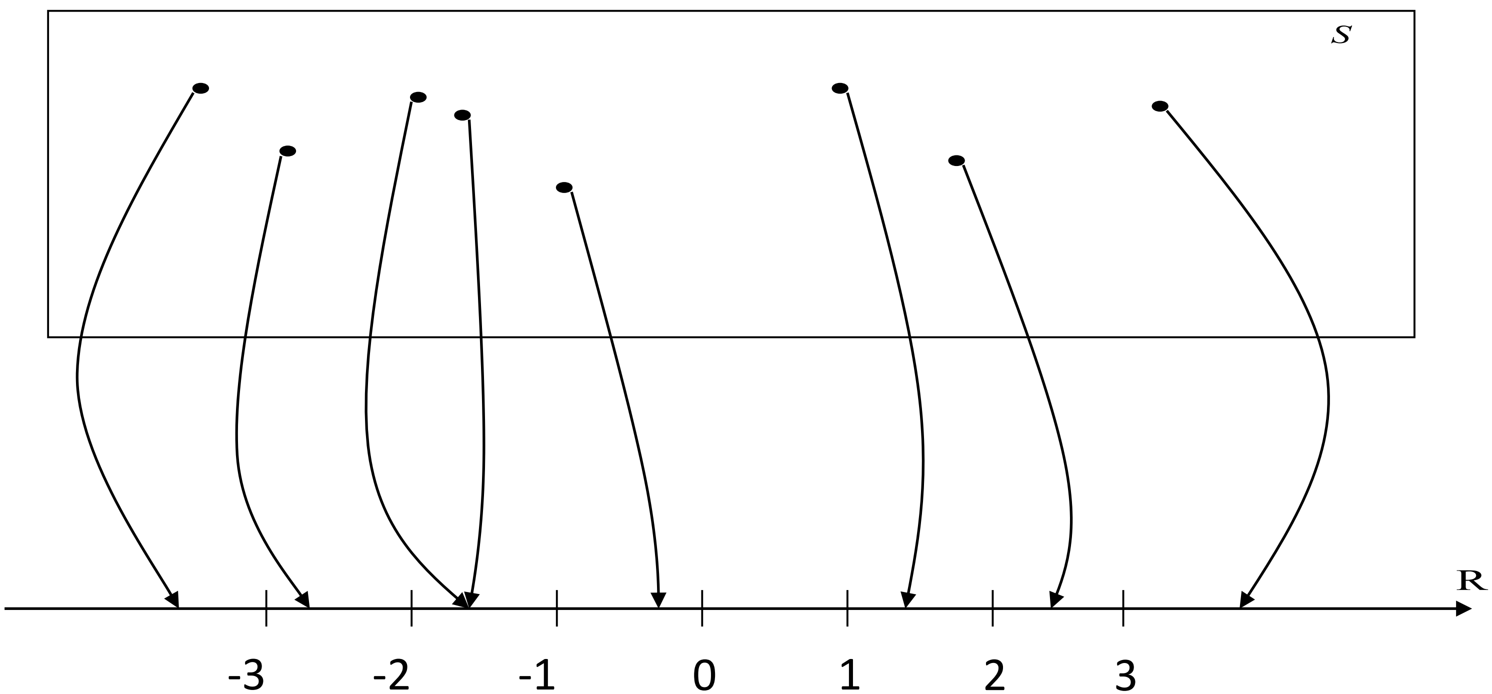
$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$p(A \cap B) = p(B|A) \cdot p(A)$$



Random variable

- Random variable
  - A numerical value to each outcome of a particular experiment



- Example 1 : Machine Breakdowns
  - Sample space : $S = \{electrical, mechanical, misuse\}$
  - Each of these failures may be associated with a repair cost
  - State space : $\{50, 200, 350\}$
  - Cost is a random variable : 50, 200, and 350
- Probability Mass Function (p.m.f.)
  - A set of probability value assigned to each of the values taken by the discrete random variable x_i
  - $0 \le p_i \le 1$ and $\sum_i p_i = 1$
  - Probability : $P(X = x_i) = p_i$

Continuous and Discrete random variables

- **Discrete** random variables have a countable number of outcomes
  - Examples: Dead/alive, treatment/placebo, dice, counts, etc.
- **Continuous** random variables have an infinite continuum of possible values.
  - Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.

- **Distribution function**:

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

- If $F_X(x)$ is a continuous function of x, then X is a continuous random variable.
  - $F_X(x)$: discrete in x → Discrete rv's
  - $F_X(x)$: piecewise continuous → Mixed rv's
  - **PROPERTIES:**
    - $0 \leq F_X(x) \leq 1, \ -\infty < x < \infty$
    - $F_X(x)$: monotonically increasing func. of x
    - $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$
-

Probability Density Function (pdf)

- X : continuous rv, then, $f(x) = \frac{dF(x)}{dx}$ is the pdf of X.

$$CDF \longleftrightarrow pdf$$

$$P(X \leq x) \ = \ F(x) = \int_{-\infty}^{x} f(u)du, \ -\infty < x < \infty$$

$$P(X \in (a,b]) \ = \ P(a < X \leq b) = \int_{a}^{b} f_X(u)du.$$

- ***pdf* properties**:

1. $f(x) \geq 0$ for all x.
2. $\int_{-\infty}^{\infty} f(x)dx = 1.$

$$F(t) = \int_{-\infty}^{t} f(x)dx$$

$$= \int_{0}^{t} f(x)dx \quad ,$$

Binomial

- Suppose that the probability of success is p

- What is the probability of failure?

  $q = 1 - p$

- Examples
  - Toss of a coin (S = head): $p = 0.5 \Rightarrow q = 0.5$
  - Roll of a die (S = 1): $p = 0.1667 \Rightarrow q = 0.8333$
  - Fertility of a chicken egg (S = fertile): $p = 0.8 \Rightarrow q = 0.2$

binomial

- Imagine that a trial is repeated *n* times

- Examples
  - A coin is tossed 5 times
  - A die is rolled 25 times
  - 50 chicken eggs are examined

- Assume *p* remains constant from trial to trial and that the trials are statistically independent of each other
- Example
  - What is the probability of obtaining 2 heads from a coin that was tossed 5 times?

$$P(HHTTT) = (1/2)^5 = 1/32$$

Poisson

- When there is a large number of trials, but a small probability of success, binomial calculation becomes impractical
  - Example: Number of deaths from horse kicks in the Army in different years

- The mean number of successes from n trials is $\mu = np$
  - Example: 64 deaths in 20 years from thousands of soldiers

If we substitute μ/n for p, and let n tend to infinity, the binomial distribution becomes the Poisson distribution:

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!}$$

poisson

- Poisson distribution is applied where random events in space or time are expected to occur
- Deviation from Poisson distribution may indicate some degree of non-randomness in the events under study
- Investigation of cause may be of interest

Exponential Distribution

The random variable X that equals the distance between successive counts of a Poisson process with mean $\lambda > 0$ is an **exponential random variable** with parameter λ. The probability density function of X is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \leq x < \infty \qquad (4\text{-}14)$$

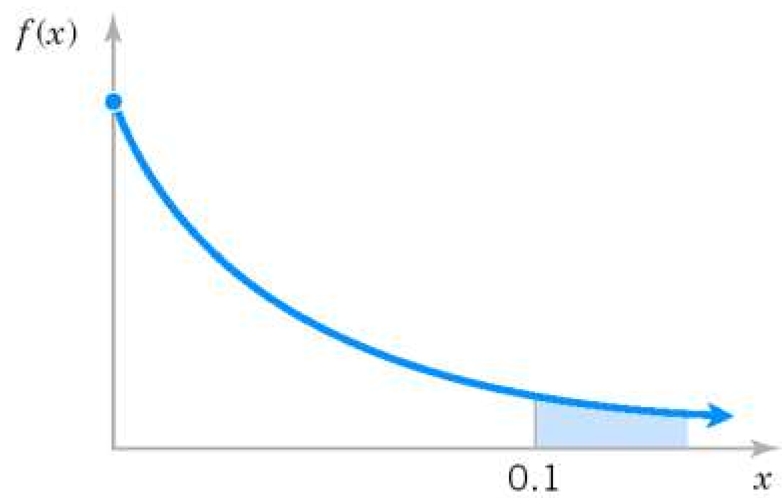
If the random variable X has an exponential distribution with parameter λ,

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2} \qquad (4\text{-}15)$$

In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in an interval of 6 minutes?

Let X denote the time in hours from the start of the interval until the first log-on. Then, X has an exponential distribution with $\lambda = 25$ log-ons per hour. We are interested in the probability that X exceeds 6 minutes. Because λ is given in log-ons per hour, we express all time units in hours. That is, 6 minutes $= 0.1$ hour. The probability requested is shown as the shaded area under the probability density function in Fig. 4-23. Therefore,

$$P(X > 0.1) = \int_{0.1}^{\infty} 25e^{-25x}\,dx = e^{-25(0.1)} = 0.082$$



Also, the cumulative distribution function can be used to obtain the same result as follows:

$$P(X > 0.1) = 1 - F(0.1) = e^{-25(0.1)}$$

An identical answer is obtained by expressing the mean number of log-ons as 0.417 log-ons per minute and computing the probability that the time until the next log-on exceeds 6 minutes. Try it.

What is the probability that the time until the next log-on is between 2 and 3 minutes? Upon converting all units to hours,

$$P(0.033 < X < 0.05) = \int_{0.033}^{0.05} 25e^{-25x}\,dx = -e^{-25x}\Big|_{0.033}^{0.05} = 0.152$$

An alternative solution is

$$P(0.033 < X < 0.05) = F(0.05) - F(0.033) = 0.152$$

Determine the interval of time such that the probability that no log-on occurs in the i val is 0.90. The question asks for the length of time x such that $P(X > x) = 0.90$. Now,

$$P(X > x) = e^{-25x} = 0.90$$

Take the (natural) log of both sides to obtain $-25x = \ln(0.90) = -0.1054$. Therefore,

$$x = 0.00421 \text{ hour} = 0.25 \text{ minute}$$

Furthermore, the mean time until the next log-on is

$$\mu = 1/25 = 0.04 \text{ hour} = 2.4 \text{ minutes}$$

The standard deviation of the time until the next log-on is

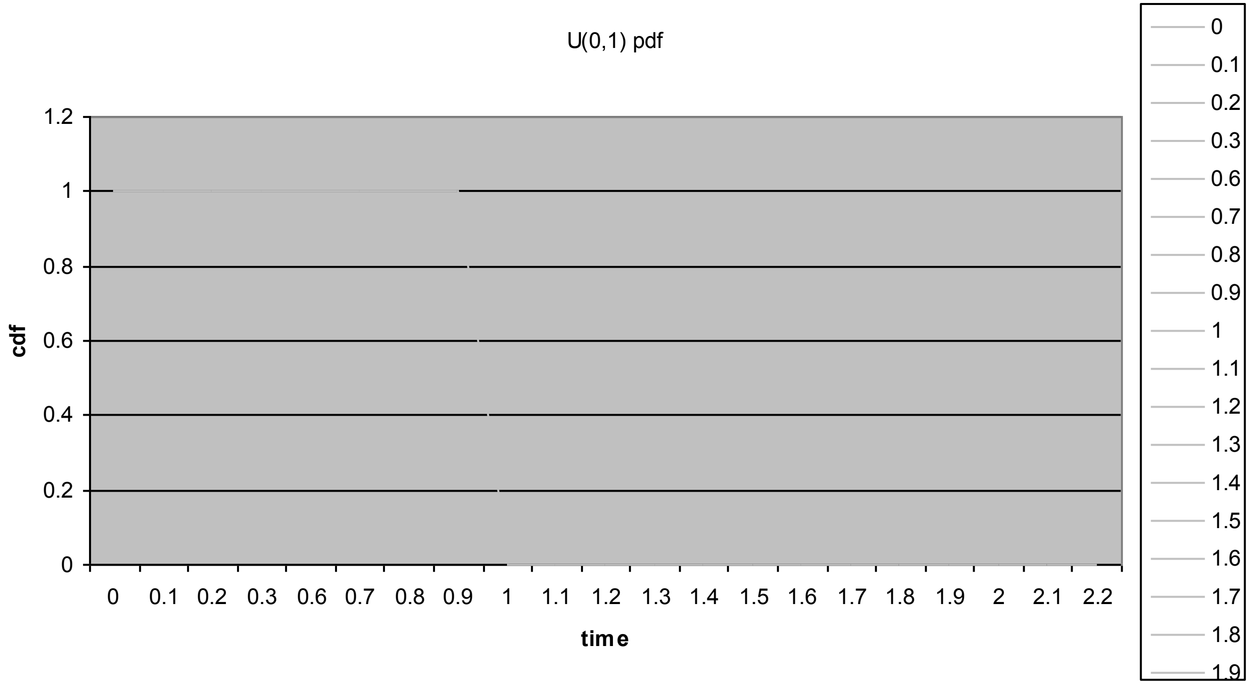
$$\sigma = 1/25 \text{ hours} = 2.4 \text{ minutes}$$

Uniform

All (pseudo) random generators generate random deviates of U(0,1) distribution; that is, if you generate a large number of random variables and plot their empirical distribution function, it will approach this distribution in the limit.

