B. Review of Probability

B.1. Probability

B.1.1. A. Random Experiments:

In the study of probability, any process of observation is referred to as an *experiment*. The results of an observation are called the *outcomes* of the experiment. An experiment is called a *random experiment* if its outcome cannot be predicted. Typical examples of a random experiment are the roll of a die, the toss of a coin, drawing a card from a deck, or selecting a message signal for transmission from several messages.

B.1.2. B. Sample Space and Events:

The set of all possible outcomes of a random experiment is called the *sample space* S. An element in S is called a *sample point*. Each outcome of a random experiment corresponds to a sample point.

A set A is called a *subset* of B, denoted by $A \subset B$ if every element of A is also an element of B. Any subset of the sample space S is called an *event*. A sample point of S is often referred to as an *elementary event*. Note that the sample space S is the subset of itself, that is, $S \subset S$. Since S is the set of all possible outcomes, it is often called the *certain event*.

B.1.3. C. Algebra of Events:

1. The *complement* of event A, denoted $\bar{A}$, is the event containing all sample points in S but not in A.
2. The *union* of events A and B, denoted $A \cup B$, is the event containing all sample points in either A or B or both.
3. The *intersection* of events A and B, denoted $A \cap B$, is the event containing all sample points in both A and B.
4. The event containing no sample point is called the *null event*, denoted Ø. Thus Ø corresponds to an impossible event.
5. Two events A and B are called *mutually exclusive* or *disjoint* if they contain no common sample point, that is, $A \cap B = \varnothing$.

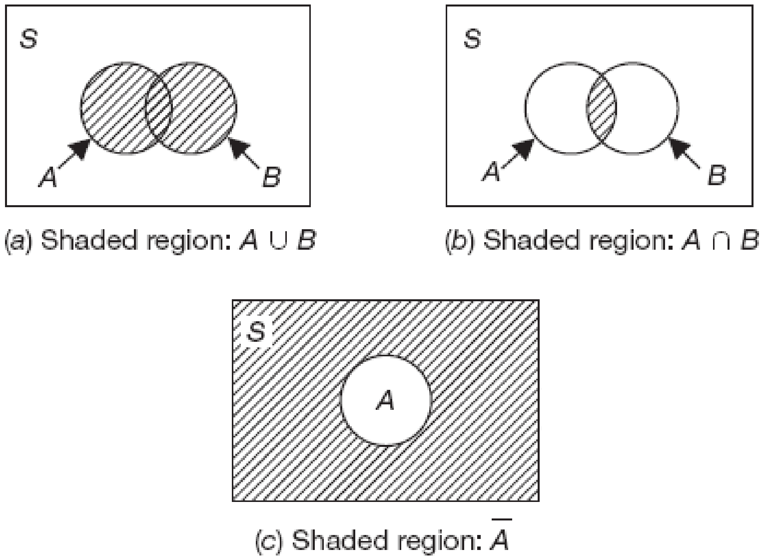
By the preceding set of definitions, we obtain the following identities:

$$\bar{S} = \varnothing \quad \bar{\varnothing} = S$$
$$S \cup A = S \quad S \cap A = A$$
$$A \cup \bar{A} = S \quad A \cap \bar{A} = \varnothing \quad \bar{\bar{A}} = A$$

B.1.4. D. Venn Diagram:

A graphical representation that is very useful for illustrating set operation is the Venn diagram. For instance, in the three Venn diagrams shown in Fig. B-1, the shaded areas represent, respectively, the events $A \cup B$, $A \cap B$, and $\bar{A}$.

Figure B-1



B.1.5. E. Probabilities of Events:

An assignment of real numbers to the events defined on S is known as the *probability measure*. In the *axiomatic* definition, the probability $P(A)$ of the event A is a real number assigned to A that satisfies the following three *axioms*:

$$\text{Axiom 1:} \qquad P(A) \geq 0$$

(B.1)

$$\text{Axiom 2:} \qquad P(S) = 1$$

(B.2)

$$\text{Axiom 3:} \qquad P(A \cup B) = P(A) + P(B) \qquad \text{if } A \cap B = \varnothing$$

(B.3)

With the preceding axioms, the following useful properties of probability can be obtained.

$$1. \quad P(\bar{A}) = 1 - P(A)$$

(B.4)

$$2. \quad P(\varnothing) = 0$$

(B.5)

$$3. \quad P(A) \leq P(B) \qquad \text{if } A \subset B$$

(B.6)

$$4. \quad P(A) \leq 1$$

(B.7)

$$5. \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(B.8)

Note that Property 4 can be easily derived from axiom 2 and property 3. Since $A \subset S$, we have

$$P(A) \leq P(S) = 1$$

Thus, combining with axiom 1, we obtain

$$0 \leq P(A) \leq 1$$

(B.9)

Property 5 implies that

$$P(A \cup B) \leq P(A) + P(B)$$

(B.10)

since $P(A \cap B) \geq 0$ by axiom 1.

One can also define $P(A)$ intuitively, in terms of relative frequency. Suppose that a random experiment is repeated n times. If an event A occurs n_A times, then its probability $P(A)$ is defined as

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

(B.11)

Note that this limit may not exist.

