CHAPTER 2

Probability and Random Variables

2.1 Introduction

At the start of Sec. 1.1.2, we had indicated that one of the possible ways of classifying the signals is: deterministic or random. By random we mean unpredictable; that is, in the case of a random signal, we cannot with certainty predict its future value, even if the entire past history of the signal is known. If the signal is of the deterministic type, no such uncertainty exists.

Consider the signal \( x(t) = A \cos(2\pi f_t t + \theta) \). If \( A, \theta \) and \( f_t \) are known, then (we are assuming them to be constants) we know the value of \( x(t) \) for all \( t \). (\( A, \theta \) and \( f_t \) can be calculated by observing the signal over a short period of time).

Now, assume that \( x(t) \) is the output of an oscillator with very poor frequency stability and calibration. Though, it was set to produce a sinusoid of frequency \( f = f_t \), frequency actually put out maybe \( f'_t \) where \( f'_t \in (f_t \pm \Delta f_t) \). Even this value may not remain constant and could vary with time. Then, observing the output of such a source over a long period of time would not be of much use in predicting the future values. We say that the source output varies in a random manner.

Another example of a random signal is the voltage at the terminals of a receiving antenna of a radio communication scheme. Even if the transmitted
(radio) signal is from a highly stable source, the voltage at the terminals of a receiving antenna varies in an unpredictable fashion. This is because the conditions of propagation of the radio waves are not under our control.

But randomness is the essence of communication. Communication theory involves the assumption that the transmitter is connected to a source, whose output, the receiver is not able to predict with certainty. If the students know ahead of time what is the teacher (source + transmitter) is going to say (and what jokes he is going to crack), then there is no need for the students (the receivers) to attend the class!

Although less obvious, it is also true that there is no communication problem unless the transmitted signal is disturbed during propagation or reception by unwanted (random) signals, usually termed as noise and interference. (We shall take up the statistical characterization of noise in Chapter 3.)

However, quite a few random signals, though their exact behavior is unpredictable, do exhibit statistical regularity. Consider again the reception of radio signals propagating through the atmosphere. Though it would be difficult to know the exact value of the voltage at the terminals of the receiving antenna at any given instant, we do find that the average values of the antenna output over two successive one minute intervals do not differ significantly. If the conditions of propagation do not change very much, it would be true of any two averages (over one minute) even if they are well spaced out in time. Consider even a simpler experiment, namely, that of tossing an unbiased coin (by a person without any magical powers). It is true that we do not know in advance whether the outcome on a particular toss would be a head or tail (otherwise, we stop tossing the coin at the start of a cricket match!). But, we know for sure that in a long sequence of tosses, about half of the outcomes would be heads (If this does not happen, we suspect either the coin or tosser (or both!)).
Statistical regularity of averages is an experimentally verifiable phenomenon in many cases involving random quantities. Hence, we are tempted to develop mathematical tools for the analysis and quantitative characterization of random signals. To be able to analyze random signals, we need to understand random variables. The resulting mathematical topics are: probability theory, random variables and random (stochastic) processes. In this chapter, we shall develop the probabilistic characterization of random variables. In chapter 3, we shall extend these concepts to the characterization of random processes.

2.2 Basics of Probability

We shall introduce some of the basic concepts of probability theory by defining some terminology relating to random experiments (i.e., experiments whose outcomes are not predictable).

2.2.1. Terminology

Def. 2.1: Outcome

The end result of an experiment. For example, if the experiment consists of throwing a die, the outcome would be anyone of the six faces, $F_1, ...., F_6$.

Def. 2.2: Random experiment

An experiment whose outcomes are not known in advance. (e.g. tossing a coin, throwing a die, measuring the noise voltage at the terminals of a resistor etc.)

Def. 2.3: Random event

A random event is an outcome or set of outcomes of a random experiment that share a common attribute. For example, considering the experiment of throwing a die, an event could be the 'face $F_1$' or 'even indexed faces' ($F_2, F_4, F_6$). We denote the events by upper case letters such as $A, B$ or $A_1, A_2$.....
Def. 2.4: Sample space

The sample space of a random experiment is a mathematical abstraction used to represent all possible outcomes of the experiment. We denote the sample space by \( S \).

Each outcome of the experiment is represented by a point in \( S \) and is called a sample point. We use \( s \) (with or without a subscript), to denote a sample point. An event on the sample space is represented by an appropriate collection of sample point(s).

Def. 2.5: Mutually exclusive (disjoint) events

Two events \( A \) and \( B \) are said to be mutually exclusive if they have no common elements (or outcomes). Hence if \( A \) and \( B \) are mutually exclusive, they cannot occur together.

Def. 2.6: Union of events

The union of two events \( A \) and \( B \), denoted \( A \cup B \), \( (A + B) \) or \( (A \text{ or } B) \) is the set of all outcomes which belong to \( A \) or \( B \) or both. This concept can be generalized to the union of more than two events.