U(a,b) → pdf constant over the (a,b) interval and CDF is the ramp function

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$


U(0,1) pdf
cdf
1.2
1
0.8
0.6
0.4
0.2
0
0 0.1 0.2 0.3 0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2 2.1 2.2
time
0
0.1
0.2
0.3
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9

Uniform distribution

$$F(x)=\begin{cases} 0\ , & x < a, \\ \frac{x-a}{b-a}\ , & a \leq x < b \\ 1\ , & x \geq b. \end{cases}$$

Gaussian (Normal) Distribution

- Bell shaped pdf – intuitively pleasing!
- Central Limit Theorem: *mean of a large number of mutually independent rv's (having arbitrary distributions) starts following Normal distribution as n* →

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; -\infty < x < \infty$$

- μ: mean, σ: std. deviation, σ^2: variance ($N(\mu, \sigma^2)$)
- μ and σ completely describe the statistics. This is significant in statistical estimation/signal processing/communication theory etc.

- *N(0,1)* is called normalized Guassian.
- *N(0,1)* is symmetric i.e.
  - $f(x)=f(-x)$
  - $F(z) = 1-F(z).$
- Failure rate *h(t)* follows IFR behavior.
  - Hence, *N( )* is suitable for modeling long-term wear or aging related failure phenomena

Exponential Distribution

$$f(t) = \sum_{i=1}^{k} \alpha_i \lambda_i \mathrm{e}^{-\lambda_i t},\ t > 0,\ \lambda_i > 0,\ \alpha_i > 0,\ \sum_{i=1}^{k} \alpha_i$$

$$F(t) = \sum_i \alpha_i (1 - \mathrm{e}^{-\lambda_i t}),\ t \geq 0$$

$$h(t) = \frac{\sum_i \alpha_i \lambda \mathrm{e}^{-\lambda_i t}}{\sum_i \alpha_i \lambda \mathrm{e}^{-\lambda_i t}},\ t \geq 0$$

Conditional Distributions

- The conditional distribution of *Y given X=1* is:
- While marginal distributions are obtained from the bivariate by summing, conditional distributions are obtained by “making a cut” through the bivariate distribution

The Expectation of a Random Variable

Expectation of a discrete random variable with p.m.f

$$P(X = x_i) = p_i$$

$$E(X) = \sum_i p_i x_i$$

Expectation of a continuous random variable with p.d.f $f(x)$

$$E(X) = \int_{\text{state space}} x f(x) dx$$

$$\text{expectation of } X = \text{mean of } X = \text{average of } X$$

$$E[X] = \bar{X} = \int_{-\infty}^{\infty} x f_X(x) dx$$ continuous r.v.

$$E[X] = \bar{X} = \sum_{i=1}^{N} x_i P(x_i)$$ discrete r.v.

$$f_X(x+a) = f_X(-x+a), \forall x \quad \Rightarrow \quad E[X] = a$$

X r.v. $\Rightarrow$ Y=g(X) r.v. Ex: $Y = g(X) = X^2$

$$P(X=0) = P(X=-1) = P(X=1) = \frac{1}{3} \qquad P(Y=0) = \frac{1}{3} \qquad P(Y=1) = \frac{2}{3}$$

Expectation

expectation of a function of a r.v. X

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$ continuous r.v.

$$E[g(X)] = \sum_{i=1}^{N} g(x_i) P(x_i)$$ discrete r.v.

conditional expectation of a r.v. X

$$E[X|B] = \int_{-\infty}^{\infty} x f_X(x|B) dx$$ continuous r.v.

$$E[X|B] = \sum_{i=1}^{N} x_i P(x_i|B)$$ discrete r.v.

Ex: $B = \{X \le b\}$

$$f_X(x \mid X \le b) = \begin{cases} \dfrac{f_X(x)}{\int_{-\infty}^{b} f_X(x)dx}, & x < b \\ 0, & x \ge b \end{cases}$$

$$E[X \mid X \le b] = \frac{\int_{-\infty}^{b} x f_X(x)dx}{\int_{-\infty}^{b} f_X(x)dx}$$

Moments

n-th moment of a r.v. X

$$m_n = E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)dx$$ continuous r.v.

$$m_n = E[X^n] = \sum_{i=1}^{N} x_i^n P(x_i)$$ discrete r.v.

$$m_0 = 1$$

$$m_1 = \bar{X}$$

properties of expectation:

(1) $E[c] = c$ $\quad c$ -- constant

(2) $E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)]$

PF: $E[c] = \int_{-\infty}^{\infty} c f_X(x) dx = c \int_{-\infty}^{\infty} f_X(x) dx = c$

$$E[ag(X) + bh(X)] = \int_{-\infty}^{\infty} \{ag(x) + bh(x)\} f_X(x) dx$$

$$= a \int_{-\infty}^{\infty} g(x) f_X(x) dx + b \int_{-\infty}^{\infty} h(x) f_X(x) dx = aE[g(X)] + bE[h(X)]$$

variance of a r.v. X

$$\sigma_X^2 = \mu_2 = E[(X-\bar{X})^2] = E[X^2 - 2\bar{X}X + \bar{X}^2]$$

$$= E[X^2] - 2\bar{X}E[X] + \bar{X}^2 = m_2 - m_1^2$$

standard deviation of a r.v. $X = \sigma_X (\geq 0)$

skewness of a r.v. $X = \dfrac{\mu_3}{\sigma_X^3}$

Ex 3.2-1 & Ex3.2-2: $f_X(x)$ symmetric about $x = \bar{X} \quad \Rightarrow \quad \mu_3 = 0$

exponential r.v.

$$f_X(x) = \begin{cases} \dfrac{1}{b} e^{-\frac{x-a}{b}}, & x > a \\ 0, & \mathrm{x} < a \end{cases}$$

$$m_1 = E[X] = \int_a^\infty x \frac{1}{b} e^{-\frac{x-a}{b}} dx = a + b$$

$$m_2 = E[X^2] = \int_a^\infty x^2 \frac{1}{b} e^{-\frac{x-a}{b}} dx = (a+b)^2 + b^2$$

$$\sigma_X^2 = \mu_2 = m_2 - m_1^2 = b^2$$

$$m_3 = E[X^3] = \int_a^\infty x^3 \frac{1}{b} e^{-\frac{x-a}{b}} dx = a^3 + 3a^2b + 6ab^2 + 6b^3$$

$$\mu_3 = E[(X - \bar{X})^3] = E[X^3 - 3X^2\bar{X} + 3X\bar{X}^2 - \bar{X}^3] = m_3 - 3m_1m_2 + 3m_1^2m_1 - m_1^3$$

$$= a^3 + 3a^2b + 6ab^2 + 6b^3 - 3(a+b)\{(a+b)^2 + b^2\} + 2(a+b)^3 = 2b^3$$

skewness of a r.v. $X = \dfrac{\mu_3}{\sigma_X^3} = \dfrac{2b^3}{b^3} = 2$

Chebychev's inequality $P[|X - \bar{X}| \geq \varepsilon] \leq \frac{\sigma_X^2}{\varepsilon^2}$

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

$$\geq \varepsilon^2 \int_{|x-\bar{X}|\geq\varepsilon} f_X(x)dx = \varepsilon^2 P[|X - \bar{X}| \geq \varepsilon]$$

Markov's inequality

$$P[X < 0] = 0 \quad \Rightarrow \quad P[X \geq a] \leq \frac{E[X]}{a}$$

Ex 3.2-3: $P[|X - \bar{X}| \geq 3\sigma_X] \leq \frac{\sigma_X^2}{9\sigma_X^2} = \frac{1}{9}$

Characteristic function of r.v. X

$$\Phi_X(\omega) = E[e^{j\omega X}] = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx$$

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

Fourier transform

$$|\Phi_X(\omega)| \le \int_{-\infty}^{\infty} |f_X(x)| |e^{j\omega x}| dx \le \int_{-\infty}^{\infty} f_X(x) dx = 1 = \Phi_X(0)$$

$$\left. \frac{d^n \Phi_X(\omega)}{d\omega^n} \right|_{\omega=0} = \left. \int_{-\infty}^{\infty} f_X(x) j^n x^n e^{j\omega x} dx \right|_{\omega=0} = j^n \int_{-\infty}^{\infty} f_X(x) x^n dx = j^n E[X^n]$$

$$m_n = (-j)^n \left. \frac{d^n \Phi_X(\omega)}{d\omega^n} \right|_{\omega=0}$$

Functions That Give Moments

Moment generating function of r.v. X

$$M_X(v) = E[e^{vX}] = \int_{-\infty}^{\infty} f_X(x) e^{vx} dx$$

$$\left. \frac{d^n M_X(v)}{dv^n} \right|_{v=0} = \int_{-\infty}^{\infty} f_X(x) x^n e^{vx} dx \Big|_{v=0} = \int_{-\infty}^{\infty} f_X(x) x^n dx = m_n$$

Ex 3.3-1 & Ex 3.3-2:

$$f_X(x) = \begin{cases} \frac{1}{b} e^{-\frac{x-a}{b}}, & x > a \\ 0, & \mathrm{x} < a \end{cases}$$

$$\Phi_X(\omega) = E[e^{j\omega X}] = \frac{1}{b} e^{\frac{a}{b}} \int_a^\infty e^{-(\frac{1}{b} - j\omega)x} dx = \frac{1}{b} e^{\frac{a}{b}} \left. \frac{e^{-(\frac{1}{b} - j\omega)x}}{-(\frac{1}{b} - j\omega)} \right|_{x=a}^{\infty}$$

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

$$\frac{d\Phi_X(\omega)}{d\omega} = \frac{jae^{j\omega a}(1 - j\omega b) + e^{j\omega a} jb}{(1 - j\omega b)^2}$$

$$M_X(v) = E[e^{vX}] = \frac{e^{va}}{1 - vb}$$

$$\frac{dM_X(v)}{dv} = \frac{ae^{va}(1 - vb) + e^{va} b}{(1 - vb)^2}$$

$$m_1 = (-j) \left. \frac{d\Phi_X(\omega)}{d\omega} \right|_{\omega=0} = a + b$$

$$m_1 = \left. \frac{dM_X(v)}{dv} \right|_{v=0} = a + b$$

Chernoff's inequality Ex 3.3-3:

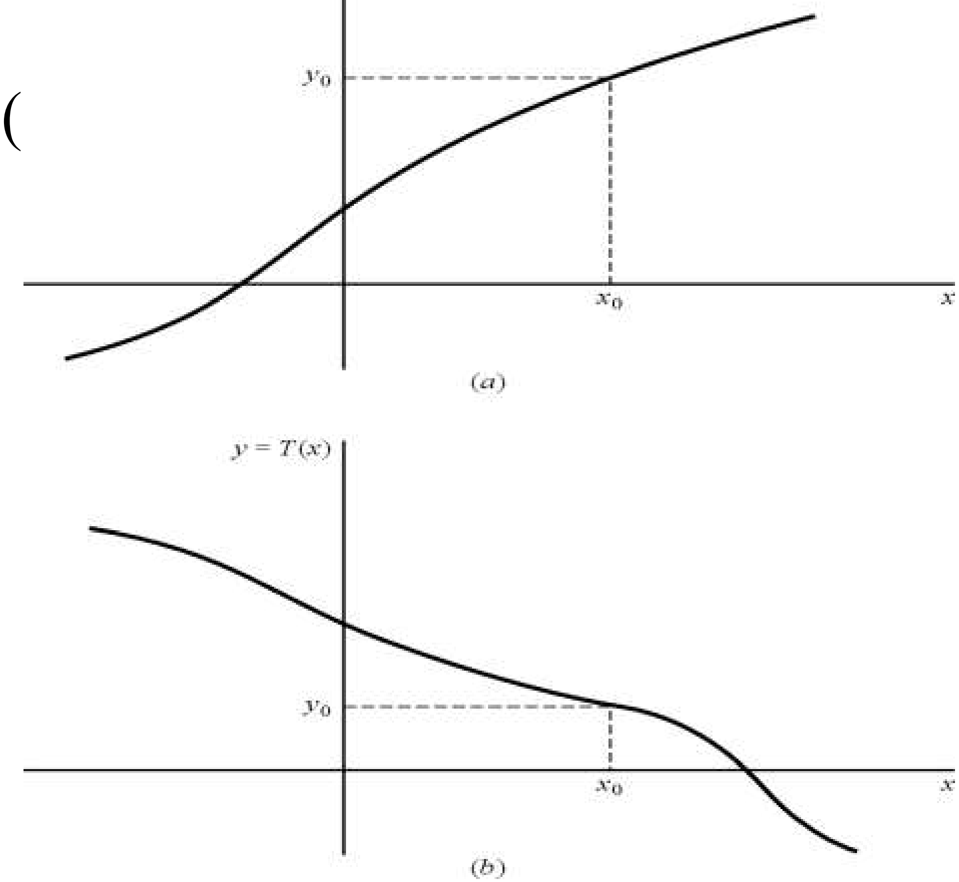
$v > 0$

$$P[X \geq a] = \int_a^\infty f_X(x)dx = \int_{-\infty}^\infty f_X(x)u(x-a)dx$$

$$\leq \int_{-\infty}^\infty f_X(x)e^{v(x-a)}dx = e^{-va}M_X(v)$$

Transformations of a Random Variable

$Y = T(X)$ $f_X($



monotone increasing $\Leftrightarrow$

$T(x_1) < T(x_2)$ for any $x_1 < x_2$

monotone decreasing $\Leftrightarrow$

$T(x_1) > T(x_2)$ for any $x_1 < x_2$

Assume monotone increasing $T(\bullet)$ $\quad Y = T(X)$

$$F_Y(y_0) = P[Y \le y_0] = P[X \le x_0] = F_X(x_0)$$

$$\int_{-\infty}^{y_0} f_Y(y)dy = \int_{-\infty}^{T^{-1}(y_0)} f_X(x)dx$$

$$f_Y(y_0) = f_X[T^{-1}(y_0)]\frac{dT^{-1}(y_0)}{dy_0}$$

$$f_Y(y) = f_X[T^{-1}(y)]\frac{dT^{-1}(y)}{dy} = f_X(x)\frac{dx}{dy}$$

Assume monotone decreasing $T(\bullet)$ $\quad Y = T(X)$

$$F_Y(y_0) = P[Y \le y_0] = P[X \ge x_0] = 1 - F_X(x_0)$$

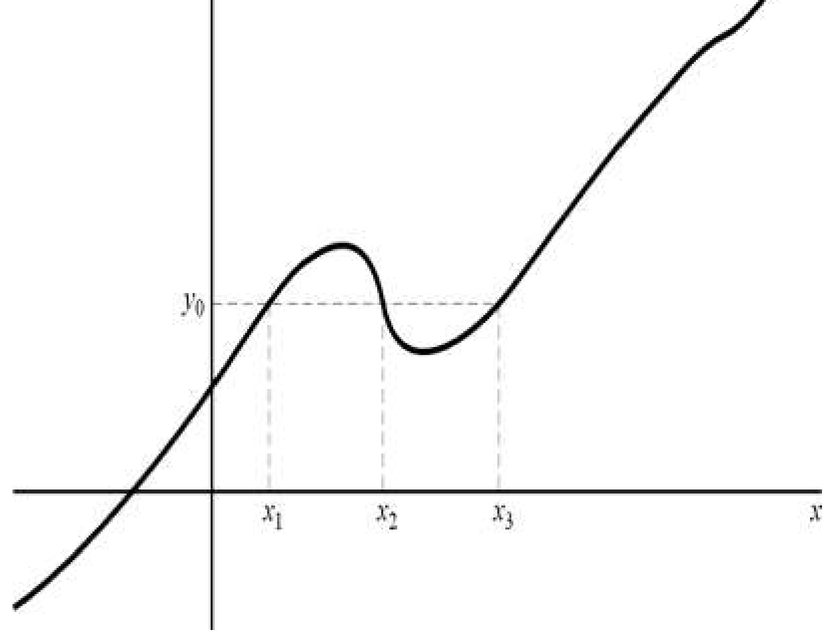
$$f_Y(y) = -f_X(x)\frac{dx}{dy}$$

monotone $T(\bullet)$ $\Rightarrow$ $$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X(x)\frac{1}{\left|\frac{dy}{dx}\right|}$$

nonmonotone $T(\bullet)$

$$Y = T(X)$$

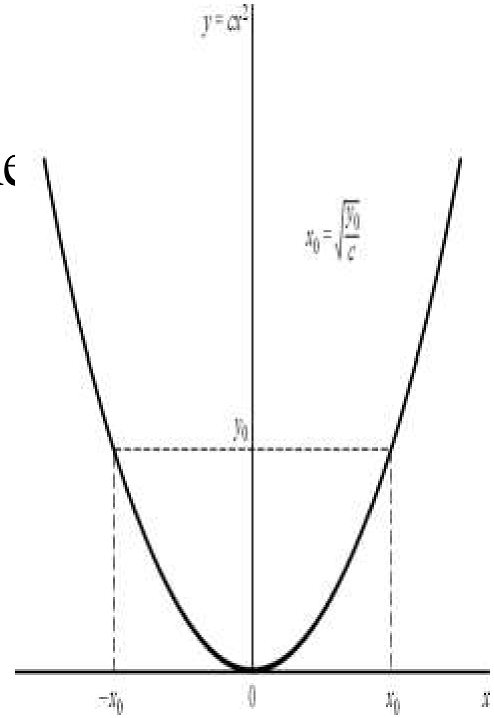
$$f_Y(y) = \sum_n \frac{f_X(x_n)}{\left|\left.\frac{dT(x)}{dx}\right|_{x=x_n}\right|}$$



Ex 3.4-2:

$$Y = T(X) = cX^2$$

nonmonotone



$$f_Y(y) = f_X(\sqrt{y/c})\left|\frac{d\sqrt{y/c}}{dy}\right|$$

$$+ f_X(-\sqrt{y/c})\left|\frac{-d\sqrt{y/c}}{dy}\right|$$

$$= \frac{f_X(\sqrt{y/c}) + f_X(-\sqrt{y/c})}{2\sqrt{cy}}, \qquad y \geq 0$$

MULTIPLE RANDOM VARIABLES and OPERATIONS:

MULTIPLE RANDOM VARIABLES :

Vector Random Variables

A vector random variable X is a function that assigns a vector of real numbers to each outcome ζ in S, the sample space of the random experiment

Events and Probabilities

EXAMPLE 4.4

Consider the tow-dimensional random variable **X** = (X, Y). Find the region of the plane corresponding to the events

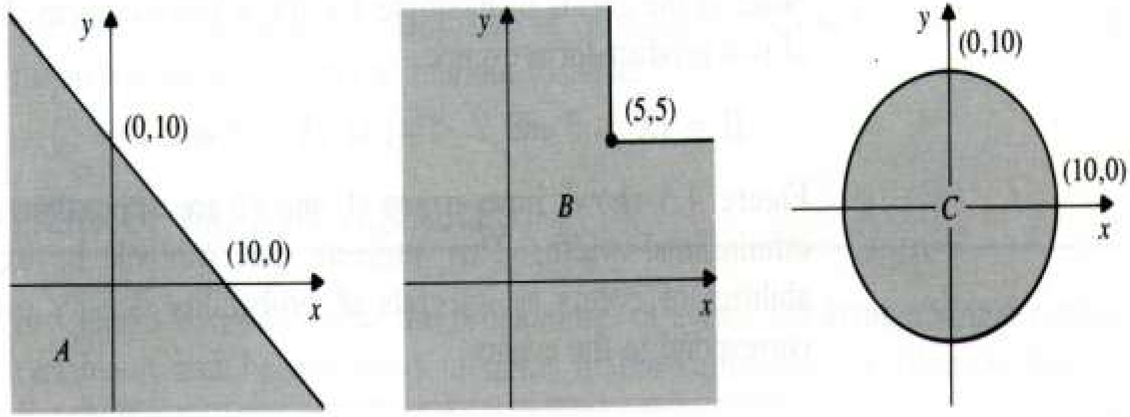
$$A = \{X + Y \leq 10\},$$

$$B = \{\min(X, Y) \leq 5\}, \text{ and}$$

$$C = \{X^2 + Y^2 \leq 100\}.$$

The regions corresponding to events A and C are straightforward to find and are shown in Fig. 4.1.

FIGURE 4.1 Examples of two-dimensional events.



Independence

If the one-dimensional random variable X and Y are "independent," if A_1 is any event that involves X only and A_2 is any event that involves Y only, then

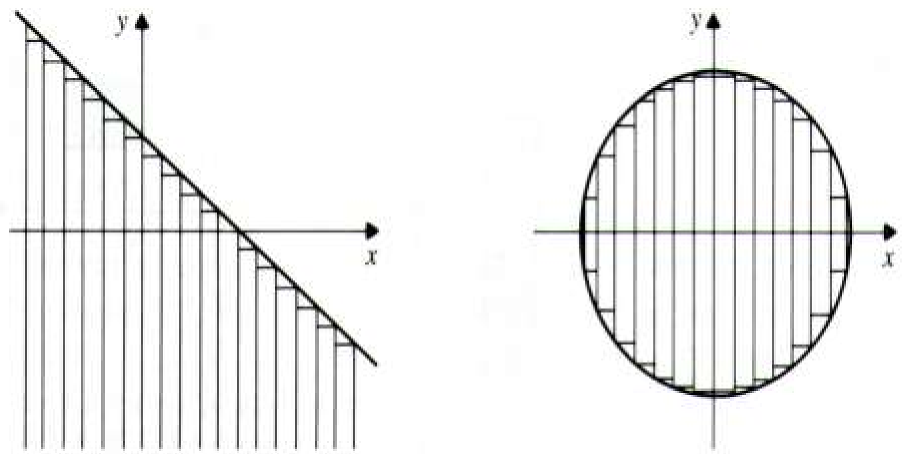
$$P[X \text{ in } A_1,\ Y \text{ in } A_2] = P[X \text{ in } A_1]P[Y \text{ in } A_2].$$

In the general case of n random variables, we say that the random variables $X_1, X_2, \ldots, X_n$ are independent if

$$P[X_1 \text{ in } A_1, \ldots, X_n \text{ in } A_n] = P[X_1 \text{ in } A_1] \cdots P[X_n \text{ in } A_n], \tag{4.3}$$

where the A_k is an event that involves X_k only.

FIGURE 4.3
Some two-dimensional non-product-form events.



