Introduction

- Provide a brief review of topics that will help us:
  - Statistically characterize network traffic flow
  - Model and estimate performance parameters
- Set stage for discussion of traffic management, routing and compression later in the course
- **NOT** a condensed class in probability theory

1

Probability – Axiomatic Definition

Common Axioms:

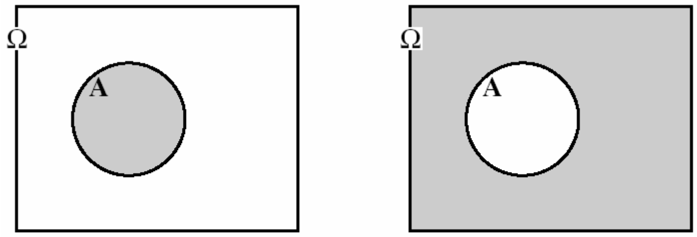
1. $0 \leq Pr[A] \leq 1$ for each even A
2. $Pr[\Omega] = 1$
3. $Pr[A \cup B] = Pr[A] + Pr[B]$ if A and B are mutually exclusive

Important Laws:

1. $Pr[\bar{A}] = 1 - Pr[A]$
2. $Pr[A \cap B] = 0$ if A and B are mutually exclusive
3. $Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$
4. $Pr[A \cup B \cup C] = Pr[A] + Pr[B] + Pr[C] - Pr[A \cap B] - Pr[A \cap C] - Pr[B \cap C] + Pr[A \cap B \cap C]$

2

Venn Diagrams

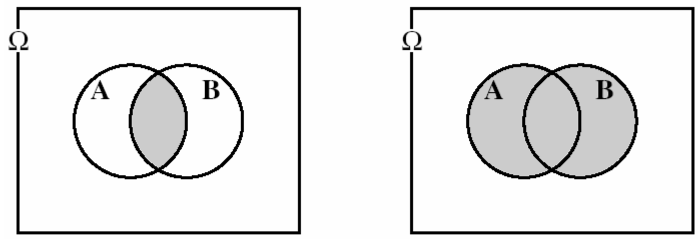


(a) A (b) NOT A $\overline{A}$

Complementation

3

Venn Diagrams

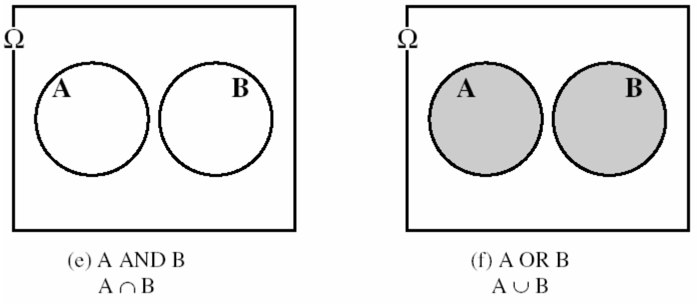


(c) A AND B $A \cap B$ (d) A OR B $A \cup B$

Intersection

4

Venn Diagrams



Mutual Exclusivity

Probability Definitions

Relative Frequency Definition:

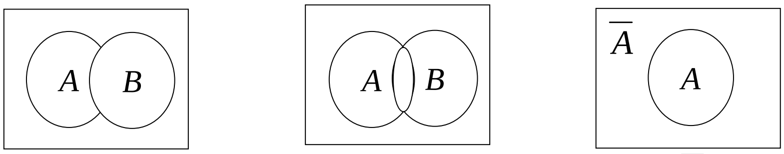
$$Pr[A] = \lim_{n \to \infty} \frac{n_A}{n}$$

where *n* is the number of trials, and n_A the number of times event *A* occurred

Classical Definition:

$$Pr[A] = \frac{N_A}{N}$$

where *N* is the number of equally likely outcomes and N_A is the number of outcomes in which event *A* occurs

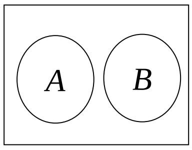


$A \cup B$ $\quad A \cap B$ $\quad \bar{A}$

Fig.1.1

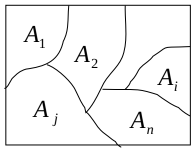
- If $A \cap B = \phi,$ the empty set, then A and B are said to be mutually exclusive (M.E).
- A partition of Ω is a collection of mutually exclusive subsets of Ω such that their union is Ω.

$$A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1} A_i = \Omega. \qquad (1\text{-}5)$$



$A \cap B = \phi$

Fig. 1.2



PILLAI

Conditional Probability

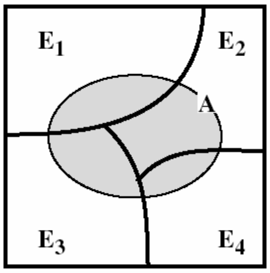
- The conditional probability of an event A, given that event B has occurred is:

$$\Pr[A \mid B] = \frac{\Pr[AB]}{\Pr[B]}$$

- Where Pr[AB] encompasses all possible outcomes that satisfy the conditions
- if A and B are independent events then Pr[AB] = Pr[A]Pr[B]

Total Probability

- Given a set of mutually exclusive events E_1, E_2, …, E_n covering all possible outcomes, and
- Given an arbitrary event A, then:

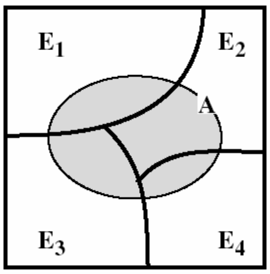


$$Pr[A] = \sum_{i=1}^{n} Pr[A \mid E_i] Pr[E_i]$$

9

Bayes's Theorem

- "Posterior odds" – the probability that an event really occurred, given evidence in favor of it:

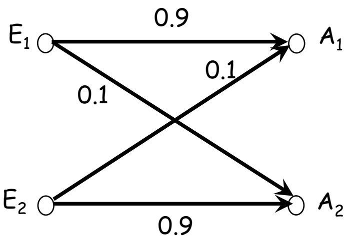


$$Pr[E_i \mid A] = \frac{Pr[A \mid E_i]\, Pr[E_i]}{Pr[A]} = \frac{Pr[A \mid E_i]\, Pr[E_i]}{\sum_{i=1}^{n} Pr[A \mid E_i] Pr[E_i]}$$

10

Example: binary symmetric Channel

- $P(A_1)=P(A_1|B_1)P(B_1)$
  $+ P(A_1|B_2)P(B_2)$
  $=0.58$
- $P(A_2)=P(A_2|B_1)P(B_1)$
  $+ P(A_2|B_2)P(B_2)$
  $=0.42$
- $P(B_1|A_1)=\{P(A_1|B_1)P(B_1)\}/P(A_1)=0.931$



11

Bayes's Theorem Example – "The Juror's Fallacy"

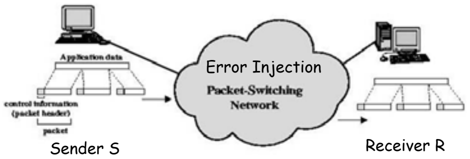
- Hit & run accident involving a taxi
- 85% of taxis are yellow, 15% are blue
- Eyewitness reported that the taxi involved in the accident was blue
- Data shows that eyewitnesses are correct on car color 80% of the time
- What is the probability that the cab was blue?

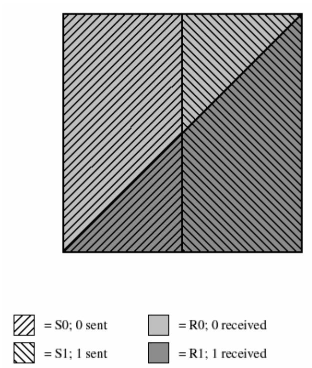
$$\Pr[\text{Blue}|\text{WB}] = \frac{\Pr[\text{WB}|\text{Blue}]\,\Pr[\text{Blue}]}{\Pr[\text{WB}|\text{Blue}]\,\Pr[\text{Blue}] + \Pr[\text{WB}|\text{Yellow}]\,\Pr[\text{Yellow}]}$$

$$= \frac{(0.8)(0.15)}{(0.8)(0.15) + (0.2)(0.85)} = 0.41$$

12

Bayes's Theorem Example





- Network injects errors (flips bits)
- Assume Pr[S1] = Pr[S0] = p = 0.5
- Assume Pr[R1] = Pr[R0] = (1-p) = 0.5
- Given error injection, such that Pr[R0 | S1] = p_a and Pr[R1 | S0] = p_b, then :

$$\Pr[S1 \mid R0] = \frac{\Pr[R0 \mid S1]\Pr[S1]}{\Pr[R0 \mid S1]\Pr[S1] + \Pr[R0 \mid S0]\Pr[S0]} = \frac{p_a\, p}{p_a\, p + (1-p_b)(1-p)}$$

13

Although simple enough, Bayes' theorem has an interesting interpretation: *P*(*A*) represents the a-priori probability of the event *A*. Suppose *B* has occurred, and assume that *A* and *B* are not independent. How can this new information be used to update our knowledge about *A*? Bayes' rule take into account the new information ("*B* has occurred") and gives out the a-posteriori probability of *A* given *B*.

We can also view the event *B* as new knowledge obtained from a fresh experiment. We know something about *A* as *P*(*A*). The new information is available in terms of *B*. The new information should be used to improve our knowledge/understanding of *A*. Bayes' theorem gives the exact mechanism for incorporating such new information.

14

PILLAI

Random Variables

- Association of real numbers with events, e.g. assigning a value to each outcome of an experiment
- A random variable *X* is a function that assigns a real number to every outcome in a sample space, and satisfies the following conditions:
  1. the set $\{X \leq x\}$ is an event for every x
  2. $Pr[X= \infty] = Pr[X = -\infty] = 0$
- A random variable can be continuous or discrete

15

Random Variables

- Continuous random variables can be described by either a distribution function or a density function
- Discrete random variables are described by a probability function $P_x(k)$
- Random variable characteristics:
  - Mean value
  - Second moment
  - Variance
  - Standard deviation