Def. 2.7: Intersection of events

The intersection of two events, \( A \) and \( B \), is the set of all outcomes which belong to \( A \) as well as \( B \). The intersection of \( A \) and \( B \) is denoted by \( A \cap B \) or simply \( AB \). The intersection of \( A \) and \( B \) is also referred to as a joint event \( A \) and \( B \). This concept can be generalized to the case of intersection of three or more events.

Def. 2.8: Occurrence of an event

An event \( A \) of a random experiment is said to have occurred if the experiment terminates in an outcome that belongs to \( A \).
**Def. 2.9: Complement of an event**

The complement of an event $A$, denoted by $\bar{A}$, is the event containing all points in $S$ but not in $A$.

**Def. 2.10: Null event**

The null event, denoted $\phi$, is an event with no sample points. Thus $\phi = \overline{S}$ (note that if $A$ and $B$ are disjoint events, then $AB = \phi$ and vice versa).

### 2.2.2 Probability of an Event

The probability of an event has been defined in several ways. Two of the most popular definitions are: i) the *relative frequency* definition, and ii) the *classical* definition.

**Def. 2.11: The relative frequency definition:**

Suppose that a random experiment is repeated $n$ times. If the event $A$ occurs $n_A$ times, then the probability of $A$, denoted by $P(A)$, is defined as

$$P(A) = \lim_{n \to \infty} \left( \frac{n_A}{n} \right)$$  \hspace{1cm} (2.1)

$\left( \frac{n_A}{n} \right)$ represents the fraction of occurrence of $A$ in $n$ trials.

For small values of $n$, it is likely that $\left( \frac{n_A}{n} \right)$ will fluctuate quite badly. But as $n$ becomes larger and larger, we expect, $\left( \frac{n_A}{n} \right)$ to tend to a definite limiting value. For example, let the experiment be that of tossing a coin and $A$ the event 'outcome of a toss is Head'. If $n$ is the order of 100, $\left( \frac{n_A}{n} \right)$ may not deviate from
by more than, say ten percent and as \( n \) becomes larger and larger, we expect \( \left( \frac{n_A}{n} \right) \) to converge to \( \frac{1}{2} \).

**Def. 2.12: The classical definition:**

The relative frequency definition given above has empirical flavor. In the classical approach, the probability of the event \( A \) is found without experimentation. This is done by counting the total number \( N \) of the possible outcomes of the experiment. If \( N_A \) of those outcomes are favorable to the occurrence of the event \( A \), then

\[
P(A) = \frac{N_A}{N}
\]

where it is assumed that all outcomes are equally likely!

Whatever may the definition of probability, we require the probability measure (to the various events on the sample space) to obey the following postulates or axioms:

**P1)** \( P(A) \geq 0 \)  \hspace{2cm} (2.3a)

**P2)** \( P(S) = 1 \)  \hspace{2cm} (2.3b)

**P3)** \( (AB) = \phi \) , then \( P(A + B) = P(A) + P(B) \)  \hspace{2cm} (2.3c)

(Note that in Eq. 2.3(c), the symbol + is used to mean two different things; namely, to denote the union of \( A \) and \( B \) and to denote the addition of two real numbers). Using Eq. 2.3, it is possible for us to derive some additional relationships:

i) If \( AB \neq \phi \) , then \( P(A + B) = P(A) + P(B) - P(AB) \)  \hspace{2cm} (2.4)

ii) Let \( A_1, A_2, \ldots, A_n \) be random events such that:

a) \( A_i A_j = \phi \), for \( i \neq j \) and  \hspace{2cm} (2.5a)
b) \( A_1 + A_2 + \ldots + A_n = S \). \hspace{1cm} (2.5b)

Then, \( P(A) = P(AA_1) + P(AA_2) + \ldots + P(AA_n) \) \hspace{1cm} (2.6)

where \( A \) is any event on the sample space.

Note: \( A_1, A_2, \ldots, A_n \) are said to be mutually exclusive (Eq. 2.5a) and exhaustive (Eq. 2.5b).

iii) \( P(\overline{A}) = 1 - P(A) \) \hspace{1cm} (2.7)

The derivation of Eq. 2.4, 2.6 and 2.7 is left as an exercise.

A very useful concept in probability theory is that of conditional probability, denoted \( P(B | A) \); it represents the probability of \( B \) occurring, given that \( A \) has occurred. In a real world random experiment, it is quite likely that the occurrence of the event \( B \) is very much influenced by the occurrence of the event \( A \). To give a simple example, let a bowl contain 3 resistors and 1 capacitor. The occurrence of the event 'the capacitor on the second draw' is very much dependent on what has been drawn at the first instant. Such dependencies between the events is brought out using the notion of conditional probability.

The conditional probability \( P(B | A) \) can be written in terms of the joint probability \( P(AB) \) and the probability of the event \( P(A) \). This relation can be arrived at by using either the relative frequency definition of probability or the classical definition. Using the former, we have

\[
P(AB) = \lim_{n \to \infty} \left( \frac{n_{AB}}{n} \right)
\]

\[
P(A) = \lim_{n \to \infty} \left( \frac{n_A}{n} \right)
\]
where \( n_{AB} \) is the number of times \( AB \) occurs in \( n \) repetitions of the experiment.