EXAMPLE B.1 Using the axioms of probability, prove Eq. (B.4).

$$S = A \cup \bar{A} \qquad \text{and} \qquad A \cap \bar{A} = \varnothing$$

Then the use of axioms 1 and 3 yields

$$P(S) = 1 = P(A) + P(\bar{A})$$

Thus

$$P(\bar{A}) = 1 - P(A)$$

EXAMPLE B.2 Verify Eq. (B.5).

$$A = A \cup \varnothing \qquad \text{and} \qquad A \cap \varnothing = \varnothing$$

Therefore, by axiom 3,

$$P(A) = P(A \cup \varnothing) = P(A) + P(\varnothing)$$

and we conclude that

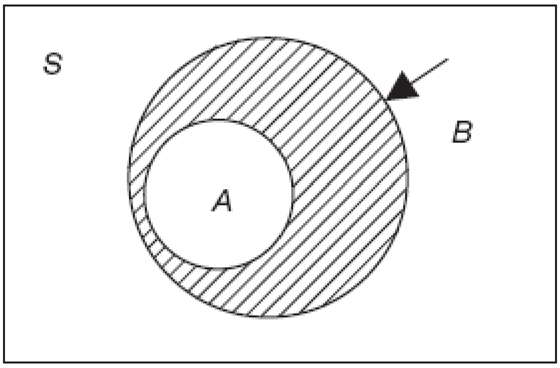
$$P(\varnothing) = 0$$

EXAMPLE B.3 Verify Eq. (B.6).

Let $A \subset B$. Then from the Venn diagram shown in Fig. B-2, we see that

$$B = A \cup (B \cap \bar{A}) \quad \text{and} \quad A \cap (B \cap \bar{A}) = \varnothing$$

Figure B-2



Hence, from axiom 3,

$$P(B) = P(A) + P(B \cap \bar{A}) \geq P(A)$$

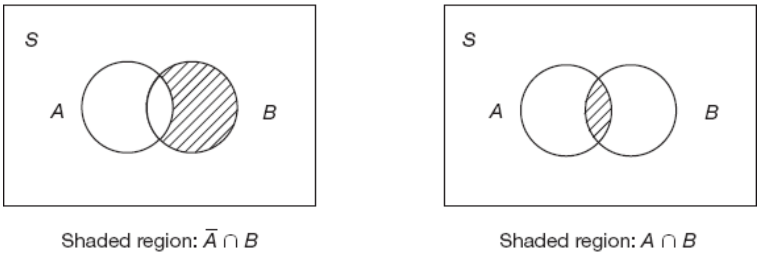
because by axiom 1, $P(B \cap -A) \geq 0$.

EXAMPLE B.4 Verify Eq. (B.8).

From the Venn diagram of Fig. B-3, each of the sets $A \cup B$ and B can be expressed, respectively, as a union of mutually exclusive sets as follows:

$$A \cup B = A \cup (\bar{A} \cap B) \qquad \text{and} \qquad B = (A \cap B) \cup (\bar{A} \cap B)$$

Figure B-3



Thus, by axiom 3,

$$P(A \cup B) = P(A) + P(\bar{A} \cap B)$$

(B.12)

and

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

(B.13)

From Eq. (B.13) we have

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

(B.14)

Substituting Eq. (B.14) into Eq. (B.12), we obtain

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

B.1.6. F. Equally Likely Events:

Consider a finite sample space *S* with finite elements

$$S = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$$

where λ_i's are elementary events. Let $P(\lambda_i) = p_i$. Then

1. $0 \le p_i \le 1 \quad i = 1, 2, \ldots, n$

$$\sum_{i=1}^{n} p_i = p_1 + p_2 + \cdots + p_n = 1$$

(B.15)

3. If $A = \bigcup_{i \in I} \lambda_i$, where *I* is a collection of subscripts, then

$$P(A) = \sum_{\lambda_i \in A} p(\lambda_i) = \sum_{i \in I} p_i$$

(B.16)

When all elementary events λ_i ($i = 1, 2, \ldots, n$) are equally likely events, that is

$$p_1 = p_2 = \ldots = p_n$$

then from Eq. (B.15), we have

$$p_i = \frac{1}{n} \quad i = 1, 2, \ldots, n$$

(B.17)

and

$$P(A) = \frac{n(A)}{n}$$

(B.18)

where $n(A)$ is the number of outcomes belonging to event *A* and *n* is the number of sample points in *S*.

B.1.7. G. Conditional Probability:

The *conditional probability* of an event *A* given the event *B*, denoted by $P(A|B)$, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(B) > 0$$

(B.19)

where $P(A \cap B)$ is the joint probability of A and B. Similarly,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad P(A) > 0$$

(B.20)

is the conditional probability of an event B given event A. From Eqs. (B.19) and (B.20) we have

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

(B.21)

Equation (B.21) is often quite useful in computing the joint probability of events.

From Eq. (B.21) we can obtain the following *Bayes rule*:

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

(B.22)

EXAMPLE B.5 Find $P(A|B)$ if (*a*) $A \cap B = \varnothing$, (*b*) $A \subset B$, and (*c*) $B \subset A$.

a. If $A \cap B = \varnothing$, then $P(A \cap B) = P(\varnothing) = 0$. Thus,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\varnothing)}{P(B)} = 0$$

b. If $A \subset B$, then $A \cap B = A$ and

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

c. If $B \subset A$, then $A \cap B = B$ and

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

B.1.8. H. Independent Events:

Two events A and B are said to be *(statistically) independent* if

$$P(A|B) = P(A) \qquad \text{and} \qquad P(B|A) = P(B)$$

(B.23)

This, together with Eq. (B.21), implies that for two statistically independent events

$$P(A \cap B) = P(A)P(B)$$

(B.24)

We may also extend the definition of independence to more than two events. The events $A_1, A_2, \ldots, A_n$ are independent if and only if for every subset $\{A_{i1}, A_{i2}, \ldots, A_{ik}\}$ $(2 \le k \le n)$ of these events,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k})$$

(B.25)

B.1.9. I. Total Probability:

The events $A_1, A_2, \ldots, A_n$ are called *mutually exclusive* and *exhaustive* if

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n = S \quad \text{and} \quad A_i \cap A_j = \varnothing \quad i \neq j$$

(B.26)

Let B be any event in S. Then

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

(B.27)

which is known as the *total probability* of event B. Let $A = A_i$ in Eq. (B.22); using Eq. (B.27) we obtain

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

(B.28)

Note that the terms on the right-hand side are all conditioned on events A_i, while that on the left is conditioned on B. Equation (B.28) is sometimes referred to as *Bayes' theorem.*

EXAMPLE B.6 Verify Eq. (B.27).