16

Bernoulli Trials

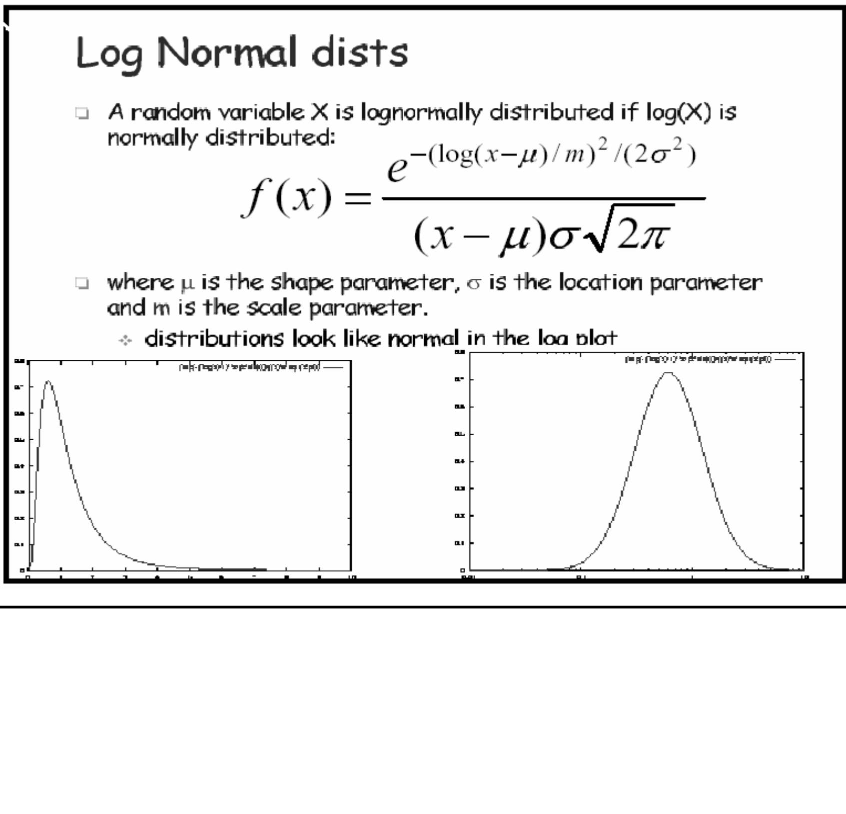
- A single trial can be classified as a "success" or a "failure"
  - P(success)=p
  - P(failure)=1-p=q
- Small set of trials; N
- Small set of outcomes (A); K
- For a N series of trials, we have

$$Po(K) = \underbrace{P(A)P(A)\cdots P(A)}_{K}\ \underbrace{P(\overline{A})P(\overline{A})\cdots P(\overline{A})}_{N-K} = p^k q^{n-k}.$$

$$P(K) = \binom{N}{K} p^K q^{N-K}$$ (binomial Distribution)

17

Probability Functions



3

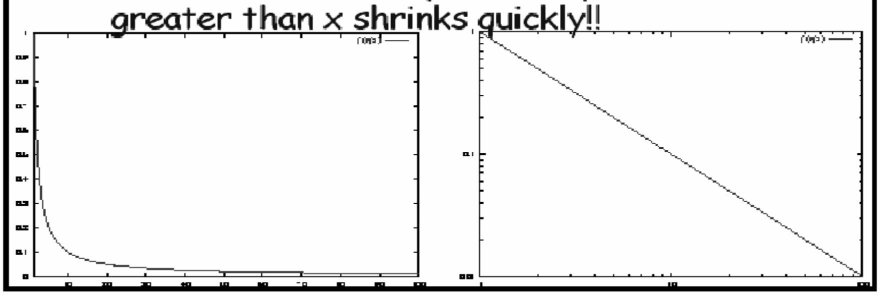
Probability Functions

Power Laws

- Experiments that have trials whose distributions have the general form:

$$P[X > x] \approx c\,x^{-a} \qquad x \to \infty,\ a, c > 0$$

- As x increases, the probability that X is greater than x shrinks quickly!!



9

Probability Functions

Power Laws

- Pareto: the simplest power law distribution

$$f(x) = \frac{ab^a}{x^{a+1}}, \quad x \geq b, \quad a > 0, \quad b > 0$$

⇒ cumulative distribution function (cdf)

$$P[X \leq x] = 1 - \frac{b^a}{x^a}$$

⇒ complementary cumulative distribution function (ccdf)

$$P[X > x] = \frac{b^a}{x^a}$$

20

Probability Functions

Properties of the Pareto

- a determines the mean and variance
- b determines minimum value x can take

$$E[x] = \begin{cases} \infty & a \le 1 \\ \dfrac{ab}{(1-a)} & a > 1 \end{cases}$$

$$\operatorname{var}[x] = \begin{cases} \infty & a \le 2 \\ \dfrac{ab^2}{(a-1)^2(a-2)} & a > 2 \end{cases}$$

Note: $1<a<2 \Rightarrow$ finite mean, *infinite variance*!

21

Probability Functions

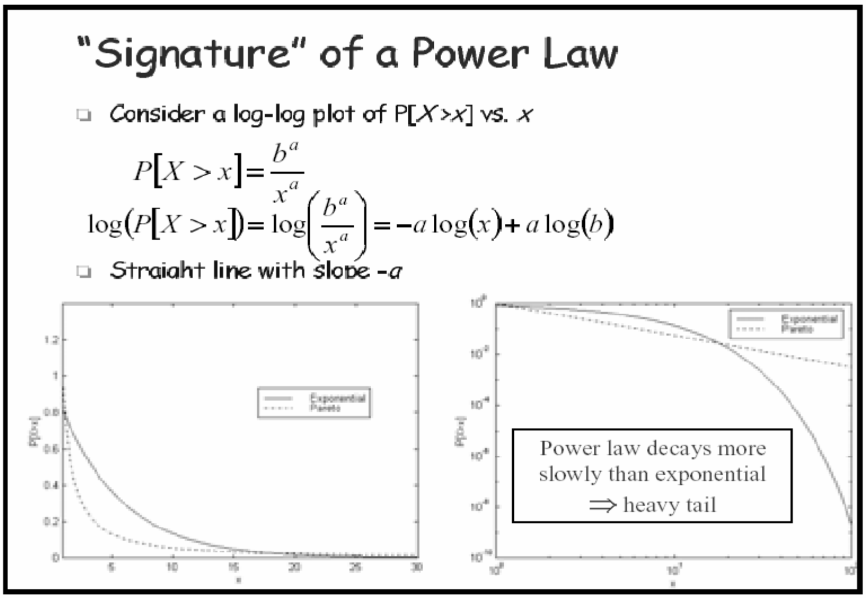
Pareto

- Pareto studied the distribution of incomes.
  - Few people have a lot of money, many are poor.
- Pareto distributions are concerned with how many events are larger than x: this number is an inverse power of x.