As \( P(B \mid A) \) refers to the probability of \( B \) occurring, given that \( A \) has occurred, we have

**Def 2.13: Conditional Probability**

\[
P(B \mid A) = \lim_{n \to \infty} \frac{n_{AB}}{n_A} = \lim_{n \to \infty} \left( \frac{n_{AB}}{n} \right) = \frac{P(AB)}{P(A)}, P(A) \neq 0
\] (2.8a)

or \( P(AB) = P(B \mid A) \cdot P(A) \)

Interchanging the role of \( A \) and \( B \), we have

\[
P(A \mid B) = \frac{P(AB)}{P(B)}, P(B) \neq 0
\] (2.8b)

Eq. 2.8(a) and 2.8(b) can be written as

\[
P(AB) = P(B \mid A) \cdot P(A) = P(B) \cdot P(A \mid B)
\] (2.9)

In view of Eq. 2.9, we can also write Eq. 2.8(a) as

\[
P(B \mid A) = \frac{P(B) \cdot P(A \mid B)}{P(A)}, P(A) \neq 0
\] (2.10a)

Similarly

\[
P(A \mid B) = \frac{P(A) \cdot P(B \mid A)}{P(B)}, P(B) \neq 0
\] (2.10b)

Eq. 2.10(a) or 2.10(b) is one form of Bayes’ rule or Bayes’ theorem.

Eq. 2.9 expresses the probability of joint event \( AB \) in terms of conditional probability, say \( P(B \mid A) \) and the (unconditional) probability \( P(A) \). Similar relation can be derived for the joint probability of a joint event involving the intersection of three or more events. For example \( P(ABC) \) can be written as
Another useful probabilistic concept is that of **statistical independence**. Suppose the events $A$ and $B$ are such that

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

(2.13)

That is, knowledge of occurrence of $A$ tells no more about the probability of occurrence of $B$ than we knew without that knowledge. Then, the events $A$ and $B$ are said to be **statistically independent**. Alternatively, if $A$ and $B$ satisfy the Eq. 2.13, then

$$P(AB) = P(A)P(B)$$

(2.14)

Either Eq. 2.13 or 2.14 can be used to define the statistical independence of two events. Note that if $A$ and $B$ are independent, then $P(AB) = P(A)P(B)$, whereas if they are disjoint, then $P(AB) = 0$. The notion of statistical independence can be generalized to the case of more than two events. A set of $k$ events $A_1, A_2, \ldots, A_k$ are said to be statistically independent if and only if (iff) the probability of every intersection of $k$ or fewer events equal the product of the probabilities of its constituents. Thus three events $A, B, C$ are independent when

$$P(ABC) = P(AB)P(C | AB)$$

$$= P(A)P(B | A)P(C | AB)$$

(2.11)

**Exercise 2.1**

Let $A_1, A_2, \ldots, A_n$ be $n$ mutually exclusive and exhaustive events and $B$ is another event defined on the same space. Show that

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

(2.12)

Eq. 2.12 represents another form of **Bayes’ theorem**.
\[ P(AB) = P(A) P(B) \]
\[ P(AC) = P(A) P(C) \]
\[ P(BC) = P(B) P(C) \]
and \[ P(ABC) = P(A) P(B) P(C) \]

We shall illustrate some of the concepts introduced above with the help of two examples.

**Example 2.1**

Priya (P1) and Prasanna (P2), after seeing each other for some time (and after a few tiffs) decide to get married, much against the wishes of the parents on both the sides. They agree to meet at the office of registrar of marriages at 11:30 a.m. on the ensuing Friday (looks like they are not aware of *Rahu Kalam* or they don't care about it).

However, both are somewhat lacking in punctuality and their arrival times are equally likely to be anywhere in the interval 11 to 12 hrs on that day. Also arrival of one person is independent of the other. Unfortunately, both are also very short tempered and will wait only 10 min. before leaving in a huff never to meet again.

a) Picture the sample space

b) Let the event A stand for “P1 and P2 meet”. Mark this event on the sample space.

c) Find the probability that the lovebirds will get married and (hopefully) will live happily ever after.

a) The sample space is the rectangle, shown in Fig. 2.1(a).
b) The diagonal $OP$ represents the simultaneous arrival of Priya and Prasanna. Assuming that $P1$ arrives at $11: x$, meeting between $P1$ and $P2$ would take place if $P2$ arrives within the interval $a$ to $b$, as shown in the figure. The event $A$, indicating the possibility of $P1$ and $P2$ meeting, is shown in Fig. 2.1(b).
Example 2.2:

Let two honest coins, marked 1 and 2, be tossed together. The four possible outcomes are $T_1T_2$, $T_1H_2$, $H_1T_2$, $H_1H_2$. ($T_1$ indicates toss of coin 1 resulting in tails; similarly $T_2$ etc.) We shall treat that all these outcomes are equally likely; that is the probability of occurrence of any of these four outcomes is $\frac{1}{4}$. (Treating each of these outcomes as an event, we find that these events are mutually exclusive and exhaustive). Let the event $A$ be 'not $H_1H_2$' and $B$ be the event 'match'. (Match comprises the two outcomes $T_1T_2$, $H_1H_2$). Find $P(B \mid A)$. Are $A$ and $B$ independent?