Pairs of Discrete Random Variable

Let the vector random variable **X** = (X,Y) assume values from some countable set $S=\{(x_j, y_k),\ \ j=1,2,\ldots,\ k=1,2,\ldots\}$. The joint probability mass function of **X** specifies the probabilities of the product-form event

$$\{X=x_j\}\cap\{Y=y_k\}:$$

$$\begin{aligned} p_{X,Y}(x_j,y_k) &= P[\{X=x_j\}\cap\{Y=y_k\}] \\ &\equiv P[X=x_j, Y=y_k] \qquad j=1,2,\ldots \quad k=1,2,\ldots \end{aligned} \qquad (4.4)$$

The probability of any event A is the sum of the pmf over the outcomes in A

$$P[X \text{ in } A]=\sum_{(x_j,y_k)}\sum_{in\ A} p_{X,Y}(x_j,y_k). \qquad (4.5)$$

$$\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k) = 1. \tag{4.6}$$

The marginal probability mass functions :

$$\begin{aligned} p_X(x_j) &= P\left[X = x_j\right] \\ &= P\left[X = x_j, Y = \text{anything}\right] \\ &= P\left[\left\{X = x_j \;\; and \;\; Y = y_1\right\} \cup \left\{X = x_j \;\; and \;\; Y = y_2\right\} \cup \cdots\right] \\ &= \sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k), \end{aligned} \tag{4.7a}$$

$$\begin{aligned} p_Y(y_k) &= P\left[Y = y_k\right] \\ &= \sum_{j=1}^{\infty} p_{X,Y}(x_j, y_k). \end{aligned} \tag{4.7b}$$

The Joint cdf of X and Y

The joint cumulative distribution function of X and Y is defined as the probability of the product-form event $\{X \le x_1\} \cap \{Y \le y_1\}$":

$$F_{X,Y}(x_1, y_1) = P[X \le x_1, Y \le y_1]. \tag{4.8}$$

The joint cdf is nondecreasing in the “northeast” direction,

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

It is impossible for either X or Y to assume a value less than $-\infty$, therefore

(ii) $F_{X,Y}(-\infty,y_1) = F_{X,Y}(x_2,-\infty) = 0$

It is certain that X and Y will assume values less than infinity, therefore

(iii) $F_{X,Y}(\infty,\infty) = 1$.

If we let one of the variables approach infinity while keeping the other fixed, we obtain the marginal cumulative distribution functions

(iv) $$F_X(x) = F_{X,Y}(x,\infty) = P[X \le x, Y \le \infty] = P[X \le x]$$

and

$$F_Y(y) = F_{X,Y}(\infty, y) = P[Y \le y].$$

Recall that the cdf for a single random variable is continuous form the right. It can be shown that the joint cdf is continuous from the "north" and from the "east"

(v) $$\lim_{x \to a^+} F_{X,Y}(x,y) = F_{X,Y}(a,y)$$

and

$$\lim_{y \to b^+} F_{X,Y}(x,y) = F_{X,Y}(x,b)$$

FIGURE 4.4
The joint cumulative distribution function is defined as the probability of the semi-infinite rectangle defined by the point (x_1, y_1).

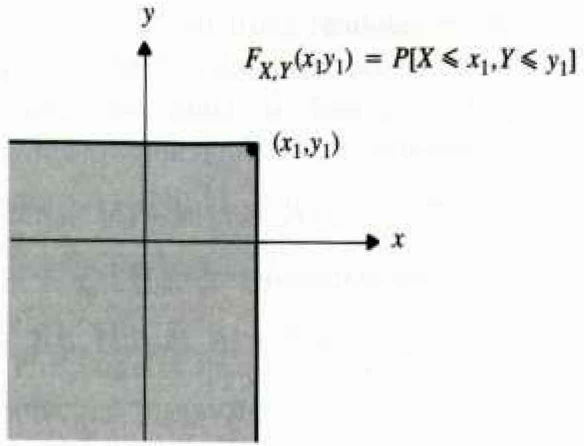
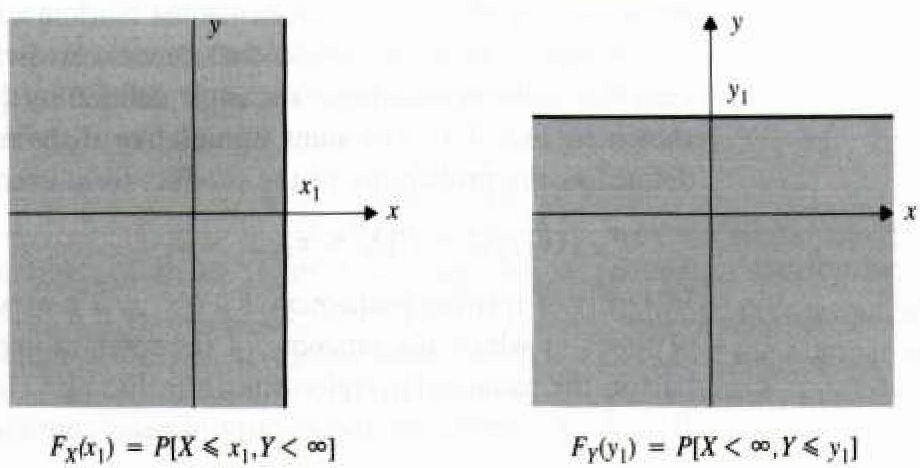


FIGURE 4.5
The marginal cdf's are the probabilities of these half-planes.



The Joint pdf of Two Jointly Continuous Random Variables

We say that the random variables X and Y are jointly continuous if the probabilities of events involving (X, Y) can be expressed as an integral of a pdf. There is a nonnegative function $f_{X,Y}(x,y)$, called the joint probability density function, that is defined on the real plane such that for every event A, a subset of the plane,

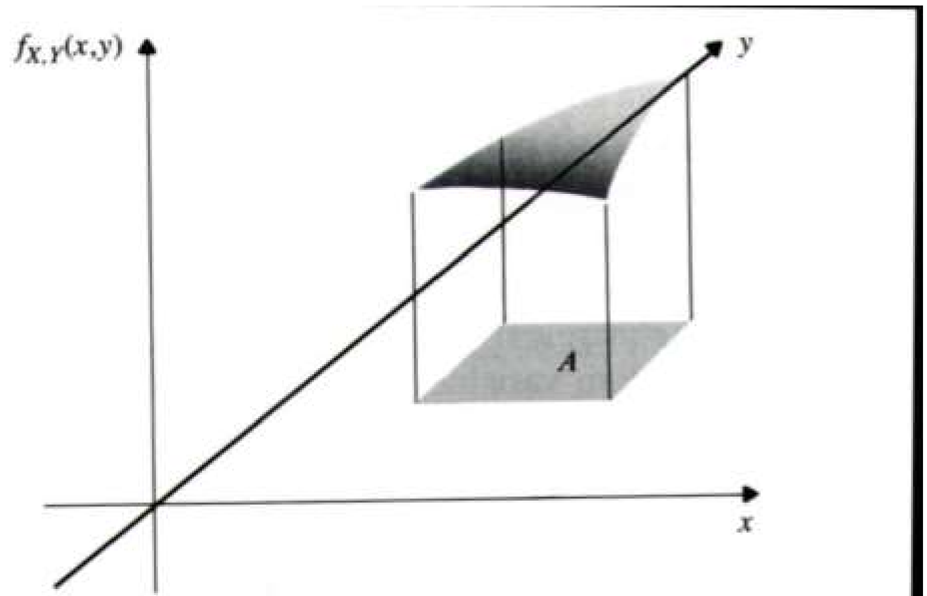
$$P[\mathbf{X} \text{ in } A] = \int_A \int f_{X,Y}(x', y')\,dx'\,dy', \qquad (4.9)$$

as shown in Fig. 4.7. When a is the entire plane, the integral must equal one :

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x', y')\,dx'\,dy'. \qquad (4.10)$$

The joint cdf can be obtained in terms of the joint pdf of jointly continuous random variables by integrating over the semi-infinite

FIGURE 4.7
The probability of A is the integral of $f_{X,Y}(x, y)$ over the region defined by A.



The marginal pdf's $f_X(x)$ and $f_Y(y)$ are obtained by taking the derivative of the corresponding marginal cdf's

$$F_X(x) = F_{X,Y}(x,\infty)$$

$$F_Y(y) = F_{X,Y}(\infty, y).$$

$$F_X(x) = \frac{d}{dx}\int_{-\infty}^{x}\left\{\int_{-\infty}^{\infty} f_{X,Y}(x',y')dy'\right\}dx'$$

$$= \int_{-\infty}^{\infty} f_{X,Y}(x,y')dy'. \qquad (4.15a)$$

$$F_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x',y)dx'. \qquad (4.15b)$$

INDEPENDENCE OF TWO RANDOM VARIABLES

X and Y are independent random variables if any event A_1 defined in terms of X is independent of any event A_2 defined in terms of Y;

$$P[X \text{ in } A_1, Y \text{ in } A_2] = P[X \text{ in } A_1]P[Y \text{ in } A_2]. \qquad (4,17)$$

Suppose that X and Y are a pair of discrete random variables. If we let $A_1 = \{X = x_j\}$ and $A_2 = \{Y = y_k\}$ then the independence of X and Y implies that

$$\begin{aligned} p_{X,Y}(x_j, y_k) &= P[X = x_j, Y = y_k] \\ &= P[X = x_j]P[Y = y_k] \\ &= p_X(x_j)p_Y(y_k) \quad \text{for all } x_j \text{ and } y_k. \end{aligned} \qquad (4.18)$$

4.4 CONDITIONAL PROBABILITY AND CONDITIONAL EXPECTATION

Conditional Probability

In Section 2.4, we know

$$P[Y \text{ in } A \mid X = x] = \frac{P[Y \text{ in } A, X = x]}{P[X = x]}. \qquad (4.22)$$

If X is discrete, then Eq. (4.22) can be used to obtain the conditional cdf of Y given $X = x_k$:

$$F_Y(y \mid x_k) = \frac{P[Y \leq y, X = x_k]}{P[X = x_k]}, \quad \text{for } P[X = x_k] > 0. \qquad (4.23)$$

The conditional pdf of Y given $X = x_k$, if the derivative exists, is given by

$$f_Y(y \mid x_k) = \frac{d}{dy} F_Y(y \mid x_k). \qquad (4.24)$$

MULTIPLE RANDOM VARIABLES

Joint Distributions

The joint cumulative distribution function of X_1, X_2,…., X_n is defined as the probability of an n-dimensional semi-infinite rectangle associate with the point (x_1,…, x_n):

$$F_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots x_n) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n]. \quad (4.38)$$

The joint cdf is defined for discrete, continuous, and random variables of mixed type

FUNCTIONS OF SEVERAL RANDOM VARIABLES

One Function of Several Random Variables

Let the random variable Z be defined as a function of several random variables:

$$Z = g(X_1, X_2, \ldots, X_n). \tag{4.51}$$

The cdf of Z is found by first finding the equivalent event of that is, the set $R_Z = \{\mathrm{x} = (x_1, \ldots, x_n)$ such that $g(\mathrm{x}) \le z\}$, then

$$F_Z(z) = P[\mathrm{X} \text{ in } R_z]$$
$$= \int \underset{\mathrm{x}\ \text{in}\ R_z}{\cdots} \int f_{X_1,\ldots,X_n}(x_1', \ldots, x_n')\, dx_1' \ldots dx_n'. \tag{4.52}$$

EXAMPLE 4.31 Sum of Two Random Variables

Let $Z = X + Y$. Find $F_Z(z)$ and $f_Z(z)$ in terms of the joint pdf of X and Y.