$$P[X > x] = \frac{b^a}{x^a}$$

22

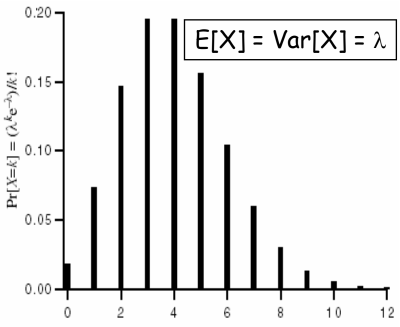
Probability Functions



23

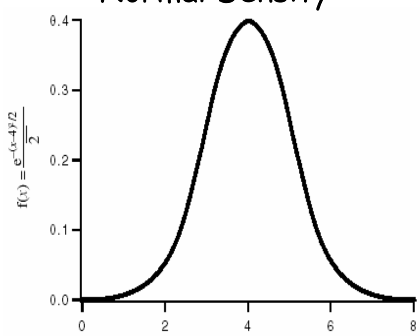
Probability Distributions

Poisson Distribution



(c) Poisson distribution (λ = 1)

Normal Density



(d) Normal probability density (μ = 4, σ = 1)

$$\Pr[X=k] = \frac{\lambda^k}{k!} e^{-\lambda}$$

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

24

Functions of Random Variables: discrete-type

Suppose X is a discrete-type r.v with

$$P(X = x_i) = p_i, \quad x = x_1, x_2, \cdots, x_i, \cdots \tag{5-39}$$

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

$$P(Y = y_i) = P(X = x_i) = p_i, \quad y = y_1, y_2, \cdots, y_i, \cdots \tag{5-40}$$

Example 5.8: Suppose $X \sim P(\lambda)$, so that

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0,1,2,\cdots \tag{5-41}$$

Define $Y = X^2 + 1$. Find the p.m.f of Y.
Solution: X takes the values $0,1,2,\cdots,k,\cdots$ so that Y only takes the value $1,3,\cdots,k^2+1,\cdots$ and

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

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

$$P(Y = j) = P\left(X = \sqrt{j-1}\right) = e^{-\lambda} \frac{\lambda^{\sqrt{j-1}}}{(\sqrt{j-1})!}, \quad j = 1,3,\cdots,k^2+1,\cdots. \tag{5-42}$$

Functions of a Random Variable

Continues Case:

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

$$Y = g(X). \qquad (5\text{-}1)$$

27

PILLAI

Example 5.1: $Y = aX + b$ (5-4)

Solution: Suppose $a > 0$.

$$F_Y(y) = P(Y(\xi) \le y) = P(aX(\xi) + b \le y) = P\left(X(\xi) \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right). \qquad (5\text{-}5)$$

and

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right). \qquad (5\text{-}6)$$

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

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

$$= 1 - F_X\left(\frac{y-b}{a}\right), \qquad (5\text{-}7)$$

and hence

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right). \text{ Then, } f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \qquad (5\text{-}8)$$

28

PILLAI

Note: As a general approach, given $Y = g(X)$, first sketch the graph $y = g(x)$, and determine the range space of *y*. Suppose $a < y < b$ is the range space of $y = g(x)$. Then clearly for $y < a$, $F_Y(y) = 0$, and for $y > b$, $F_Y(y) = 1$, so that $F_Y(y)$ can be nonzero only in $a < y < b$. Next, determine whether there are discontinuities in the range space of *y*. If so evaluate $P(Y(\xi) = y_i)$ at these discontinuities. In the continuous region of *y*, use the basic approach

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

and determine appropriate events in terms of the r.v *X* for every *y*. Finally, we must have $F_Y(y)$ for $-\infty \le y \le +\infty$, and obtain

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

29

PILLAI

However, if $Y = g(X)$ is a continuous function, it is easy to establish a direct procedure to obtain $f_Y(y)$. A continuos function *g*(*x*) with $g'(x)$ nonzero at all but a finite number of points, has only a finite number of maxima and minima, and it eventually becomes monotonic as $|x| \to \infty$. Consider a specific *y* on the *y*-axis, and a positive increment Δy as shown in Fig. 5.4

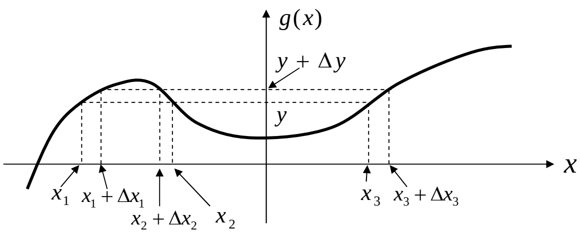


Fig. 5.4

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

30

PILLAI

For small $\Delta y, \Delta x_i$, making use of the approximation in (5-26), we get

$$f_Y(y)\Delta y = f_X(x_1)\Delta x_1 + f_X(x_2)(-\Delta x_2) + f_X(x_3)\Delta x_3. \qquad (5\text{-}28)$$

In this case, $\Delta x_1 > 0,\ \ \Delta x_2 < 0$ and $\Delta x_3 > 0,$ so that (5-28) can be rewritten as

$$f_Y(y) = \sum_i f_X(x_i)\frac{|\Delta x_i|}{\Delta y} = \sum_i \frac{1}{|\Delta y / \Delta x_i|} f_X(x_i) \qquad (5\text{-}29)$$

and as $\Delta y \to 0,$ (5-29) can be expressed as

$$f_Y(y) = \sum_i \frac{1}{|dy/dx|_{x_i}} f_X(x_i) = \sum_i \frac{1}{|g'(x_i)|} f_X(x_i). \qquad (5\text{-}30)$$

The summation index *i* in (5-30) depends on *y*, and for every *y* the equation $y = g(x_i)$ must be solved to obtain the total number of solutions at every *y*, and the actual solutions $x_1, x_2, \cdots$ all in terms of *y*.