We know that $P(B \mid A) = \frac{P(AB)}{P(A)}$.

$AB$ is the event 'not $H_1H_2$' and 'match'; i.e., it represents the outcome $T_1T_2$. Hence $P(AB) = \frac{1}{4}$. The event $A$ comprises of the outcomes ‘$T_1T_2$, $T_1H_2$ and $H_1T_2$’; therefore,

$$P(A) = \frac{3}{4}$$

$$P(B \mid A) = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Intuitively, the result $P(B \mid A) = \frac{1}{3}$ is satisfying because, given 'not $H_1H_2$' the toss would have resulted in anyone of the three other outcomes which can be
treated to be equally likely, namely $\frac{1}{3}$. This implies that the outcome $T_1T_2$ given 'not $H_1H_2$', has a probability of $\frac{1}{3}$.

As $P(B) = \frac{1}{2}$ and $P(B | A) = \frac{1}{3}$, $A$ and $B$ are dependent events.

2.3 Random Variables

Let us introduce a transformation or function, say $X$, whose domain is the sample space (of a random experiment) and whose range is in the real line; that is, to each $s_i \in S$, $X$ assigns a real number, $X(s_i)$, as shown in Fig.2.2.

Fig. 2.2: A mapping $X(\ )$ from $S$ to the real line.

The figure shows the transformation of a few individual sample points as well as the transformation of the event $A$, which falls on the real line segment $[a_1, a_2]$.

2.3.1 Distribution function:

Taking a specific case, let the random experiment be that of throwing a die. The six faces of the die can be treated as the six sample points in $S$; that is,
\( F_i = s_i, \ i = 1, 2, \ldots, 6 \). Let \( X(s_i) = i \). Once the transformation is induced, then the events on the sample space will become transformed into appropriate segments of the real line. Then we can enquire into the probabilities such as

\[
P\{s: X(s) < a_i\}
\]

\[
P\{s: b_i < X(s) \leq b_2\}
\]
or

\[
P\{s: X(s) = c\}
\]

These and other probabilities can be arrived at, provided we know the Distribution Function of \( X \), denoted by \( F_X(\ ) \) which is given by

\[
F_X(x) = P\{s: X(s) \leq x\} \quad (2.15)
\]

That is, \( F_X(x) \) is the probability of the event, comprising all those sample points which are transformed by \( X \) into real numbers less than or equal to \( x \). (Note that, for convenience, we use \( x \) as the argument of \( F_X(\ ) \). But, it could be any other symbol and we may use \( F_X(\alpha), F_X(\alpha_i) \) etc.) Evidently, \( F_X(\ ) \) is a function whose domain is the real line and whose range is the interval \([0, 1]\)).

As an example of a distribution function (also called Cumulative Distribution Function CDF), let \( S \) consist of four sample points, \( s_i \) to \( s_4 \), with each with sample point representing an event with the probabilities \( P(s_i) = \frac{1}{4}, P(s_2) = \frac{1}{8}, P(s_3) = \frac{1}{8} \) and \( P(s_4) = \frac{1}{2} \). If \( X(s_i) = i - 1.5, i = 1, 2, 3, 4 \), then the distribution function \( F_X(x) \), will be as shown in Fig. 2.3.
Fig. 2.3: An example of a distribution function

\( F_X(x) \) satisfies the following properties:

i) \( F_X(x) \geq 0, \quad -\infty < x < \infty \)

ii) \( F_X(-\infty) = 0 \)

iii) \( F_X(\infty) = 1 \)

iv) If \( a > b \), then \( F_X(a) - F_X(b) = P\{s: b < X(s) \leq a\} \)

v) If \( a > b \), then \( F_X(a) \geq F_X(b) \)

The first three properties follow from the fact that \( F_X(x) \) represents the probability and \( P(S) = 1 \). Properties iv) and v) follow from the fact

\[ \{s: X(s) \leq b\} \cup \{s: b < X(s) \leq a\} = \{s: X(s) \leq a\} \]

Referring to the Fig. 2.3, note that \( F_X(x) = 0 \) for \( x < -0.5 \) whereas \( F_X(-0.5) = \frac{1}{4} \). In other words, there is a discontinuity of \( \frac{1}{4} \) at the point \( x = -0.5 \). In general, there is a discontinuity in \( F_X \) of magnitude \( P_a \) at a point \( x = a \), if and only if

\[ P\{s: X(s) = a\} = P_a \quad \text{(2.16)} \]
The properties of the distribution function are summarized by saying that $F_x(\ )$ is monotonically non-decreasing, is continuous to the right at each point $x^1$, and has a step of size $P_a$ at point $a$ if and if Eq. 2.16 is satisfied.