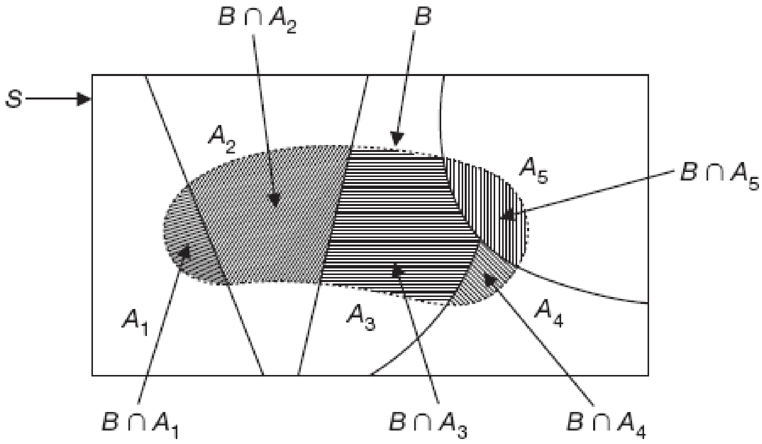
Since $B \cap S = B$ [and using Eq. (B.26)], we have

$$\begin{aligned} B &= B \cap S = B \cap (A_1 \cup A_2 \cup \cdots \cup A_N) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_N) \end{aligned}$$

Now the events $B \cap A_k$ $(k = 1, 2, \ldots, N)$ are mutually exclusive, as seen from the Venn diagram of Fig. B-4. Then by axiom 3 of the probability definition and Eq. (B.21), we obtain

$$P(B) = P(B \cap S) = \sum_{k=1}^{N} P(B \cap A_k) = \sum_{k=1}^{N} P(B|A_k)P(A_k)$$

Figure B-4

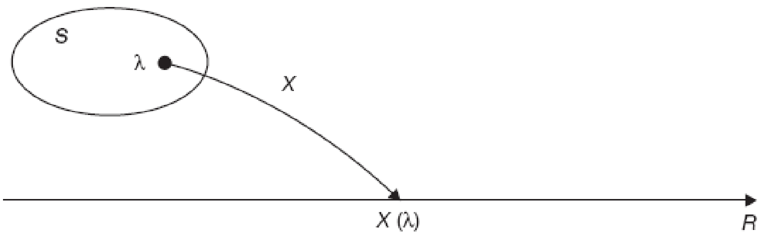


B.2. Random Variables

B.2.1. A. Random Variables:

Consider a random experiment with sample space S. A *random variable* $X(\lambda)$ is a single-valued real function that assigns a real number called the *value* of $X(\lambda)$ to each sample point λ of S. Often we use a single letter X for this function in place of $X(\lambda)$ and use r.v. to denote the random variable. A schematic diagram representing a r.v. is given in Fig. B-5.

Figure B-5 *Random variable X as a function.*



The sample space S is termed the *domain* of the r.v. X, and the collection of all numbers [values of $X(\lambda)$] is termed the *range* of the r.v. X. Thus, the range of X is a certain subset of the set of all real numbers and it is usually denoted by R_X. Note that two or more different sample points might give the same value of $X(\lambda)$, but two different numbers in the range cannot be assigned to the same sample point.

The r.v. X induces a probability measure on the real line as follows:

$$P(X = x) = P\{\lambda : X(\lambda) = x\}$$
$$P(X \le x) = P\{\lambda : X(\lambda) \le x\}$$
$$P(x_1 < X \le x_2) = P\{\lambda : x_1 < X(\lambda) \le x_2\}$$

If X can take on only a *countable* number of distinct values, then X is called a *discrete* random variable. If X can assume any values within one or more intervals on the real line, then X is called a *continuous* random variable. The number of telephone calls arriving at an office in a finite time is an example of a discrete random variable, and the exact time of arrival of a telephone call is an example of a continuous random variable.

B.2.2. B. Distribution Function:

The *distribution function* [or *cumulative distribution function* (cdf)] of *X* is the function defined by

$$F_X(x) = P(X \le x) \qquad -\infty < x < \infty$$

(B.29)

B.2.2.1. Properties of $F_X(x)$:

$$1. \quad 0 \le F_X(x) \le 1$$

(B.30)

$$2. \quad F_X(x_1) \le F_X(x_2) \qquad \text{if } x_1 < x_2$$

(B.31)

$$3. \quad F_X(\infty) = 1$$

(B.32)

$$4. \quad F_X(-\infty) = 0$$

(B.33)

$$5. \quad F_X(a^+) = F_X(a) \quad a^+ = \lim_{0<\varepsilon\to 0} a + \varepsilon$$

(B.34)

From definition (B.29) we can compute other probabilities:

$$P(a < X \le b) = F_X(b) - F_X(a)$$

(B.35)

$$P(X > a) = 1 - F_X(a)$$

(B.36)