For example, if $Y = X^2,$ then for all $y > 0,\ \ x_1 = -\sqrt{y}$ and $x_2 = +\sqrt{y}$ represent the two solutions for each *y*. Notice that the solutions x_i are all in terms of *y* so that the right side of (5-30) is only a function of *y*. Referring back to the example $Y = X^2$ (Example 5.2) here for each $y > 0,$ there are two solutions given by $x_1 = -\sqrt{y}$ and $x_2 = +\sqrt{y}.$ ($f_Y(y) = 0$ for $y < 0$). Moreover

$$\frac{dy}{dx} = 2x \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x = x_i} = 2\sqrt{y}$$

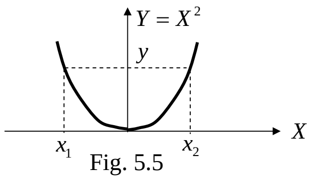


Fig. 5.5

and using (5-30) we get

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right), & y > 0, \\ 0, & \text{otherwise}\ , \end{cases} \qquad (5\text{-}31)$$

which agrees with (5-14).

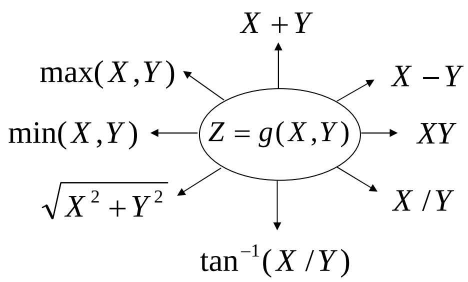
8. One Function of Two Random Variables

Given two random variables X and Y and a function $g(x,y)$, we form a new random variable Z as

$$Z = g(X, Y). \qquad (8\text{-}1)$$

Given the joint p.d.f $f_{XY}(x, y)$, how does one obtain $f_Z(z)$, the p.d.f of Z ? Problems of this type are of interest from a practical standpoint. For example, a receiver output signal usually consists of the desired signal buried in noise, and the above formulation in that case reduces to $Z = X + Y$.

It is important to know the statistics of the incoming signal for proper receiver design. In this context, we shall analyze problems of the following type:



(8-2)

Referring back to (8-1), to start with

$$F_Z(z) = P\big(Z(\xi) \le z\big) = P\big(g(X,Y) \le z\big) = P\big[(X,Y) \in D_z\big]$$
$$= \int\int_{x,y \in D_z} f_{XY}(x, y)\,dxdy\,, \qquad (8\text{-}3)$$

where D_z in the *XY* plane represents the region such that $g(x,y) \le z$ is satisfied. Note that D_z need not be simply connected (Fig. 8.1). From (8-3), to determine $F_Z(z)$ it is enough to find the region D_z for every *z*, and then evaluate the integral there.

We shall illustrate this method through various examples.

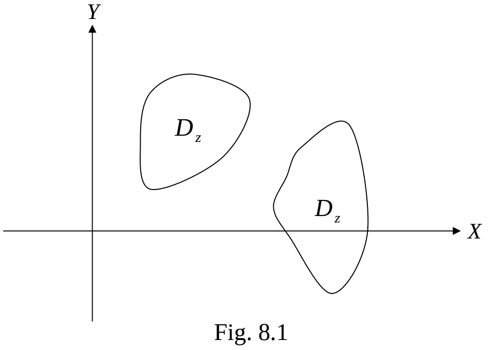


Fig. 8.1

Example 8.1: *Z* = *X* + *Y*. Find $f_Z(z)$.
Solution:

$$F_Z(z) = P\left(X + Y \le z\right) = \int_{y=-\infty}^{+\infty} \int_{x=-\infty}^{z-y} f_{XY}(x,y)dxdy, \qquad (8\text{-}4)$$

since the region D_z of the *xy* plane where $x + y \le z$ is the shaded area in Fig. 8.2 to the left of the line $x + y = z.$ Integrating over the horizontal strip along the *x*-axis first (inner integral) followed by sliding that strip along the *y*-axis from $-\infty$ to $+\infty$ (outer integral) we cover the entire shaded area.

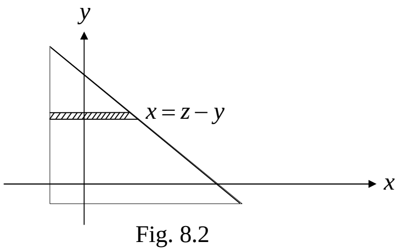


Fig. 8.2

We can find $f_Z(z)$ by differentiating $F_Z(z)$ directly. In this context, it is useful to recall the differentiation rule in (7-15) - (7-16) due to Leibnitz. Suppose

$$H(z) = \int_{a(z)}^{b(z)} h(x,z)dx. \tag{8-5}$$

Then

$$\frac{dH(z)}{dz} = \frac{db(z)}{dz} h(b(z),z) - \frac{da(z)}{dz} h(a(z),z) + \int_{a(z)}^{b(z)} \frac{\partial h(x,z)}{\partial z} dx. \tag{8-6}$$

Using (8-6) in (8-4) we get

$$f_Z(z) = \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-y} f_{XY}(x,y)dx \right) dy = \int_{-\infty}^{+\infty} \left(1 \cdot f_{XY}(z-y,y) - 0 + \frac{\partial f_{XY}(x,y)}{\partial z} \right) dy$$

$$= \int_{-\infty}^{+\infty} f_{XY}(z-y,y)dy. \tag{8-7}$$

Alternatively, the integration in (8-4) can be carried out first along the *y*-axis followed by the *x*-axis as in Fig. 8.3.