Functions such as $X(\ )$ for which distribution functions exist are called Random Variables (RV). In other words, for any real $x$, $\{s: X(s) \leq x\}$ should be an event in the sample space with some assigned probability. (The term “random variable” is somewhat misleading because an RV is a well defined function from $S$ into the real line.) However, every transformation from $S$ into the real line need not be a random variable. For example, let $S$ consist of six sample points, $s_1$ to $s_6$. The only events that have been identified on the sample space are: $A = \{s_1, s_2\}$, $B = \{s_3, s_4, s_5\}$ and $C = \{s_6\}$ and their probabilities are $P(A) = \frac{2}{6}$, $P(B) = \frac{1}{2}$ and $P(C) = \frac{1}{6}$. We see that the probabilities for the various unions and intersections of $A$, $B$ and $C$ can be obtained.

Let the transformation $X$ be $X(s_i) = i$. Then the distribution function fails to exist because $P[s: 3.5 < x \leq 4.5] = P(s_4)$ is not known as $s_4$ is not an event on the sample space.

---

1 Let $x = a$. Consider, with $\Delta > 0$,

$$\lim_{\Delta \to 0} P[a < X(s) \leq a + \Delta] = \lim_{\Delta \to 0} [F_X(a + \Delta) - F_X(a)]$$

We intuitively feel that as $\Delta \to 0$, the limit of the set $\{s: a < X(s) \leq a + \Delta\}$ is the null set and can be proved to be so. Hence,

$$F_x(a_i) - F_x(a) = 0, \text{ where } a_i = \lim_{\Delta \to 0} (a + \Delta)$$

That is, $F_x(x)$ is continuous to the right.
2.3.2 Probability density function

Though the CDF is a very fundamental concept, in practice we find it more convenient to deal with Probability Density Function (PDF). The PDF, \( f_x(x) \) is defined as the derivative of the CDF; that is

\[
f_x(x) = \frac{dF_x(x)}{dx} \quad (2.17a)
\]

or

\[
F_x(x) = \int_{-\infty}^{x} f_x(\alpha) \, d\alpha \quad (2.17b)
\]

The distribution function may fail to have a continuous derivative at a point \( x = a \) for one of the two reasons:

i) the slope of the \( F_x(x) \) is discontinuous at \( x = a \)

ii) \( F_x(x) \) has a step discontinuity at \( x = a \)

The situation is illustrated in Fig. 2.4.
As can be seen from the figure, \( F_X(x) \) has a discontinuous slope at \( x = 1 \) and a step discontinuity at \( x = 2 \). In the first case, we resolve the ambiguity by taking \( f_x \) to be a derivative on the right. (Note that \( F_X(x) \) is continuous to the right.) The second case is taken care of by introducing the impulse in the probability domain. That is, if there is a discontinuity in \( F_X \) at \( x = a \) of magnitude \( P_a \), we include an impulse \( P_a \delta(x - a) \) in the PDF. For example, for the CDF shown in Fig. 2.3, the PDF will be,

\[
1/8 \quad 1/8 \quad 1/8 \quad 1/8
\]

\[
X_f(x) = \delta + \delta - \delta - \delta
\]

\[
\text{(2.18)}
\]

In Eq. 2.18, \( f_X(x) \) has an impulse of weight \( 1/8 \) at \( x = 1/2 \) as \( P\left[X = \frac{1}{2}\right] = \frac{1}{8} \). This impulse function cannot be taken as the limiting case of an even function (such as \( \frac{1}{\varepsilon} \) \( g_a\left(\frac{x}{\varepsilon}\right) \)) because,

\[
\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{1/2 - \varepsilon}^{1/2} f_X(x) \, dx = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{1/2 - \varepsilon}^{1/2} \delta\left(x - \frac{1}{2}\right) \, dx \neq \frac{1}{16}
\]
However, \( \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} f_x(x) \, dx = \frac{1}{8} \). This ensures,

\[
F_x(x) = \begin{cases} 
\frac{2}{8}, & -\frac{1}{2} \leq x < \frac{1}{2} \\
\frac{3}{8}, & \frac{1}{2} \leq x < \frac{3}{2}
\end{cases}
\]

Such an impulse is referred to as the left-sided delta function.

As \( F_x \) is non-decreasing and \( F_x(\infty) = 1 \), we have

i) \( f_x(x) \geq 0 \) \hspace{1cm} (2.19a)

ii) \( \int_{-\infty}^{\infty} f_x(x) \, dx = 1 \) \hspace{1cm} (2.19b)

Based on the behavior of CDF, a random variable can be classified as:

i) continuous (ii) discrete and (iii) mixed. If the CDF, \( F_x(x) \), is a continuous function of \( x \) for all \( x \), then \( X \) is a continuous random variable. If \( F_x(x) \) is a staircase, then \( X \) corresponds to a discrete variable. We say that \( X \) is a mixed random variable if \( F_x(x) \) is discontinuous but not a staircase. Later on in this lesson, we shall discuss some examples of these three types of variables.