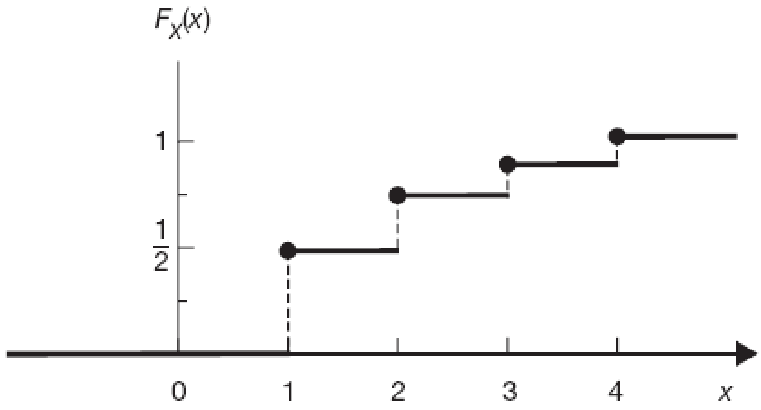
$$P(X < b) = F_X(b^-) \quad b^- = \lim_{0<\varepsilon\to 0} b - \varepsilon$$

(B.37)

B.2.3. C. Discrete Random Variables and Probability Mass Functions:

Let *X* be a discrete r.v. with cdf $F_X(x)$. Then $F_X(x)$ is a staircase function (see Fig. B-6), and $F_X(x)$ changes values only in jumps (at most a countable number of them) and is constant between jumps.

Figure B-6



Suppose that the jumps in $F_X(x)$ of a discrete r.v. X occur at the points $x_1, x_2, \ldots$, where the sequence may be either finite or countably infinite, and we assume $x_i < x_j$ if $i < j$. Then

$$F_X(x_i) - F_X(x_{i-1}) = P(X \le x_i) - P(X \le x_{i-1}) = P(X = x_i)$$

(B.38)

Let

$$p_X(x) = P(X = x)$$

(B.39)

The function $p_X(x)$ is called the *probability mass function* (pmf) of the discrete r.v. X.

B.2.3.1. Properties of $p_X(x)$:

$$1. \quad 0 \le p_X(x_i) \le 1 \qquad i = 1, 2, \ldots$$

(B.40)

$$2. \quad p_X(x) = 0 \qquad \text{if } x \ne x_i (i = 1, 2, \ldots)$$

(B.41)

$$3. \quad \sum_i p_X(x_i) = 1$$

(B.42)

The cdf $F_X(x)$ of a discrete r.v. X can be obtained by

$$F_X(x) = P(X \le x) = \sum_{x_i \le x} p_X(x_i)$$

(B.43)

B.2.4. D. Examples of Discrete Random Variables:

B.2.4.1. 1. Bernoulli Distribution:

A r.v. X is called a *Bernoulli* r.v. with parameter p if its pmf is given by

$$p_X(k) = P(X = k) = p^k(1 - p)^{1-k} \qquad k = 0, 1$$

(B.44)

where $0 \le p \le 1$. By Eq. (B.29), the cdf $F_X(x)$ of the Bernoulli r.v. X is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

(B.45)

B.2.4.2. 2. Binomial Distribution:

A r.v. X is called a *binomial* r.v. with parameters (n, p) if its pmf is given by

$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \qquad k = 0, 1, \ldots, n$$

(B.46)

where $0 \le p \le 1$ and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

which is known as the binomial coefficient. The corresponding cdf of X is

$$F_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} \qquad n \le x < n + 1$$

(B.47)

B.2.4.3. 3. Poisson Distribution:

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

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

(B.48)

The corresponding cdf of X is

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

(B.49)

B.2.5. E. Continuous Random Variables and Probability Density Functions:

Let X be a r.v. with cdf $F_X(x)$. Then $F_X(x)$ is continuous and also has a derivative $dF_X(x)/dx$ that exists everywhere except at possibly a finite number of points and is piecewise continuous. Thus, if X is a continuous r.v., then

$$P(X = x) = 0$$

(B.50)

In most applications, the r.v. is either discrete or continuous. But if the cdf $F_X(x)$ of a r.v. X possesses both features of discrete and continuous r.v.'s, then the r.v. X is called the *mixed* r.v.

Let

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

(B.51)

The function $f_X(x)$ is called the *probability density function* (pdf) of the continuous r.v. X.

B.2.5.1. Properties of $f_X(x)$:

$$1. \quad f_X(x) \ge 0$$

(B.52)

$$2. \quad \int_{-\infty}^{\infty} f_X(x)dx = 1$$

(B.53)

$$3. \quad f_X(x) \text{ is piecewise continuous.}$$

$$4. \quad P(a < X \le b) = \int_a^b f_X(x)\, dx$$

(B.54)

The cdf $F_X(x)$ of a continuous r.v. X can be obtained by

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(\xi)\, d\xi$$

(B.55)

B.2.6. F. Examples of Continuous Random Variables:

B.2.6.1. 1. Uniform Distribution:

A r.v. *X* is called a *uniform* r.v. over (*a*, *b*) if its pdf is given by

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

(B.56)

The corresponding cdf of *X* is

$$F_X(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$

(B.57)

B.2.6.2. 2. Exponential Distribution:

A r.v. *X* is called an *exponential* r.v. with parameter λ (> 0) if its pdf is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$$

(B.58)

The corresponding cdf of *X* is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

(B.59)

B.2.6.3. 3. Normal (or Gaussian) Distribution:

A r.v. $X = N(\mu; \sigma^2)$ is called a *normal* (or *Gaussian*) r.v. if its pdf is given by

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

(B.60)

The corresponding cdf of *X* is

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-(\xi-\mu)^2/(2\sigma^2)}\, d\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)t} e^{-\xi^2/2}\, d\xi$$

(B.61)

B.3. Two-Dimensional Random Variables

B.3.1. A. Joint Distribution Function:

Let S be the sample space of a random experiment. Let X and Y be two r.v.'s defined on S. Then the pair (X, Y) is called a two-dimensional r.v. if each of X and Y associates a real number with every element of S. The *joint cumulative distribution* function (or joint cdf) of X and Y, denoted by $F_{XY}(x,y)$, is the function defined by

$$F_{XY}(x, y) = P(X \le x, Y \le y)$$

(B.62)

Two r.v.'s X and Y will be called *independent* if

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

(B.63)

for every value of x and y.