The cdf of Z is

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x'} f_{X,Y}(x', y') dy' dx'.$$

The pdf of Z is

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x', z - x') dx'. \quad (4.53)$$

Thus the pdf for the sum of two random variables is given by a superposition integral. If X and Y are independent random variables, then by Eq. (4.21) the pdf is given by the convolution integral of the margial pdf's of X and Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x') f_Y(z - x') dx'. \quad (4.54)$$

pdf of Linear Transformations

We consider first the linear transformation of two random variables

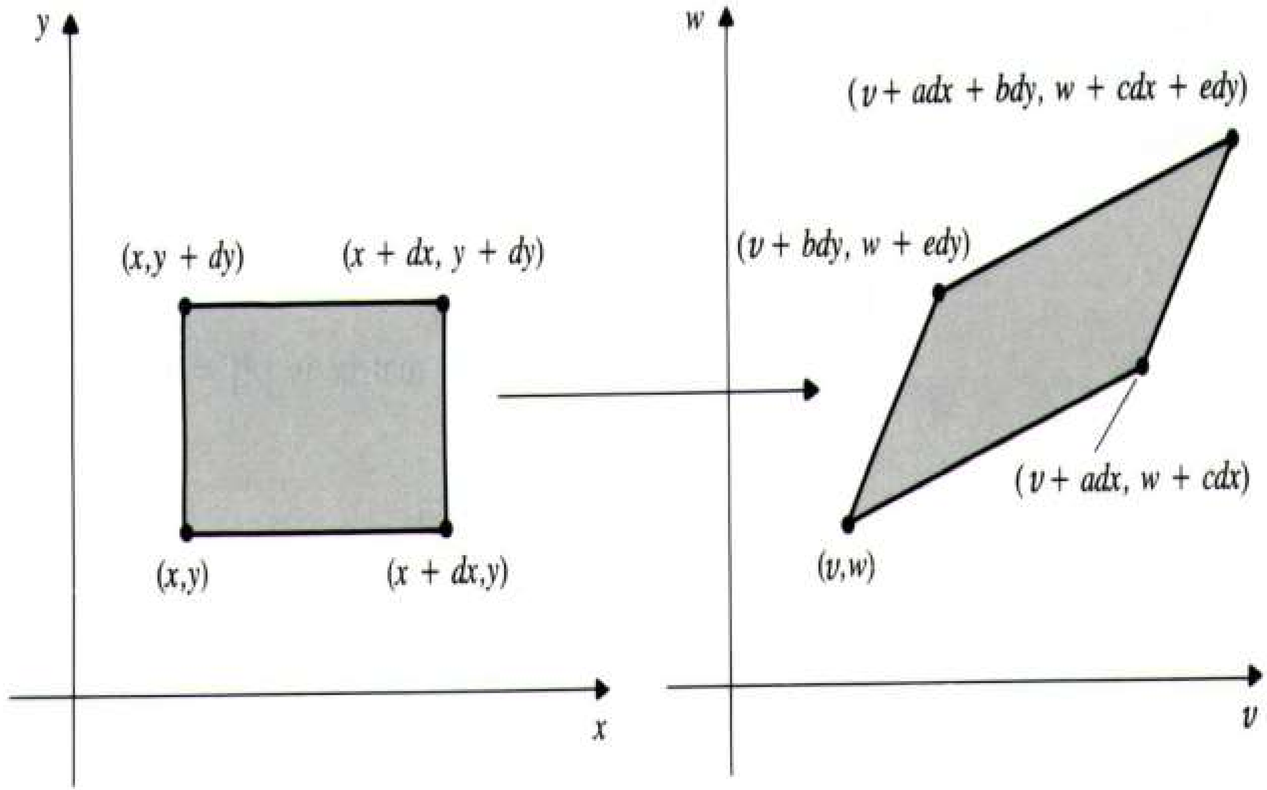
$$V = aX + bY \qquad \begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & e \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$
$$W = cX + eY$$

Denote the above matrix by A. We will assume A has an inverse, so each point (v, w) has a unique corresponding point (x, y) obtained from

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} v \\ w \end{bmatrix}. \tag{4.56}$$

In Fig. 4.15, the infinitesimal rectangle and the parallelogram are equivalent events, so their probabilities must be equal. Thus

$$f_{X,Y}(x, y)dxdy \cong f_{V,W}(v, w)dP$$



$$v = ax + by$$
$$w = cx + ey$$

FIGURE 4.15
Image of an infinitesimal rectangle under a linear transformation.

where dP is the area of the parallelogram. The joint pdf of V and W is thus given by

$$f_{V,W}(v,w) = \frac{f_{X,Y}(x,y)}{\left|\frac{dP}{dxdy}\right|}, \qquad (4.57)$$

where x an y are related to (v, w) by Eq. (4.56) It can be shown that $dP = (|ae - bc|)dxdy$, so the "stretch factor" is

$$\left|\frac{dP}{dxdy}\right| = \frac{|ae-bc|(dxdy)}{(dxdy)} = |ae-bc| = |A|,$$

where $|A|$ is the determinant of A.

Let the n-dimensional vector **Z** be

$$\mathbf{Z} = A\mathbf{X},$$

where A is an $n \times n$ invertible matrix. The joint of **Z** is then

EXPECTED VALUE OF FUNCTIONS OF RANDOM VARIABLES

The expected value of $Z = g(X, Y)$ can be found using the following expressions

$$E[Z] = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) & X, Y \text{ jointly continuous} \\ \sum_{i}\sum_{n} g(x_i, y_n) p_{X,Y}(x_i, y_n) & X, Y \text{ discrete.} \end{cases} \tag{4.64}$$

*Joint Characteristic Function

The joint characteristic function of n random variables is defined as

$$\Phi_{X_1,X_2,\ldots X_n}(w_1,w_2,\ldots w_n)=E\left[e^{j\left(w_1X_1+w_2X_2+\cdots+w_nX_n\right)}\right]. \tag{4.73a}$$

$$\Phi_{X,Y}(w_1,w_2)=E\left[e^{j\left(w_1X+w_2Y\right)}\right]. \tag{4.73b}$$

If X and Y are jointly continuous random variables, then

$$\Phi_{X,Y}(w_1,w_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(x,y)e^{j\left(w_1x+w_2y\right)}dxdy. \tag{4.73c}$$

The inversion formula for the Fourier transform implies that the joint pdf is given by

$$f_{X,Y}(x,y)=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi_{X,Y}(w_1,w_2)e^{j\left(w_1x+w_2y\right)}dw_1dw_2\,. \tag{4.74}$$

JOINTLY GAUSSIAN RANDOM VARIABLES

The random variables X and Y are said to be jointly Gaussian if their joint pdf has the form

$$f_{X,Y}(x,y) = \frac{\exp\left\{\frac{-1}{2\left(1-\rho_{X,Y}^2\right)}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right) + \left(\frac{y-m_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}} \qquad (4.79)$$

$-\infty < x < \infty$ and $-\infty < y < \infty$

The pdf is constant for values x and y for which the argument of the exponent is constant

$$\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right) + \left(\frac{y-m_2}{\sigma_2}\right)^2\right] = \text{constant}$$

When $\rho_{X,Y} = 0$, X and Y are independent ; when $\rho_{X,Y} \neq 0$, the major axis of the ellipse is oriented along the angle

$$\theta = \frac{1}{2}\arctan\left(\frac{2\rho_{X,Y}\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}\right). \qquad (4.80)$$

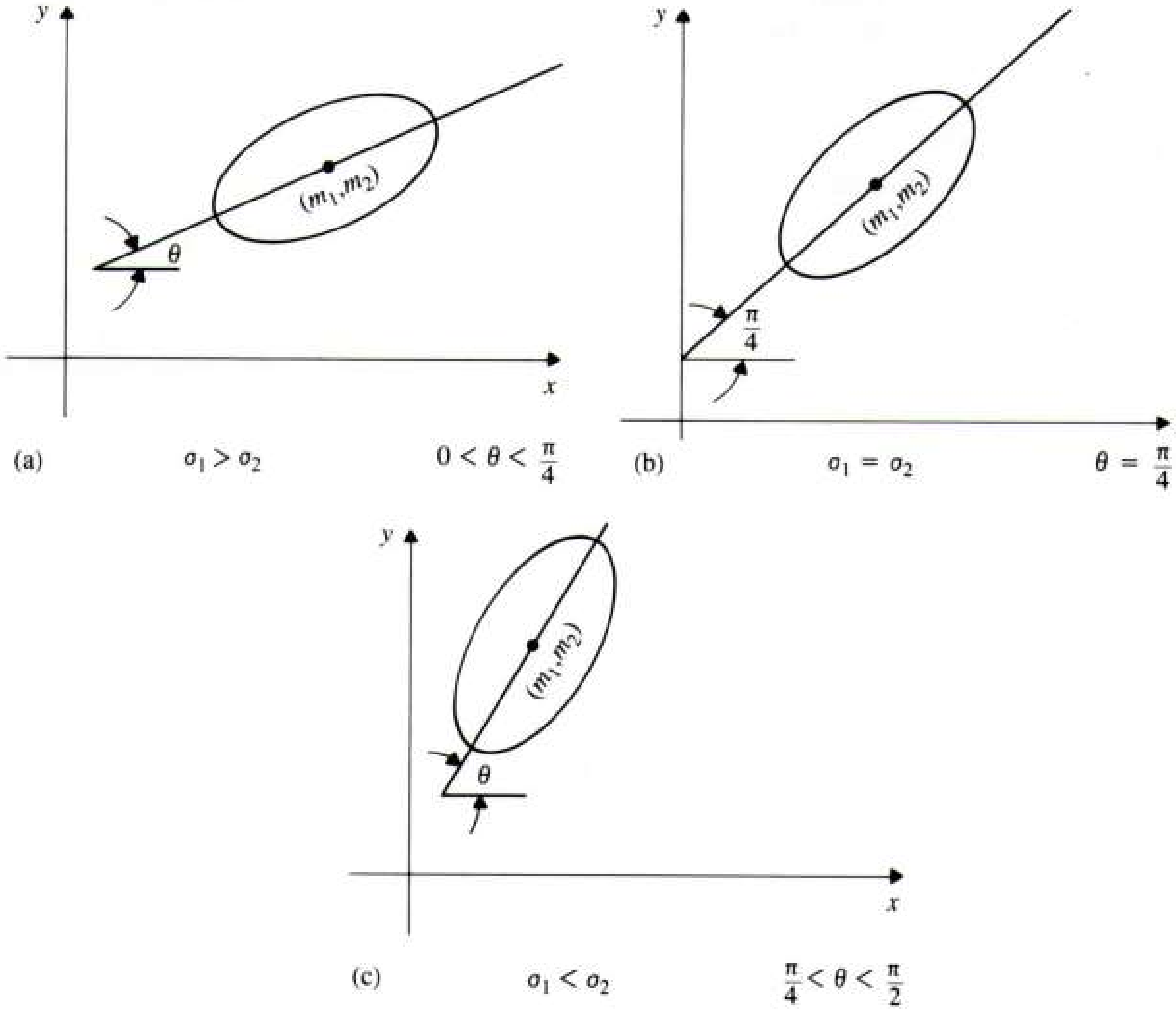
Note that the angle is 45º when the variance are equal.
The marginal pdf of X is found by integrating $f_{X,Y}(x, y)$ over all y

$$f_X(x) = \frac{e^{-(x-m_1)^2/2\sigma_1^2}}{\sqrt{2\pi}\sigma_1}, \qquad (4.81)$$

that is, X is a Gaussian random variable with mean m_1 and variance

$$\sigma_1^2$$

FIGURE 4.19
Orientation of contours of equal value of joint Gaussian pdf for $\rho_{X,Y} > 0$.



n Jointly Gaussian Random Variables

The random variables $X_1, X_2, \ldots, X_n$ are said to be jointly Gaussian if their joint pdf is given by

$$f_{\mathbf{X}}(\mathbf{x}) \equiv f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots x_n) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T K^{-1}(\mathbf{x}-\mathbf{m})\right\}}{(2\pi)^{n/2}|k|^{1/2}}, \quad \textbf{(4.83)}$$

where **x** and **m** are column vectors defined by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \\ E[X_3] \\ E[X_4] \end{bmatrix}$$

and K is the covariance matrix that is defined by

$$K = \begin{bmatrix} \text{VAR}(X_1) & \text{COV}(X_2, X_1) & \cdots & \text{COV}(X_1, X_n) \\ \text{COV}(X_2, X_1) & \text{VAR}(X_2) & \cdots & \text{COV}(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ \text{COV}(X_n, X_1) & \cdots & & \text{VAR}(X_n) \end{bmatrix} \quad (4.84)$$

Transformations of Random Vectors

Let $X_1,\ldots, X_n$ be random variables associate with some experiment, and let the random variables $Z_1,\ldots, Z_n$ be defined by n functions of **X** = $(X_1,\ldots, X_n)$:

$$Z_1 = g_1(\mathrm{X}) \qquad Z_2 = g_2(\mathrm{X}) \quad \ldots \quad Z_n = g_n(\mathrm{X}).$$