We can induce more than one transformation on a given sample space. If we induce \( k \) such transformations, then we have a set of \( k \) co-existing random variables.

### 2.3.3 Joint distribution and density functions

Consider the case of two random variables, say \( X \) and \( Y \). They can be characterized by the (two-dimensional) joint distribution function, given by

\[
F_{x,y}(x, y) = P\{s : X(s) \leq x, Y(s) \leq y\}
\]

\[\text{Eq. (2.20)}\]

\(^1\) As the domain of the random variable \( X(\ ) \) is known, it is convenient to denote the variable simply by \( X \).

2.19
That is, $F_{X,Y}(x, y)$ is the probability associated with the set of all those sample points such that under $X$, their transformed values will be less than or equal to $x$ and at the same time, under $Y$, the transformed values will be less than or equal to $y$. In other words, $F_{X,Y}(x_1, y_1)$ is the probability associated with the set of all sample points whose transformation does not fall outside the shaded region in the two dimensional (Euclidean) space shown in Fig. 2.5.

![Fig. 2.5: Space of $\{X(s), Y(s)\}$ corresponding to $F_{X,Y}(x_1, y_1)$](image)

Looking at the sample space $S$, let $A$ be the set of all those sample points $s \in S$ such that $X(s) \leq x_1$. Similarly, if $B$ is comprised of all those sample points $s \in S$ such that $Y(s) \leq y_1$; then $F(x_1, y_1)$ is the probability associated with the event $A \cap B$.

Properties of the two dimensional distribution function are:

i) $F_{X,Y}(x, y) \geq 0$, $-\infty < x < \infty$, $-\infty < y < \infty$

ii) $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$

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

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

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

2.20
vi) If \( x_2 > x_1 \) and \( y_2 > y_1 \), then
\[
F_{X,Y}(x_2, y_2) \geq F_{X,Y}(x_2, y_1) \geq F_{X,Y}(x_1, y_1)
\]

We define the two dimensional joint PDF as
\[
f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) \tag{2.21a}
\]
or
\[
F_{X,Y}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(\alpha, \beta) \, d\alpha \, d\beta \tag{2.21b}
\]

The notion of joint CDF and joint PDF can be extended to the case of \( k \) random variables, where \( k \geq 3 \).

Given the joint PDF of random variables \( X \) and \( Y \), it is possible to obtain the one dimensional PDFs, \( f_X(x) \) and \( f_Y(y) \). We know that,
\[
F_{X,Y}(x_1, \infty) = F_X(x_1).
\]

That is,
\[
f_X(x_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X,Y}(\alpha, \beta) \, d\beta \, d\alpha.
\]

\[
f_X(x_1) = \frac{d}{dx} \left\{ \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X,Y}(\alpha, \beta) \, d\beta \, d\alpha \right\} \tag{2.22}
\]

Eq. 2.22 involves the derivative of an integral. Hence,
\[
f_X(x_1) \left. \frac{dF_X(x)}{dx} \right|_{x=x_1} = \int_{-\infty}^{\infty} f_{X,Y}(x_1, \beta) \, d\beta
\]
or
\[
f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \tag{2.23a}
\]

Similarly,
\[
f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \tag{2.23b}
\]

(In the study of several random variables, the statistics of any individual variable is called the marginal. Hence it is common to refer \( F_X(x) \) as the marginal...
distribution function of $X$ and $f_X(x)$ as the marginal density function. $F_Y(y)$ and $f_Y(y)$ are similarly referred to.)

### 2.3.4 Conditional density

Given $f_{X,Y}(x, y)$, we know how to obtain $f_X(x)$ or $f_Y(y)$. We might also be interested in the PDF of $X$ given a specific value of $y = y_1$. This is called the conditional PDF of $X$, given $y = y_1$, denoted $f_{X|Y}(x | y_1)$ and defined as

$$f_{X|Y}(x | y_1) = \frac{f_{X,Y}(x, y_1)}{f_Y(y_1)} \quad (2.24)$$

where it is assumed that $f_Y(y_1) \neq 0$. Once we understand the meaning of conditional PDF, we might as well drop the subscript on $y$ and denote it by $f_{X|Y}(x | y)$. An analogous relation can be defined for $f_{Y|X}(y | x)$. That is, we have the pair of equations,

$$(2.25a) \quad f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

and $$(2.25b) \quad f_{Y|X}(y | x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

or $$(2.25c) \quad f_{X,Y}(x, y) = f_{X|Y}(x | y) f_Y(y)$$

$$= f_{Y|X}(y | x) f_X(x) \quad (2.25d)$$

The function $f_{X|Y}(x | y)$ may be thought of as a function of the variable $x$ with variable $y$ arbitrary, but fixed. Accordingly, it satisfies all the requirements of an ordinary PDF; that is,

$$(2.26a) \quad f_{X|Y}(x | y) \geq 0$$

and $$(2.26b) \quad \int_{-\infty}^{\infty} f_{X|Y}(x | y) \, dx = 1$$
2.3.5 Statistical independence