B.3.2. B. Marginal Distribution Function:

Since $\{X \le \infty\}$ and $\{Y \le \infty\}$ are certain events, we have

$$\{X \le x, Y \le \infty\} = \{X \le x\} \qquad \{X \le \infty, Y \le y\} = \{Y \le y\}$$

so that

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

(B.64)

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

(B.65)

The cdf's $F_X(x)$ and $F_Y(y)$, when obtained by Eqs. (B.64) and (B.65), are referred to as the *marginal* cdf's of X and Y, respectively.

B.3.3. C. Joint Probability Mass Functions:

Let (X, Y) be a discrete two-dimensional r.v. and (X, Y) takes on the values (x_i, y_j) for a certain allowable set of integers i and j. Let

$$p_{XY}(x_i, y_j) = P(X = x_i, Y = y_j)$$

(B.66)

The function $p_{XY}(x_i, y_j)$ is called the *joint probability mass function* (joint pmf) of (X, Y).

B.3.3.1. Properties of $p_{XY}(x_i,y_j)$:

$$1.\quad 0 \le p_{XY}(x_i, y_j) \le 1$$

(B.67)

$$2.\quad \sum_{x_i}\sum_{y_j} p_{XY}(x_i, y_j) = 1$$

(B.68)

The joint cdf of a discrete two-dimensional r.v. (X, Y) is given by

$$F_{XY}(x, y) = \sum_{x_i \le x}\sum_{y_j \le y} p_{XY}(x_i, y_j)$$

(B.69)

B.3.4. D. Marginal Probability Mass Functions:

Suppose that for a fixed value $X = x_i$, the r.v. Y can only take on the possible values y_j (j = 1, 2, …,n).

Then

$$p_X(x_i) = \sum_{y_j} p_{XY}(x_i, y_j)$$

(B.70)

Similarly,

$$p_Y(y_j) = \sum_{x_j} p_{XY}(x_i, y_j)$$

(B.71)

The pmf's $p_X(x_i)$ and $p_Y(y_j)$, when obtained by Eqs. (B.70) and (B.71), are referred to as the *marginal* pmf's of X and Y, respectively. If X and Y are independent r.v.'s, then

$$p_{XY}(x_i, y_j) = p_X(x_i)p_Y(y_j)$$

(B.72)

B.3.5. E. Joint Probability Density Functions:

Let (X, Y) be a continuous two-dimensional r.v. with cdf $F_{XY}(x, y)$ and let

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

(B.73)

The function $f_{XY}(x, y)$ is called the *joint probability density function* (joint pdf) of (X, Y). By integrating Eq. (B.73), we have

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(\xi, \eta)\, d\xi\, d\eta$$

(B.74)

B.3.5.1. Properties of $f_{XY}(x, y)$:

$$1. \quad f_{XY}(x, y) \geq 0$$

(B.75)

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

(B.76)

B.3.6. F. Marginal Probability Density Functions:

By Eqs. (B.64), (B.65), and definition (B.51), we obtain

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

(B.77)

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

(B.78)

The pdf's $f_X(x)$ and $f_Y(x)$, when obtained by Eqs. (B.77) and (B.78), are referred to as the *margina*l pdf's of X and Y, respectively. If X and Y are independent r.v.'s, then

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

(B.79)

The conditional pdf of X given the event $\{Y = y\}$ is

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

(B.80)

where $f_Y(y)$ is the marginal pdf of Y.

B.4. Functions of Random Variables

B.4.1. A. Random Variable $g(X)$:

Given a r.v. X and a function $g(x)$, the expression

$$Y = g(X)$$

(B.81)

defines a new r.v. Y. With y a given number, we denote D_y the subset of R_X (range of X) such that $g(x) \leq y$. Then

$$(Y \leq y) = [g(X) \leq y] = (X \in D_y)$$

where $(X \in D_y)$ is the event consisting of all outcomes λ such that the point $X(\lambda) \in D_y$. Hence,

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P(X \in D_y)$$

(B.82)

If X is a continuous r.v. with pdf $f_X(x)$, then

$$F_Y(y) = \int_{Dy} f_X(x)\, dx$$

(B.83)

Determination of $f_Y(y)$ from $f_X(x)$:

Let X be a continuous r.v. with pdf $f_X(x)$. If the transformation $y = g(x)$ is one-to-one and has the inverse transformation

$$x = g^{-1}(y) = h(y)$$

(B.84)

then the pdf of Y is given by

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X[h(y)]\left|\frac{dh(y)}{dy}\right|$$

(B.85)

Note that if $g(x)$ is a continuous monotonic increasing or decreasing function, then the transformation $y = g(x)$ is one-to-one. If the transformation $y = g(x)$ is not one-to-one, $f_Y(y)$ is obtained as follows:

Denoting the real roots of $y = g(x)$ by x_k, that is,

$$y = g(x_1) = \ldots = g(x_k) = \ldots$$

then

$$f_Y(y) = \sum_k \frac{f_X(x_k)}{|g'(x_k)|}$$

(B.86)

where $g'(x)$ is the derivative of $g(x)$.