The joint cdf of $Z_1,\ldots, Z_n$ at the point **z** = $(z1,\ldots, zn)$ is equal to the probability of the region of **x** where

$$F_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) = P\left[g_1(\mathrm{X}) \le z_1,\ldots,g_n(\mathrm{X}) \le z_n\right]. \qquad (4.55a)$$

$$F_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) = \int\cdots\int_{\mathrm{x}':g_k(\mathrm{x}')\le z_k} f_{X_1,\ldots,X_n}(x_1',\ldots,x_n')\,dx_1'\cdots dx_n'\,. \qquad (4.55b)$$

pdf of Linear Transformations

We consider first the linear transformation of two random variables

$$V = aX + bY$$
$$W = cX + eY$$

$$\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & e \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

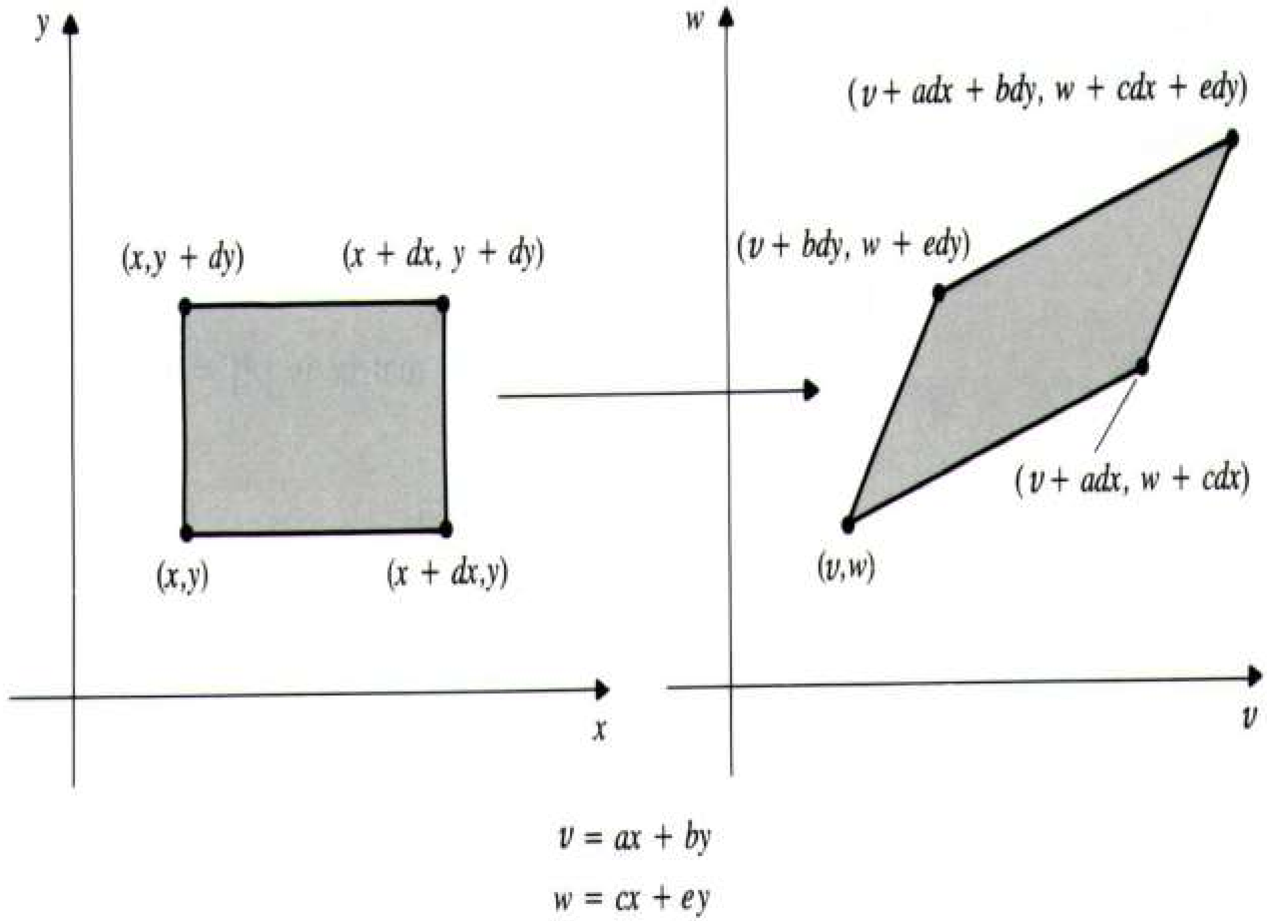
Denote the above matrix by A. We will assume A has an inverse, so each point (v, w) has a unique corresponding point (x, y) obtained from

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} v \\ w \end{bmatrix}. \qquad (4.56)$$

In Fig. 4.15, the infinitesimal rectangle and the parallelogram are equivalent events, so their probabilities must be equal. Thus

$$f_{X,Y}(x, y)dxdy \cong f_{V,W}(v, w)dP$$

FIGURE 4.15
Image of an infinitesimal rectangle under a linear transformation.



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 continuous or discrete (countably infinite or finite) as well.

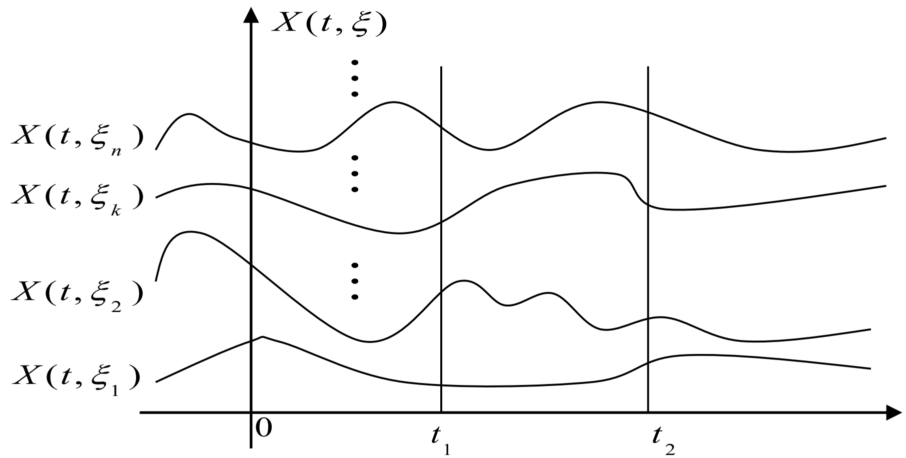


Fig. 14.1

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

$$X_1 = X(t_1,\xi_i)$$

is a random variable. The ensemble of all such realizations $X(t,\xi)$ over time represents the stochastic process $X(t)$. (see Fig 14.1). For example

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

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

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

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. Their joint distribution is given by

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

and

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

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:

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

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

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

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:

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

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

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

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j^* R_{XX}(t_i,t_j) \ge 0. \tag{14-8}$$

Eq. (14-8) follows by noticing that $E\{|Y|^2\} \ge 0 \;\text{ for }\; Y = \sum_{i=1}^n a_i X(t_i).$

The function

$$C_{XX}(t_1,t_2) = R_{XX}(t_1,t_2) - \mu_X(t_1)\mu_X^*(t_2) \tag{14-9}$$

represents the **autocovariance** function of the process *X*(*t*).

Example 14.1

Let

$$z = \int_{-T}^{T} X(t)dt.$$

Then

$$E[|z|^2] = \int_{-T}^{T}\int_{-T}^{T} E\{X(t_1)X^*(t_2)\}dt_1dt_2$$

$$= \int_{-T}^{T}\int_{-T}^{T} R_{XX}(t_1,t_2)dt_1dt_2 \tag{14-10}$$

Stationary Stochastic Processes

Stationary processes exhibit statistical properties that are invariant to shift in the time index. Thus, for example, second-order stationarity implies that the statistical properties of the pairs $\{X(t_1), X(t_2)\}$ and $\{X(t_1+c), X(t_2+c)\}$ are the same for *any* c. Similarly first-order stationarity implies that the statistical properties of $X(t_i)$ and $X(t_i+c)$ are the same for any c.

In strict terms, the statistical properties are governed by the joint probability density function. Hence a process is n[th]-order **Strict-Sense Stationary (S.S.S)** if

$$f_X(x_1, x_2, \cdots x_n,\; t_1, t_2 \cdots, t_n) \equiv f_X(x_1, x_2, \cdots x_n,\; t_1 + c, t_2 + c \cdots, t_n + c)$$

for *any* c, where the left side represents the joint density function of the random variables $X_1 = X(t_1),\ X_2 = X(t_2),\ \cdots,\ X_n = X(t_n)$ and the right side corresponds to the joint density function of the random variables $X_1' = X(t_1 + c),\ X_2' = X(t_2 + c),\ \cdots,\ X_n' = X(t_n + c)$.
A process $X(t)$ is said to be **strict-sense stationary** if (14-14) is true for all $t_i,\ \ i = 1, 2, \cdots, n,\ \ n = 1, 2, \cdots$ and *any* c.

For a **first-order strict sense stationary process**, from (14-14) we have

$$f_X(x,t) \equiv f_X(x,t+c) \tag{14-15}$$

for any c. In particular $c = -t$ gives

$$f_X(x,t) = f_X(x) \tag{14-16}$$

i.e., the first-order density of $X(t)$ is independent of t. In that case

Similarly, for a **second-order strict-sense stationary process** we have from (14-14)

$$E[X(t)] = \int_{-\infty}^{+\infty} x f(x)dx = \mu, \quad a\ constant. \tag{14-17}$$

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

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

$$f_X(x_1,x_2,\ t_1,t_2) \equiv f_X(x_1,x_2,\ t_1-t_2) \tag{14-18}$$

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

$$\begin{aligned} R_{XX}(t_1, t_2) &= E\{X(t_1)X^*(t_2)\} \\ &\triangleq \int\int x_1 x_2^* f_X(x_1, x_2, \tau = t_1 - t_2) dx_1 dx_2 \\ &= R_{XX}(t_1 - t_2) = R_{XX}(\tau) = R_{XX}^*(-\tau), \end{aligned} \quad (14\text{-}19)$$

i.e., the autocorrelation function of a second order strict-sense stationary process depends only on the difference of the time indices $\tau = t_1 - t_2.$

Notice that (14-17) and (14-19) are consequences of the stochastic process being first and second-order strict sense stationary.

On the other hand, the basic conditions for the first and second order stationarity – Eqs. (14-16) and (14-18) – are usually difficult to verify.

In that case, we often resort to a looser definition of stationarity, known as **Wide-Sense Stationarity (W.S.S)**, by making use of

(14-17) and (14-19) as the necessary conditions. Thus, a process $X(t)$ is said to be **Wide-Sense Stationary** if

(i)

and $E\{X(t)\} = \mu$ (14-20)

(ii)

$$E\{X(t_1)X^*(t_2)\} = R_{XX}(t_1 - t_2), \tag{14-21}$$

i.e., for wide-sense stationary processes, the mean is a constant and the autocorrelation function depends only on the difference between the time indices. Notice that (14-20)-(14-21) does not say anything about the nature of the probability density functions, and instead deal with the average behavior of the process. Since (14-20)-(14-21) follow from (14-16) and (14-18), strict-sense stationarity always implies wide-sense stationarity. However, the converse is *not true* in general, the only exception being the Gaussian process.

This follows, since if $X(t)$ is a Gaussian process, then by definition $X_1 = X(t_1), X_2 = X(t_2), \cdots, X_n = X(t_n)$ are jointly Gaussian random variables for any $t_1, t_2 \cdots, t_n$ whose joint characteristic function is given by

$$\phi_{\underline{X}}(\omega_1, \omega_2, \cdots, \omega_n) = e^{j\sum_{k=1}^{n} \mu(t_k)\omega_k - \sum_{l,k}^{n} \sum C_{xx}(t_i, t_k)\omega_i \omega_k / 2} \tag{14-22}$$

where $C_{XX}(t_i, t_k)$ is as defined on (14-9). If *X*(*t*) is wide-sense stationary, then using (14-20)-(14-21) in (14-22) we get

$$\phi_{\underline{X}}(\omega_1, \omega_2, \cdots, \omega_n) = e^{j\sum_{k=1}^{n} \mu \omega_k - \frac{1}{2}\sum_{1=1}^{n} \sum_{k=1}^{n} C_{xx}(t_i - t_k)\omega_i \omega_k} \tag{14-23}$$

and hence if the set of time indices are shifted by a constant *c* to generate a new set of jointly Gaussian random variables $X_1' = X(t_1 + c)$, $X_2' = X(t_2 + c), \cdots, X_n' = X(t_n + c)$ then their joint characteristic function is identical to (14-23). Thus the set of random variables $\{X_i\}_{i=1}^n$ and $\{X_i'\}_{i=1}^n$ have the same joint probability distribution for all *n* and all *c*, establishing the strict sense stationarity of Gaussian processes from its wide-sense stationarity.