In the case of random variables, the definition of statistical independence is somewhat simpler than in the case of events. We call $k$ random variables $X_1, X_2, \ldots, X_k$ statistically independent iff, the $k$-dimensional joint PDF factors into the product

$$\prod_{i=1}^{k} f_{X_i}(x_i)$$

Hence, two random variables $X$ and $Y$ are statistically independent, iff,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

and three random variables $X, Y, Z$ are independent, iff

$$f_{X,Y,Z}(x, y, z) = f_X(x) f_Y(y) f_Z(z)$$

Statistical independence can also be expressed in terms of conditional PDF. Let $X$ and $Y$ be independent. Then,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Making use of Eq. 2.25(c), we have

$$f_{X|Y}(x \mid y) f_Y(y) = f_X(x) f_Y(y)$$

or

$$f_{X|Y}(x \mid y) = f_X(x)$$

(2.30a)

Similarly,

$$f_{Y|X}(y \mid x) = f_Y(y)$$

(2.30b)

Eq. 2.30(a) and 2.30(b) are alternate expressions for the statistical independence between $X$ and $Y$. We shall now give a few examples to illustrate the concepts discussed in sec. 2.3.

Example 2.3

A random variable $X$ has

$$F_X(x) = \begin{cases} 0 , & x < 0 \\ Kx^2 , & 0 \leq x \leq 10 \\ 100K , & x > 10 \end{cases}$$

i) Find the constant $K$

ii) Evaluate $P(X \leq 5)$ and $P(5 < X \leq 7)$
iii) What is \( f_X(x) \)?

\[
F_X(\infty) = 100K = 1 \Rightarrow K = \frac{1}{100}.
\]

ii) \( P(x \leq 5) = F_X(5) = \left( \frac{1}{100} \right) \times 25 = 0.25 \)
\[
P(5 < X \leq 7) = F_X(7) - F_X(5) = 0.24
\]
\[
f_X(x) = \frac{dF_X(x)}{dx} = \begin{cases} 0, & x < 0 \\ 0.02x, & 0 \leq x \leq 10 \\ 0, & x > 10 \end{cases}
\]

Note: Specification of \( f_X(x) \) or \( F_X(x) \) is complete only when the algebraic expression as well as the range of \( X \) is given.

\[\blacklozenge\]

\textbf{Example 2.4}

Consider the random variable \( X \) defined by the PDF
\[
f_X(x) = ae^{-bx}, \quad -\infty < x < \infty \text{ where } a \text{ and } b \text{ are positive constants.}
\]

i) Determine the relation between \( a \) and \( b \) so that \( f_X(x) \) is a PDF.

ii) Determine the corresponding \( F_X(x) \)

iii) Find \( P[1 < X \leq 2] \).

i) As can be seen \( f_X(x) \geq 0 \) for \( -\infty < x < \infty \). In order for \( f_X(x) \) to represent a legitimate PDF, we require
\[
\int_{-\infty}^{\infty} ae^{-bx} dx = \int_{0}^{\infty} ae^{-bx} dx = 1.
\]
That is, \( \int_{0}^{\infty} ae^{-bx} dx = \frac{1}{2} \); hence \( b = 2a \).

ii) the given PDF can be written as
\[
f_X(x) = \begin{cases} a e^{bx}, & x < 0 \\ a e^{-bx}, & x \geq 0 \end{cases}
\]
For $x < 0$, we have

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{2} b e^{bx} \, dx = \frac{1}{2} e^{bx}.$$  

Consider $x > 0$. Take a specific value of $x = 2$

$$F_X(2) = \int_{-\infty}^{2} f_X(x) \, dx$$

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

But for the problem on hand, $\int_{2}^{\infty} f_X(x) \, dx = \int_{-\infty}^{2} f_X(x) \, dx$

Therefore, $F_X(x) = 1 - \frac{1}{2} e^{-bx}$, $x > 0$

We can now write the complete expression for the CDF as

$$F_X(x) = \begin{cases} \frac{1}{2} e^{bx}, & x < 0 \\ 1 - \frac{1}{2} e^{-bx}, & x \geq 0 \end{cases}$$

iii) $F_X(2) - F_X(1) = \frac{1}{2} (e^{-b} - e^{-2b})$

**Example 2.5**

Let $f_{X,Y}(x, y) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq y, \ 0 \leq y \leq 2 \\ 0, & \text{otherwise} \end{cases}$

Find (a) i) $f_{Y|X}(y \mid 1)$ and ii) $f_{Y|X}(y \mid 1.5)$

(b) Are $X$ and $Y$ independent?