EXAMPLE B.7 Let $Y = aX + b$. Show that if $X = N(\mu; \sigma^2)$, then $Y = N(a\mu + b; a^2\sigma^2)$.

The equation $y = g(x) = ax + b$ has a single solution $x_1 = (y - b)/a$, and $g'(x) = a$. The range of y is $(-\infty, \infty)$. Hence, by Eq. (B.86)

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

(B.87)

Since $X = N(\mu; \sigma^2)$, by Eq. (B.60)

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

(B.88)

Hence, by Eq. (B.87)

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}|a|\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{y-b}{a}-\mu\right)^2\right] \\ &= \frac{1}{\sqrt{2\pi}|a|\sigma} \exp\left[-\frac{1}{2a^2\sigma^2}(y-a\mu-b)^2\right] \end{aligned}$$

(B.89)

which is the pdf of $N(a\mu + b; a^2\sigma^2)$. Thus, if $X = N(\mu; \sigma^2)$, then $Y = N(a\mu + b; a^2\sigma^2)$.

EXAMPLE B.8 Let $Y = X^2$. Find $f_Y(y)$ if $X = N(0; 1)$.

If $y < 0$, then the equation $y = x^2$ has no real solutions; hence, $f_Y(y) = 0$.

If $y > 0$, then $y = x^2$ has two solutions

$$x_1 = \sqrt{y} \quad x_2 = -\sqrt{y}$$

Now, $y = g(x) = x^2$ and $g'(x) = 2x$. Hence, by Eq. (B.86)

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right]u(y)$$

(B.90)

Since $X = N(0; 1)$ from Eq. (B.60), we have

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

(B.91)

Since $f_X(x)$ is an even function from Eq. (B.90), we have

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X\left(\sqrt{y}\right)u(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}u(y)$$

(B.92)

B.4.2. B. One Function of Two Random Variables:

Given two random variables X and Y and a function $g(x, y)$, the expression

$$Z = g(X, Y)$$

(B.93)

is a new random variable. With z a given number, we denote by D_z the region of the xy plane such that $g(x, y) \le z$. Then

$$[Z \le z] = \{g(X, Y) \le z\} = \{(X, Y) \in D_z\}$$

where $\{ (X, Y) \in D_z\}$ is the event consisting of all outcomes λ such that the point $\{X(\lambda), Y(\lambda)\}$ is in D_z.

Hence,

$$F_Z(z) = P(Z \le z) = P\{(X, Y) \in D_z\}$$

(B.94)

If X and Y are continuous r.v.'s with joint pdf $f_{XY}(x, y)$, then

$$f_Z(z) = \iint_{D_Z} f_{XY}(x, y)\, dx\, dy$$

(B.95)

EXAMPLE B.9 Consider two r.v.'s X and Y with joint pdf $f_{XY}(x, y)$. Let $Z = X + Y$.

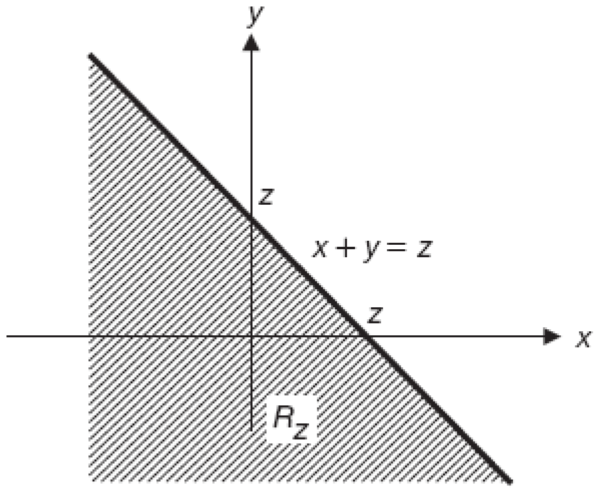
a. Determine the pdf of Z.
b. Determine the pdf of Z if X and Y are independent.

a. The range R_Z of Z corresponding to the event $(Z \le z) = (X + Y \le z)$ is the set of points (x, y) which lie on and to the left of the line $z = x + y$ (Fig. B-7). Thus, we have

$$F_Z(z) = P(X + Y \le z) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f_{XY}(x, y)\, dy \right] dx$$

(B.96)

Figure B-7



Then

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} \left[\frac{d}{dz} \int_{-\infty}^{z-x} f_{XY}(x, y)\, dy \right] dx$$
$$= \int_{-\infty}^{\infty} f_{XY}(x, z-x)\, dx$$

(B.97)

b. If X and Y are independent, then Eq. (B.97) reduces to

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

(B.98)

The integral on the right-hand side of Eq. (B.98) is known as a *convolution* of $f_X(z)$ and $f_Y(z)$. Since the convolution is commutative, Eq. (B.98) can also be written as

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y)\, f_X(z-y)\, dy$$

(B.99)

B.4.3. C. Two Functions of Two Random Variables:

Given two r.v.'s. X and Y and two functions $g(x, y)$ and $h(x, y)$, the expression

$$Z = g(X, Y) \qquad W = h(X, Y)$$

(B.100)

defines two new r.v.'s Z and W. With z and w two given numbers we denote D_{zw} the subset of R_{XY} [range of (X, Y)] such that $g(x, y) \le z$ and $h(x, y) \le w$. Then

$$(Z \le z, W \le w) = [g(x, y) \le z, h(x, y) \le w] = \{(X, Y) \in D_{zw}\}$$

where $\{(X, Y) \in D_{zw}\}$ is the event consisting of all outcomes λ such that the point $\{X(\lambda), Y(\lambda)\} \in D_{zw}$.