To summarize if *X*(*t*) is a Gaussian process, then

wide-sense stationarity (w.s.s) $\Rightarrow$ strict-sense stationarity (s.s.s).

Notice that since the joint p.d.f of Gaussian random variables depends only on their second order statistics, which is also the basis

Systems with Stochastic Inputs

A deterministic system[1] 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*).

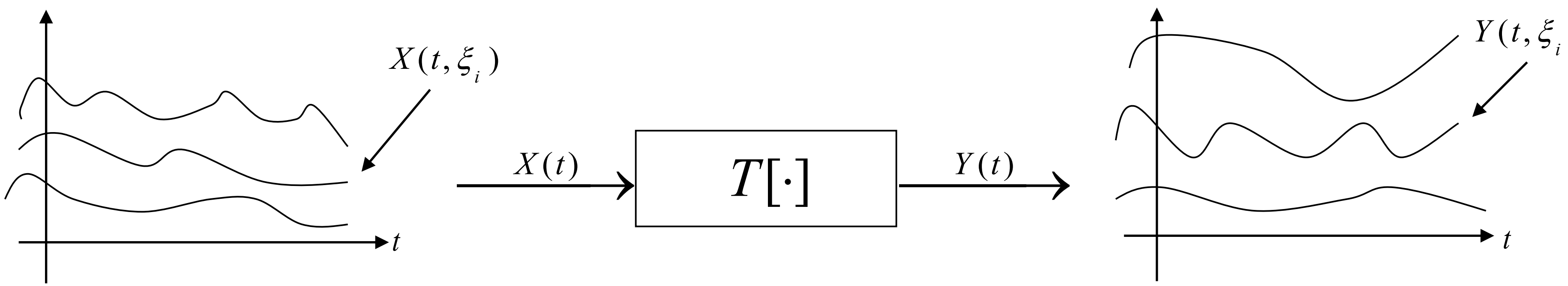


Fig. 14.3

Our goal is to study the output process statistics in terms of the input process statistics and the system function.

[1]A stochastic system on the other hand operates on both the variables *t* and ξ.

Linear Systems: $L[\cdot]$ represents a linear system if

$$L\{a_1X(t_1)+a_2X(t_2)\}=a_1L\{X(t_1)\}+a_2L\{X(t_2)\}. \qquad (14\text{-}28)$$

Let

$$Y(t)=L\{X(t)\} \qquad (14\text{-}29)$$

represent the output of a linear system.

Time-Invariant System: $L[\cdot]$ represents a time-invariant system if

$$Y(t)=L\{X(t)\}\Rightarrow L\{X(t-t_0)\}=Y(t-t_0) \qquad (14\text{-}30)$$

i.e., shift in the input results in the same shift in the output also.

If $L[\cdot]$ satisfies both (14-28) and (14-30), then it corresponds to a linear time-invariant (LTI) system.

LTI systems can be uniquely represented in terms of their output to a delta function

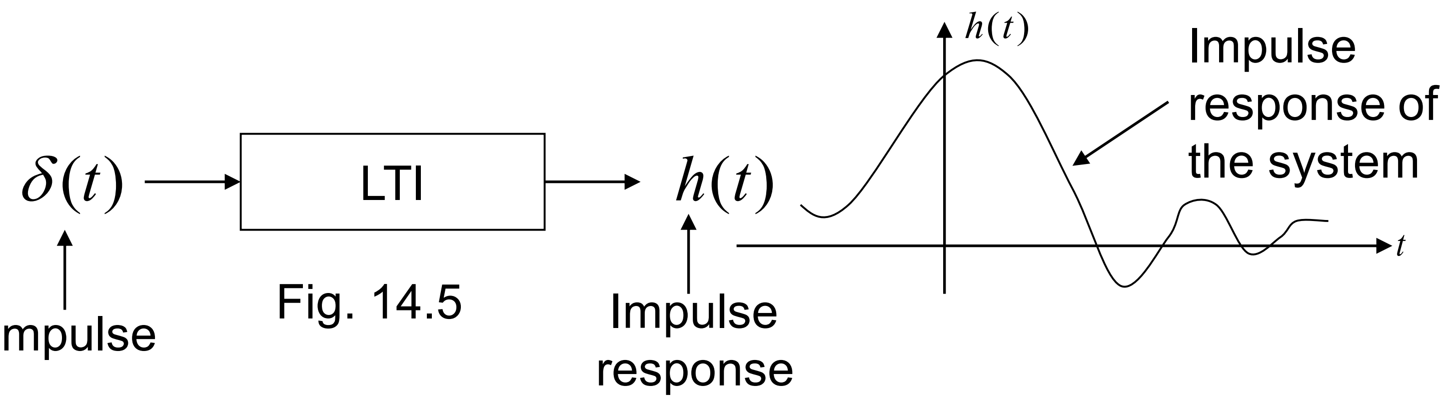
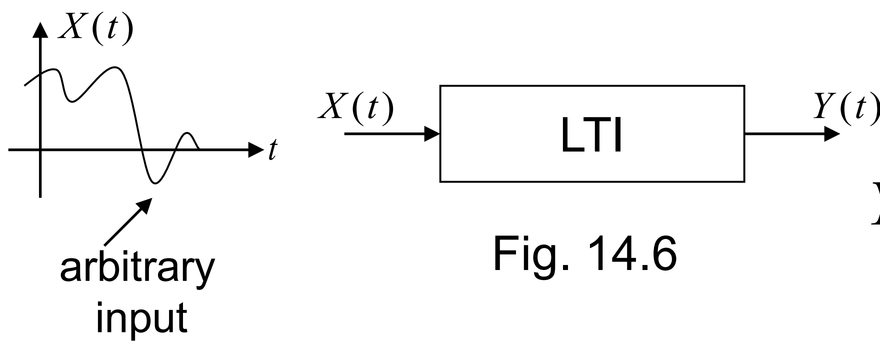


Fig. 14.5

then



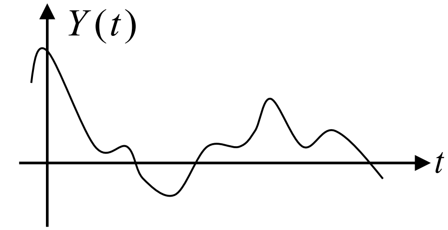


Fig. 14.6

$$Y(t) = \int_{-\infty}^{+\infty} h(t-\tau)X(\tau)d\tau$$

$$= \int_{-\infty}^{+\infty} h(\tau)X(t-\tau)d\tau \quad (14\text{-}31)$$

Eq. (14-31) follows by expressing *X*(*t*) as

$$X(t) = \int_{-\infty}^{+\infty} X(\tau)\delta(t-\tau)d\tau \qquad (14\text{-}32)$$

and applying (14-28) and (14-30) to $Y(t) = L\{X(t)\}.$ Thus

$$Y(t) = L\{X(t)\} = L\{\int_{-\infty}^{+\infty} X(\tau)\delta(t-\tau)d\tau\}$$

$$= \int_{-\infty}^{+\infty} L\{X(\tau)\delta(t-\tau)d\tau\}$$

By Linearity

$$= \int_{-\infty}^{+\infty} X(\tau)L\{\delta(t-\tau)\}d\tau$$

By Time-invariance

$$= \int_{-\infty}^{+\infty} X(\tau)h(t-\tau)d\tau = \int_{-\infty}^{+\infty} h(\tau)X(t-\tau)d\tau. \quad (14\text{-}33)$$

Output Statistics: Using (14-33), the mean of the output process is given by

$$\mu_Y(t) = E\{Y(t)\} = \int_{-\infty}^{+\infty} E\{X(\tau)h(t-\tau)d\tau\}$$

$$= \int_{-\infty}^{+\infty} \mu_X(\tau)h(t-\tau)d\tau = \mu_X(t) * h(t). \qquad (14\text{-}34)$$

Similarly the cross-correlation function between the input and output processes is given by

$$R_{XY}(t_1,t_2) = E\{X(t_1)Y^*(t_2)\}$$

$$= E\{X(t_1)\int_{-\infty}^{+\infty} X^*(t_2-\alpha)h^*(\alpha)d\alpha\}$$

$$= \int_{-\infty}^{+\infty} E\{X(t_1)X^*(t_2-\alpha)\}h^*(\alpha)d\alpha$$

$$= \int_{-\infty}^{+\infty} R_{XX}(t_1,t_2-\alpha)h^*(\alpha)d\alpha \qquad (14\text{-}35)$$

$$= R_{XX}(t_1,t_2) * h^*(t_2).$$

Finally the output autocorrelation function is given by

$$
\begin{aligned}
R_{YY}(t_1,t_2) &= E\{Y(t_1)Y^*(t_2)\} \\
&= E\{\int_{-\infty}^{+\infty} X(t_1-\beta)h(\beta)d\beta\, Y^*(t_2)\} \\
&= \int_{-\infty}^{+\infty} E\{X(t_1-\beta)Y^*(t_2)\}h(\beta)d\beta \\
&= \int_{-\infty}^{+\infty} R_{XY}(t_1-\beta,t_2)h(\beta)d\beta \\
&= R_{XY}(t_1,t_2) * h(t_1),
\end{aligned}
\tag{14-36}
$$

or

$$
R_{YY}(t_1,t_2) = R_{XX}(t_1,t_2) * h^*(t_2) * h(t_1). \tag{14-37}
$$

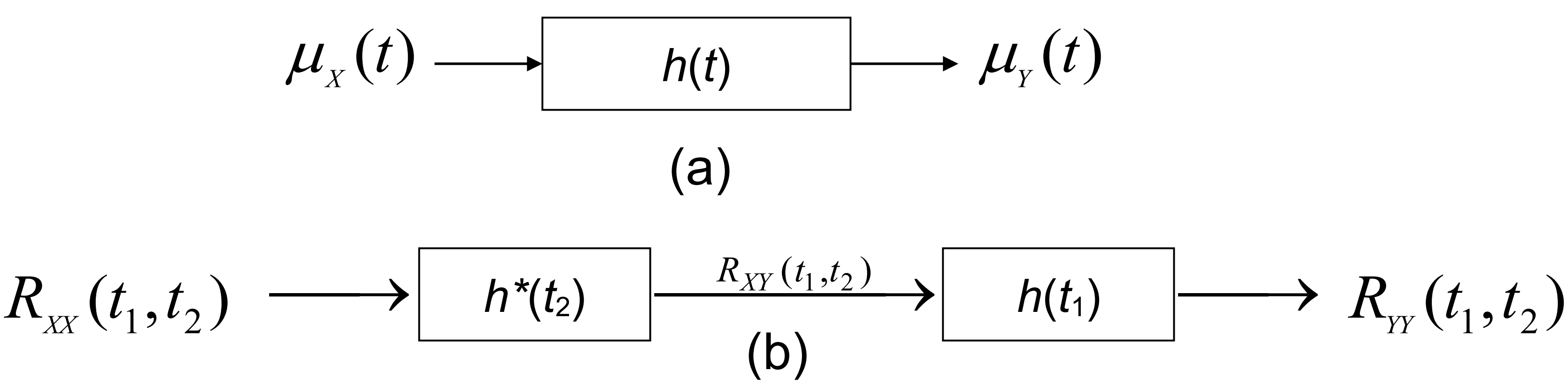


Fig. 14.7

In particular if $X(t)$ is wide-sense stationary, then we have $\mu_X(t) = \mu_X$ so that from (14-34)

$$\mu_Y(t) = \mu_X \int_{-\infty}^{+\infty} h(\tau)d\tau = \mu_X c, \quad a\ constant. \tag{14-38}$$