37

PILLAI

In that case

$$F_Z(z) = \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{z-x} f_{XY}(x,y)dxdy, \tag{8-8}$$

and differentiation of (8-8) gives

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{x=-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{y=-\infty}^{z-x} f_{XY}(x,y)dy \right) dx$$

$$= \int_{x=-\infty}^{+\infty} f_{XY}(x,z-x)dx. \tag{8-9}$$

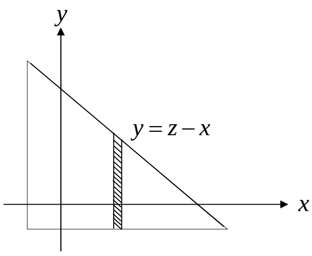


Fig. 8.3

If *X* and *Y* are independent, then

$$f_{XY}(x,y) = f_X(x) f_Y(y) \tag{8-10}$$

and inserting (8-10) into (8-8) and (8-9), we get

$$f_Z(z) = \int_{y=-\infty}^{+\infty} f_X(z-y) f_Y(y)dy = \int_{x=-\infty}^{+\infty} f_X(x) f_Y(z-x)dx. \tag{8-11}$$

38

PILLAI

The above integral is the standard convolution of the functions $f_X(z)$ and $f_Y(z)$ expressed two different ways. We thus reach the following conclusion: If two r.vs are independent, then the density of their sum equals the convolution of their density functions.

As a special case, suppose that $f_X(x) = 0$ for $x < 0$ and $f_Y(y) = 0$ for $y < 0,$ then we can make use of Fig. 8.4 to determine the new limits for $D_z.$

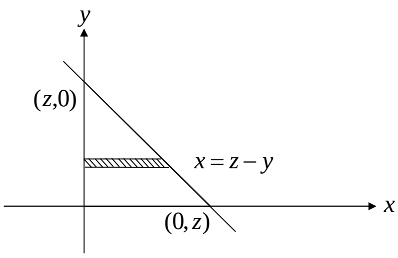


Fig. 8.4

39
PILLAI

In that case

$$F_Z(z) = \int_{y=0}^{z} \int_{x=0}^{z-y} f_{XY}(x, y)dxdy$$

or

$$f_Z(z) = \int_{y=0}^{z} \left(\frac{\partial}{\partial z} \int_{x=0}^{z-y} f_{XY}(x, y)dx \right) dy = \begin{cases} \int_0^z f_{XY}(z-y, y)dy, & z > 0, \\ 0, & z \le 0. \end{cases} \qquad \text{(8-12)}$$

On the other hand, by considering vertical strips first in Fig. 8.4, we get

$$F_Z(z) = \int_{x=0}^{z} \int_{y=0}^{z-x} f_{XY}(x, y)dydx$$

or

$$f_Z(z) = \int_{x=0}^{z} f_{XY}(x, z-x)dx = \begin{cases} \int_{y=0}^{z} f_X(x) f_Y(z-x)dx, & z > 0, \\ 0, & z \le 0, \end{cases} \qquad \text{(8-13)}$$

if X and Y are independent random variables.

40
PILLAI

Example 8.2: Suppose *X* and *Y* are independent exponential r.vs with common parameter λ, and let *Z* = *X* + *Y*. Determine $f_Z(z)$.

Solution: We have $f_X(x) = \lambda e^{-\lambda x}U(x), \quad f_Y(y) = \lambda e^{-\lambda y}U(y),$ (8-14)

and we can make use of (13) to obtain the p.d.f of *Z* = *X* + *Y*.

$$f_Z(z) = \int_0^z \lambda^2 e^{-\lambda x} e^{-\lambda(z-x)} dx = \lambda^2 e^{-\lambda z} \int_0^z dx = z\lambda^2 e^{-\lambda z} U(z). \qquad (8\text{-}15)$$

As the next example shows, care should be taken in using the convolution formula for r.vs with finite range.

Example 8.3: *X* and *Y* are independent uniform r.vs in the common interval (0,1). Determine $f_Z(z)$, where *Z* = *X* + *Y*.

Solution: Clearly, $Z = X + Y \Rightarrow 0 < z < 2$ here, and as Fig. 8.5 shows there are two cases of *z* for which the shaded areas are quite different in shape and they should be considered separately.

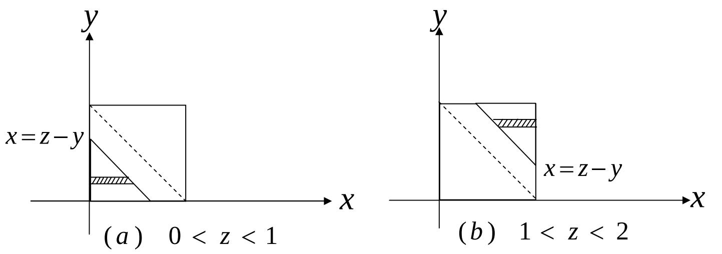


Fig. 8.5

For $0 \le z < 1$,

$$F_Z(z) = \int_{y=0}^{z} \int_{x=0}^{z-y} 1 \, dxdy = \int_{y=0}^{z} (z-y)dy = \frac{z^2}{2}, \quad 0 \le z < 1. \qquad (8\text{-}16)$$

For $1 \le z < 2$, notice that it is easy to deal with the unshaded region. In that case

$$F_Z(z) = 1 - P(Z > z) = 1 - \int_{y=z-1}^{1} \int_{x=z-y}^{1} 1 \, dxdy$$

$$= 1 - \int_{y=z-1}^{1} (1 - z + y)dy = 1 - \frac{(2-z)^2}{2}, \quad 1 \le z < 2. \qquad (8\text{-}17)$$