Hence,

$$F_{ZW}(z, w) = P(Z \le z, W \le w) = P\{(X, Y) \in D_{zw}\}$$

(B.101)

In the continuous case we have

$$f_{ZW}(z, w) = \int_{D_{ZW}}\int f_{XY}(x, y)\, dx\, dy$$

(B.102)

Determination of $f_{ZW}(z, w)$ from $f_{XY}(x, y)$:

Let X and Y be two continuous r.v.'s with joint pdf $f_{XY}(x, y)$. If the transformation

$$z = g(x, y) \qquad w = h(x, y)$$

(B.103)

is one-to-one and has the inverse transformation

$$x = q(z, w) \qquad y = r(z, w)$$

(B.104)

then the joint pdf of Z and W is given by

$$f_{ZW}(z, w) = f_{XY}(x, y)\,|J(x, y)|^{-1}$$

(B.105)

where $x = q(z, w)$, $y = r(z, w)$ and

$$J(x, y) = \begin{vmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \\ \dfrac{\partial h}{\partial x} & \dfrac{\partial h}{\partial y} \end{vmatrix} = \begin{vmatrix} \dfrac{\partial z}{\partial x} & \dfrac{\partial z}{\partial y} \\ \dfrac{\partial w}{\partial x} & \dfrac{\partial w}{\partial y} \end{vmatrix}$$

(B.106)

which is the Jacobian of the transformation (B.103).

EXAMPLE B.10 Consider the transformation

$$R = \sqrt{X^2 + Y^2} \quad \Theta = \tan^{-1}\frac{Y}{X}$$

$$\overline{J}(x, y) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin 0 \\ \sin\theta & r\cos 0 \end{vmatrix} = r$$

(B.107)

Eq. (B.105) yields

$$f_{R\Theta}(r, \theta) = r f_{XY}(r\cos\theta, r\sin\theta)$$

(B.108)

B.5. Statistical Averages

B.5.1. A. Expectation:

The *expectation* (or *mean*) of a r.v. *X*, denoted by *E*(*X*) or μ_X, is defined by

$$\mu_X = E(x) = \begin{cases} \sum_i x_i p_X(x_i) & X\text{: discrete} \\ \int_{-\infty}^{\infty} x f_X(x)\,dx & X\text{: continuous} \end{cases}$$

(B.109)

The expectation of $Y = g(X)$ is given by

$$E(Y) = E[g(X)] = \begin{cases} \sum_i g(x_i) p_X(x_i) & \text{(discrete case)} \\ \int_{-\infty}^{\infty} g(x) f_X(x)\,dx & \text{(continuous case)} \end{cases}$$

(B.110)

The expectation of $Z = g(X, Y)$ is given by

$$E(Z) = E[g(X, Y)] = \begin{cases} \sum_i \sum_j g(x_i, y_j) p_{XY}(x_i, y_j) & \text{(discrete case)} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\,dx\,dy & \text{(continuous case)} \end{cases}$$

(B.111)

Note that the *expectation operation* is linear, that is,

$$E[X + Y] = E[X] + E[Y]$$

(B.112)

$$E[cX] = cE[X]$$

(B.113)

where *c* is a constant.

EXAMPLE B.11 If *X* and *Y* are independent, then show that

$$E[XY] = E[X]E[Y]$$

(B.114)

and

$$E[g_1(X)g_2(X)] = E[g_1(X)]E[g_2(Y)]$$

(B.115)

If *X* and *Y* are independent, then by Eqs. (B.79) and (B.111) we have

$$\begin{aligned} E[X\,Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\,dxdy \\ &= \int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy = E[X]\,E[Y] \end{aligned}$$

Similarly,

$$\begin{aligned} E[g_1(X)g_2(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y) f_X(x)\, f_Y(y)\,dxdy \\ &= \int_{-\infty}^{\infty} g_1(x) f_X(x)\,dx \int_{-\infty}^{\infty} g_2(y) f_Y(y)\,dy = E[g_1(X)E[g_2(Y)] \end{aligned}$$

B.5.2. B. Moment:

The *n*th *moment* of a r.v. *X* is defined by

$$E(X^n) = \begin{cases} \sum_i x_i^n p_X(x_i) & X\text{: discrete} \\ \int_{-\infty}^{\infty} x^n f_X(x)\,dx & X\text{: continuous} \end{cases}$$

(B.116)

B.5.3. C. Variance:

The *variance* of a r.v. *X*, denoted by σ^2_X or Var(*X*), is defined by

$$\mathrm{Var}(X) = \sigma_X^2 = E\,[(X - \mu_X)^2]$$

(B.117)

Thus,

$$\sigma_X^2 = \begin{cases} \sum_i (x_i - \mu_X)^2 \, p_X(x_i) & X: \text{discrete} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx & X: \text{continuous} \end{cases}$$

(B.118)

The positive square root of the variance, or σ_X, is called the *standard deviation* of *X*. The variance or standard variation is a measure of the "spread" of the values of *X* from its mean μ_X. By using Eqs. (B.112) and (B.113), the expression in Eq. (B.117) can be simplified to

$$\sigma_X^2 = E[X^2] - \mu_X^2 = E[X^2] - (E[X])^2$$

(B.119)

Mean and variance of various random variables are tabulated in Table B-1.