Also $R_{XX}(t_1,t_2) = R_{XX}(t_1 - t_2)$ so that (14-35) reduces to

$$R_{XY}(t_1,t_2) = \int_{-\infty}^{+\infty} R_{XX}(t_1 - t_2 + \alpha)h^*(\alpha)d\alpha$$

$$= R_{XX}(\tau) * h^*(-\tau) = R_{XY}(\tau), \quad \tau = t_1 - t_2. \tag{14-39}$$

Thus $X(t)$ and $Y(t)$ are jointly w.s.s. Further, from (14-36), the output autocorrelation simplifies to

$$R_{YY}(t_1,t_2) \triangleq \int_{-\infty}^{+\infty} R_{XY}(t_1 - \beta - t_2)h(\beta)d\beta, \quad \tau = t_1 - t_2$$

$$= R_{XY}(\tau) * h(\tau) = R_{YY}(\tau). \tag{14-40}$$

From (14-37), we obtain

$$R_{YY}(\tau) = R_{XX}(\tau) * h^*(-\tau) * h(\tau). \tag{14-41}$$

From (14-38)-(14-40), the output process is also wide-sense stationary. This gives rise to the following representation

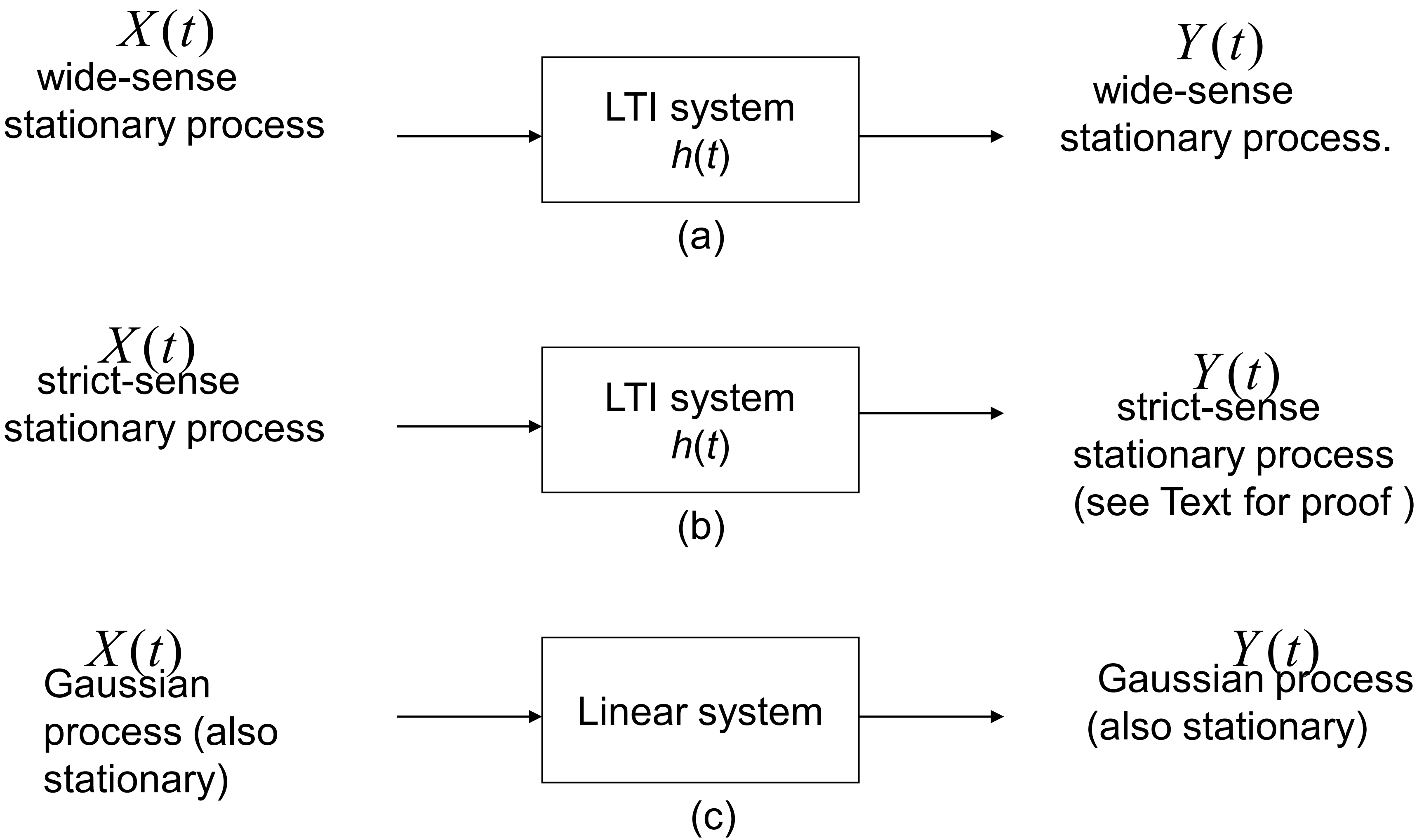


Fig. 14.8

Discrete Time Stochastic Processes:

A discrete time stochastic process $X_n = X(nT)$ is a sequence of random variables. The mean, autocorrelation and auto-covariance functions of a discrete-time process are gives by

$$\mu_n = E\{X(nT)\} \tag{14-57}$$

and

$$R(n_1,n_2) = E\{X(n_1T)X^*(n_2T)\} \tag{14-58}$$

respectively. As before strict sense stationarity and wide-sense stationarity definitions apply here also.
For example, $X(nT)$ is wide sense stationary if

$$C(n_1,n_2) = R(n_1,n_2) - \mu_{n_1}\mu_{n_2}^* \tag{14-59}$$

and

$$E\{X(nT)\} = \mu, \quad a\ constant \tag{14-60}$$

$$E[X\{(k+n)T\}X^*\{(k)T\}] = R(n) = r_n \triangleq r_{-n}^* \tag{14-61}$$

Power Spectrum

For a deterministic signal *x*(t), the spectrum is well defined: If $X(\omega)$ represents its Fourier transform, i.e., if

$$X(\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t}dt, \qquad (18\text{-}1)$$

then $|X(\omega)|^2$ represents its energy spectrum. This follows from Parseval's theorem since the signal energy is given by

$$\int_{-\infty}^{+\infty} x^2(t)dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |X(\omega)|^2 \, d\omega = E. \qquad (18\text{-}2)$$

Thus $|X(\omega)|^2 \Delta\omega$ represents the signal energy in the band $(\omega, \omega + \Delta\omega)$ (see Fig 18.1).

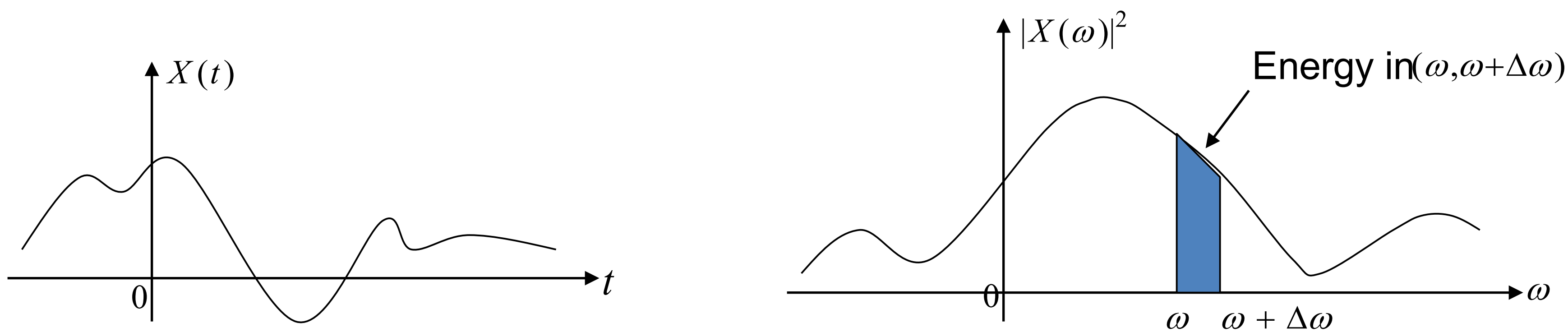


Fig 18.1

However for stochastic processes, a direct application of (18-1) generates a sequence of random variables for every ω. Moreover, for a stochastic process, $E\{|X(t)|^2\}$ represents the ensemble average power (instantaneous energy) at the instant t.

To obtain the spectral distribution of power versus frequency for stochastic processes, it is best to avoid infinite intervals to begin with, and start with a finite interval $(-T, T)$ in (18-1). Formally, partial Fourier transform of a process $X(t)$ based on $(-T, T)$ is given by

so that

$$X_T(\omega) = \int_{-T}^{T} X(t) e^{-j\omega t} dt \tag{18-3}$$

represents the power distribution associated with that realization based on $(-T, T)$. Notice that (18-4) represents a random variable for every and its ensemble average gives, the average power distribution based on $(-T, T)$. Thus

$$\frac{|X_T(\omega)|^2}{2T} = \frac{1}{2T} \left| \int_{-T}^{T} X(t) e^{-j\omega t} dt \right|^2 \quad \omega, \tag{18-4}$$

$$P_T(\omega) = E\left\{\frac{|X_T(\omega)|^2}{2T}\right\} = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} E\{X(t_1)X^*(t_2)\}e^{-j\omega(t_1-t_2)}dt_1dt_2$$

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

(18-5)

represents the power distribution of $X(t)$ based on $(-T, T)$. For wide sense stationary (w.s.s) processes, it is possible to further simplify (18-5). Thus if $X(t)$ is assumed to be w.s.s, then and (18-5) simplifies to

$$R_{XX}(t_1,t_2) = R_{XX}(t_1 - t_2)$$

Let $\tau = t_1 - t_2$ and proceeding as in (14-24), we get

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

to be the power distribution of the w.s.s. process $X(t)$ based on $(-T, T)$. Finally letting $T \to \infty$ in (18-6), we obtain

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

(18-6)

$$= \int_{-2T}^{2T} R_{XX}(\tau)e^{-j\omega\tau}\left(1 - \tfrac{|\tau|}{2T}\right)d\tau \;\geq\; 0$$

$$S_{XX}(\omega) = \lim_{T \to \infty} P_T(\omega) = \int_{-\infty}^{+\infty} R_{XX}(\tau) e^{-j\omega\tau} d\tau \geq 0 \qquad (18\text{-}7)$$

to be the *power spectral density* of the w.s.s process *X*(*t*). Notice that

$$R_{XX}(\omega) \xleftrightarrow{\text{F}\cdot\text{T}} S_{XX}(\omega) \geq 0. \qquad (18\text{-}8)$$

i.e., the autocorrelation function and the power spectrum of a w.s.s Process form a Fourier transform pair, a relation known as the **Wiener-Khinchin Theorem**. From (18-8), the inverse formula gives

$$R_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{XX}(\omega) e^{j\omega\tau} d\omega \qquad (18\text{-}9)$$

and in particular for $\tau = 0,$ we get

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{XX}(\omega) d\omega = R_{XX}(0) = E\{|X(t)|^2\} = P, \quad \text{the total power.} \qquad (18\text{-}10)$$

From (18-10), the area under $S_{XX}(\omega)$ represents the total power of the process *X*(*t*), and hence $S_{XX}(\omega)$ truly represents the power spectrum. (Fig 18.2).

If $X(t)$ is a real w.s.s process, then $R_{XX}(\tau) = R_{XX}(-\tau)$ so that

$$\begin{aligned} S_{XX}(\omega) &= \int_{-\infty}^{+\infty} R_{XX}(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{+\infty} R_{XX}(\tau) \cos \omega\tau d\tau \\ &= 2\int_{0}^{\infty} R_{XX}(\tau) \cos \omega\tau d\tau = S_{XX}(-\omega) \geq 0 \end{aligned}$$

so that the power spectrum is an even function, (in addition to being real and nonnegative). (18-13)