Using (8-16) - (8-17), we obtain

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \begin{cases} z & 0 \le z < 1, \\ 2-z, & 1 \le z < 2. \end{cases} \tag{8-18}$$

By direct convolution of $f_X(x)$ and $f_Y(y)$, we obtain the same result as above. In fact, for $0 \le z < 1$ (Fig. 8.6(a))

$$f_Z(z) = \int f_X(z-x) f_Y(x) dx = \int_0^z 1 \; dx = z. \tag{8-19}$$

and for $1 \le z < 2$ (Fig. 8.6(b))

$$f_Z(z) = \int_{z-1}^1 1 \; dx = 2 - z. \tag{8-20}$$

Fig 8.6 (c) shows $f_Z(z)$ which agrees with the convolution of two rectangular waveforms as well.

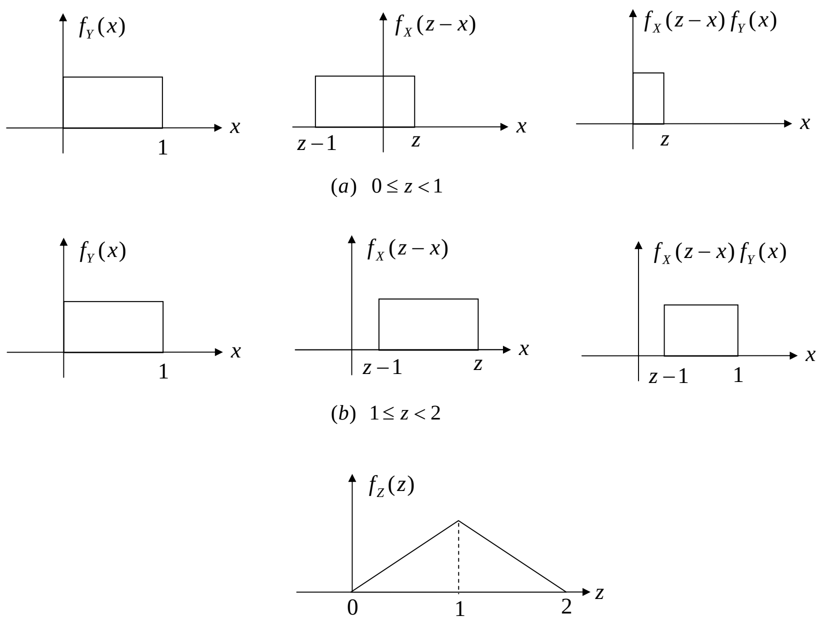


Fig. 8.6 (c)

Mean, Variance, Moments and Characteristic Functions

For a r.v *X*, its p.d.f $f_X(x)$ represents complete information about it, and for any Borel set *B* on the *x*-axis

$$P\big(X(\xi) \in B\big) = \int_B f_X(x)\,dx. \tag{6-1}$$

Note that $f_X(x)$ represents very detailed information, and quite often it is desirable to characterize the r.v in terms of its average behavior. In this context, we will introduce two parameters - mean and variance - that are universally used to represent the overall properties of the r.v and its p.d.f.

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

$$\eta_X = \overline{X} = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\,dx. \tag{6-2}$$

If *X* is a discrete-type r.v, then using (3-25) we get

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

$$= \sum_i x_i p_i = \sum_i x_i\, P(X = x_i). \tag{6-3}$$

Mean represents the average (mean) value of the r.v in a very large number of trials. For example if $X \sim U(a,b)$, then using (3-31) ,

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \tag{6-4}$$

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

Moments

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

47

$$\mu_n = E[(X-\mu)^n] \qquad \text{(6-25)}$$

are known as the central moments of *X*. Clearly, the mean $\mu = m_1$, and the variance $\sigma^2 = \mu_2$. It is easy to relate m_n and μ_n. Infact

$$\mu_n = E[(X-\mu)^n] = E\left(\sum_{k=0}^{n}\binom{n}{k}X^k(-\mu)^{n-k}\right)$$

$$= \sum_{k=0}^{n}\binom{n}{k}E\left(X^k\right)(-\mu)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}\, m_k\,(-\mu)^{n-k}. \qquad \text{(6-26)}$$

In general, the quantities

$$E[(X-a)^n] \qquad \text{(6-27)}$$

are known as the generalized moments of *X* about *a*, and

$$E[|X|^n] \qquad \text{(6-28)}$$

are known as the absolute moments of *X*.

48

PILLAI

$$\Phi_X(\omega) \overset{\Delta}{=} E\left(e^{jX\omega}\right) = \int_{-\infty}^{+\infty} e^{jx\omega} f_X(x)dx. \qquad (6\text{-}31)$$

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

For discrete r.vs the characteristic function reduces to

$$\Phi_X(\omega) = \sum_k e^{jk\omega} P(X = k). \qquad (6\text{-}32)$$

Thus for example, if $X \sim P(\lambda)$, then its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} e^{jk\omega} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{j\omega})^k}{k!} = e^{-\lambda} e^{\lambda e^{j\omega}} = e^{\lambda(e^{j\omega}-1)}. \qquad (6\text{-}33)$$

Similarly, if X is a binomial r.v , its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{n} e^{jk\omega} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^{j\omega})^k q^{n-k} = (pe^{j\omega} + q)^n. \qquad (6\text{-}34)$$