Table B-1 Properties of the Fourier Transform

RANDOM VARIABLE X	MEAN μ_X	VARIANCE σ_X^2
Bernoulli (p)	p	p(1-p)
Binomial (n, p)	np	np(1 − p)
Poisson (λ)	λ	λ
Uniform (a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential (λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gaussion (normal)	μ	σ^2

B.5.4. D. Covariance and Correlation Coefficient:

The (*k*, *n*)th moment of a two-dimensional r.v. (*X*, *Y*) is defined by

$$m_{kn} = E(X^k Y^n) = \begin{cases} \sum_{y_j} \sum_{x_i} x_i^k y_j^n \, p_{XY}(x_i, y_j) & X: \text{discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^n f_{XY}(x, y)\, dx\, dy & X: \text{continuous} \end{cases}$$

(B.120)

The (1, 1)th joint moment of (*X*, *Y*),

$$m_{11} = E(X\,Y)$$

(B.121)

is called the *correlation* of X and Y. If $E(X\,Y) = 0$, then we say that X and Y are *orthogonal*. The *covariance* of X and Y, denoted by $\mathrm{Cov}(X, Y)$ or σ_{XY}, is defined by

$$\mathrm{Cov}(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

(B.122)

Expanding Eq. (B.122), we obtain

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

(B.123)

If $\mathrm{Cov}(X, Y) = 0$, then we say that X and Y are *uncorrelated*. From Eq. (B.123) we see that X and Y are uncorrelated if

$$E(X\,Y) = E(X)E(Y)$$

(B.124)

Note that if X and Y are independent, then it can be shown that they are uncorrelated. However, the converse is not true in general; that is, the fact that X and Y are uncorrelated does not, in general, imply that they are independent. The *correlation coefficient*, denoted by $\rho(X, Y)$ or ρ_{XY}, is defined by

$$\rho(X, Y) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

(B.125)

It can be shown that (Example B.15)

$$|\rho_{XY}| \le 1 \qquad \text{or} \qquad -1 \le \rho_{XY} \le 1$$

(B.126)

B.5.5. E. Some Inequalities:

B.5.5.1. 1. Markov Inequality:

If $f_X(x) = 0$ for $x < 0$, then for any $\alpha > 0$,

$$P(X \ge \alpha) \le \frac{\mu_X}{\alpha}$$

(B.127)

B.5.5.2. 2. Chebyshev Inequality:

For any $\epsilon > 0$, then

$$P(|X - \mu_X| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2}$$

(B.128)

where $\mu_X = E[X]$ and σ^2_X is the variance of X. This is known as the *Chebyshev inequality*.

B.5.5.3. 3. Cauchy-Schwarz Inequality:

Let X and Y be real random variables with finite second moments. Then

$$(E[XY])^2 \leq E[X^2]\, E[Y^2]$$

(B.129)

This is known as the *Cauchy-Schwarz inequality*.

EXAMPLE B.12 Verify Markov inequality, Eq. (B.127).

From Eq. (B.54)

$$P(X \geq \alpha) = \int_\alpha^\infty f_X(x)\, dx$$

Since $f_X(x) = 0$ for $x < 0$,

$$\mu_X = E[X] = \int_0^\infty x f_X(x)\, dx \geq \int_\alpha^\infty x f_X(x)\, dx \geq \alpha \int_\alpha^\infty f_X(x)\, dx$$

Hence,

$$\int_\alpha^\infty f_X(x)\, dx = P(X \geq \alpha) \leq \frac{\mu_X}{\alpha}$$

EXAMPLE B.13 Verify Chebyshev inequality, Eq. (B.128).

From Eq. (B.54)

$$P(|X - \mu_X| \geq \epsilon) = \int_{-\infty}^{\mu_X - \epsilon} f_X(x)\, dx + \int_{\mu_X + \varepsilon}^\infty f_X(x)\, dx = \int_{|x - \mu_X| \geq \epsilon} f_X(x)\, dx$$

By Eq. (B.118)

$$\sigma_X^2 = \int_{-\infty}^\infty (x - \mu_X)^2\, f_X(x)\, dx \geq \int_{|x - \mu_X| \geq \epsilon} (x - \mu_X)^2\, f_X(x)\, dx \geq \epsilon^2 \int_{|x - \mu_X| \geq \epsilon} f_X(x)\, dx$$

Hence,

$$\int_{|x - \mu_X| \geq \epsilon} f_X(x)\, dx \geq \frac{\sigma_X^2}{\epsilon^2}$$

or

$$P(|X-\mu_X| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2}$$

EXAMPLE B.14 Verify Cauchy-Schwarz inequality Eq. (B.129).

Because the mean-square value of a random variable can never be negative,

$$E[(X-\alpha Y)^2] \geq 0$$

for any value of α. Expanding this, we obtain

$$E[X^2] - 2\alpha E[XY] + \alpha^2 E[Y^2] \geq 0$$

Choose a value of α for which the left-hand side of this inequality is minimum

$$\alpha = \frac{E[XY]}{E[Y^2]}$$

which results in the inequality

$$E[X^2] - \frac{(E[XY])^2}{E[Y^2]} \geq 0$$

or

$$(E[XY])^2 \leq E[X^2]\, E[Y^2]$$

EXAMPLE B.15 Verify Eq. (B.126).

From the Cauchy-Schwarz inequality Eq. (B.129) we have

$$\{E[(X-\mu_X)(Y-\mu_Y)]\}^2 \leq E[(X-\mu_X)^2]\,E[(Y-\mu_Y)^2]$$

or

$$\sigma_{XY}^2 \leq \sigma_X^2\, \sigma_Y^2$$

Then

$$\rho_{XY}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} \leq 1$$

from which it follows that

$$|\rho_{XY}| \leq 1$$