To illustrate the usefulness of the characteristic function of a r.v in computing its moments, first it is necessary to derive the relationship between them. Towards this, from (6-31)

$$\Phi_X(\omega) = E\left(e^{jX\omega}\right) = E\left[\sum_{k=0}^{\infty} \frac{(j\omega X)^k}{k!}\right] = \sum_{k=0}^{\infty} j^k \frac{E(X^k)}{k!} \omega^k$$

$$= 1 + jE(X)\omega + j^2 \frac{E(X^2)}{2!} \omega^2 + \cdots + j^k \frac{E(X^k)}{k!} \omega^k + \cdots. \qquad (6\text{-}35)$$

Taking the first derivative of (6-35) with respect to ω, and letting it to be equal to zero, we get

$$\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0} = jE(X) \quad \text{or} \quad E(X) = \frac{1}{j} \left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0}. \qquad (6\text{-}36)$$

Similarly, the second derivative of (6-35) gives

$$E(X^2) = \frac{1}{j^2} \left.\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2}\right|_{\omega=0}, \qquad (6\text{-}37)$$

Probability Distributions – Relevance to Networks 2

- Service times of queues (t_{trans}) in packet switching routers can be effectively modeled as exponential
- Arrival pattern of packets at a router is often Poisson in nature
  - and, arrival interval is exponential
- Central Limit Theorem: the distribution of a very large number of independent RVs is approximately normal, independent of individual distributions

51

Multiple Random Variables

- Independence:
  - $F(x,y) = F(x)F(y)$, and $f(x,y) = f(x)f(y)$
- Covariance:
  - $Cov(X,Y) = E[XY] - E[X]E[Y]$
- Correlation coefficient
  - $r(X,Y) = Cov(X,Y) / \sigma_X \sigma_Y$
    - positively correlated: $r(X,Y) > 0$
    - negatively correlated: $r(X,Y) < 0$
    - uncorrelated: $r(X,Y) = Cov(X,Y) = 0$

52

Stochastic (Random) Processes

- <u>Family of Random Variables</u>
  - {*x(t), t* ∈ *T*}, indexed by parameter t over index set *T*
  - index set is typically taken as time dimension
  - continuous- or discrete-time, *t*
  - continuous- or discrete-value, x(*t*)

14. Stochastic Processes

Introduction

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.

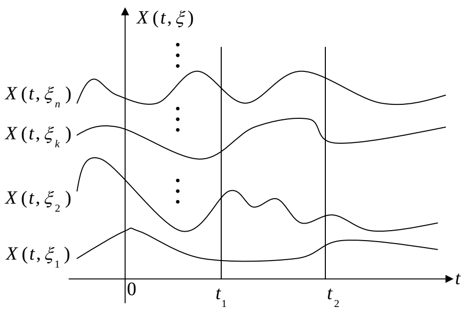


Fig. 14.1

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

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

where φ is a uniformly distributed random variable in $(0, 2\pi)$, represents a stochastic process. Stochastic processes are everywhere: Brownian motion, stock market fluctuations, various queuing systems all represent stochastic phenomena.

If $X(t)$ is a stochastic process, then for fixed t, $X(t)$ represents a random variable. Its distribution function is given by

$$F_x(x,t) = P\{X(t) \le x\} \qquad (14\text{-}1)$$

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

$$f_x(x,t) \triangleq \frac{dF_x(x,t)}{dx} \qquad (14\text{-}2)$$

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

$$F_x(x_1, x_2, t_1, t_2) = P\{X(t_1) \le x_1, X(t_2) \le x_2\} \qquad (14\text{-}3)$$

and

$$f_x(x_1, x_2, t_1, t_2) \triangleq \frac{\partial^2 F_x(x_1, x_2, t_1, t_2)}{\partial x_1 \, \partial x_2} \qquad (14\text{-}4)$$

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, $i = 1, 2, \cdots, n$ and for all n. (an almost impossible task in reality).

Mean of a Stochastic Process:

$$\mu(t) \triangleq E\{X(t)\} = \int_{-\infty}^{+\infty} x f_X(x,t)dx \qquad (14\text{-}5)$$

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

$$R_{XX}(t_1,t_2) \triangleq E\{X(t_1)X^*(t_2)\} = \int\int x_1 x_2^* f_X(x_1,x_2,t_1,t_2)dx_1dx_2 \quad (14\text{-}6)$$

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. $R_{XX}(t_1,t_2) = R_{XX}^*(t_2,t_1) = [E\{X(t_2)X^*(t_1)\}]^*$ (14-7)

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

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

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

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

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

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

Example 14.1

Let

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

Then

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

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

Example 14.2

$$X(t) = a\cos(\omega_0 t + \varphi), \qquad \varphi \sim U(0, 2\pi). \qquad (14\text{-}11)$$

This gives

$$\mu_x(t) = E\{X(t)\} = aE\{\cos(\omega_0 t + \varphi)\}$$
$$= a\cos\omega_0 t\, E\{\cos\varphi\} - a\sin\omega_0 t\, E\{\sin\varphi\} = 0, \qquad (14\text{-}12)$$

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

Similarly

$$R_{xx}(t_1, t_2) = a^2 E\{\cos(\omega_0 t_1 + \varphi)\cos(\omega_0 t_2 + \varphi)\}$$
$$= \frac{a^2}{2} E\{\cos\omega_0(t_1 - t_2) + \cos(\omega_0(t_1 + t_2) + 2\varphi)\}$$
$$= \frac{a^2}{2}\cos\omega_0(t_1 - t_2). \qquad (14\text{-}13)$$

Brownian Motion Processes

- Stochastic process that describes Random Movement of particles
  - Let B(t) denote displacement in one dimension after time t
  - Let B(t) – B(s) denote net movement over time interval (s,t)
  - Then, B(t) – B(s) has normal distribution
- Brownian probability density function:

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