(a) $f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$
\( f_X(x) \) can be obtained from \( f_{X,Y} \) by integrating out the unwanted variable \( y \) over the appropriate range. The maximum value taken by \( y \) is 2; in addition, for any given \( x, y \geq x \). Hence,

\[
f_X(x) = \int_{0}^{2} \frac{1}{x^2} \, dy = 1 - \frac{x}{2}, \quad 0 \leq x \leq 2
\]

Hence,

(i) \( f_{Y|X}(y | 1) = \frac{1}{2} \cdot \frac{1}{2} = \begin{cases} 1, & 1 \leq y \leq 2 \\ 0, & otherwise \end{cases} \)

(ii) \( f_{Y|X}(y | 1.5) = \frac{1}{2} \cdot \frac{1}{4} = \begin{cases} 2, & 1.5 \leq y \leq 2 \\ 0, & otherwise \end{cases} \)

b) the dependence between the random variables \( X \) and \( Y \) is evident from the statement of the problem because given a value of \( X = x_i \), \( Y \) should be greater than or equal to \( x_i \) for the joint PDF to be non-zero. Also we see that \( f_{Y|X}(y | x) \) depends on \( x \) whereas if \( X \) and \( Y \) were to be independent, then \( f_{Y|X}(y | x) = f_Y(y) \).
Exercise 2.3

For the two random variables $X$ and $Y$, the following density functions have been given. (Fig. 2.6)

Find

a) $f_{X,Y}(x, y)$

b) Show that

$$f_Y(y) = \begin{cases} 
\frac{y}{100}, & 0 \leq y \leq 10 \\
\frac{1}{5} - \frac{y}{100}, & 10 < y \leq 20 
\end{cases}$$

2.4 Transformation of Variables

The process of communication involves various operations such as modulation, detection, filtering etc. In performing these operations, we typically generate new random variables from the given ones. We will now develop the necessary theory for the statistical characterization of the output random variables, given the transformation and the input random variables.
2.4.1 Functions of one random variable

Assume that $X$ is the given random variable of the continuous type with the PDF, $f_X(x)$. Let $Y$ be the new random variable, obtained from $X$ by the transformation $Y = g(X)$. By this we mean the number $Y(s_i)$ associated with any sample point $s_i$ is

$$Y(s_i) = g(X(s_i))$$

Our interest is to obtain $f_Y(y)$. This can be obtained with the help of the following theorem (Thm. 2.1).

**Theorem 2.1**

Let $Y = g(X)$. To find $f_Y(y)$, for the specific $y$, solve the equation $y = g(x)$. Denoting its real roots by $x_1, x_2, \ldots, x_n$, we will show that

$$f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \ldots + \frac{f_X(x_n)}{|g'(x_n)|}$$

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

**Proof:** Consider the transformation shown in Fig. 2.7. We see that the equation $y_i = g(x)$ has three roots namely, $x_1, x_2$ and $x_3$. 

2.28
We know that \( f_y(y) \, dy = P[y < Y \leq y + dy] \). Therefore, for any given \( y_1 \), we need to find the set of values \( x \) such that \( y_1 < g(x) \leq y_1 + dy \) and the probability that \( X \) is in this set. As we see from the figure, this set consists of the following intervals:

\[
x_1 < x \leq x_1 + d\,x_1, \quad x_2 + d\,x_2 < x \leq x_2, \quad x_3 < x \leq x_3 + d\,x_3
\]

where \( dx_1 > 0, \, dx_3 > 0, \) but \( dx_2 < 0 \).

From the above, it follows that

\[
P[y_1 < Y \leq y_1 + dy] = P[x_1 < X \leq x_1 + d\,x_1] + P[x_2 + d\,x_2 < X \leq x_2] + P[x_3 < X \leq x_3 + d\,x_3]
\]

This probability is the shaded area in Fig. 2.7.

\[
P[x_1 < X \leq x_1 + d\,x_1] = f_x(x_1) \, dx_1, \quad dx_1 = \frac{dy}{g'(x_1)}
\]

\[
P[x_2 + d\,x_2 < X \leq x_2] = f_x(x_2) \, dx_2, \quad dx_2 = \frac{dy}{g'(x_2)}
\]

\[
P[x_3 < X \leq x_3 + d\,x_3] = f_x(x_3) \, dx_3, \quad dx_3 = \frac{dy}{g'(x_3)}
\]
We conclude that

\[ f_r(y) \, dy = \frac{f_x(x_1)}{g'(x_1)} \, dy + \frac{f_x(x_2)}{\left| g'(x_2) \right|} \, dy + \frac{f_x(x_3)}{g'(x_3)} \, dy \quad (2.32) \]

and Eq. 2.31 follows, by canceling \( dy \) from the Eq. 2.32.

Note that if \( g(x) = y_1 = \text{constant for every } x \) in the interval \( (x_0, x_1) \), then we have \( P[Y = y_1] = P(x_0 < X \leq x_1) = F_x(x_1) - F_x(x_0) \); that is \( F_y(y) \) is discontinuous at \( y = y_1 \). Hence \( f_y(y) \) contains an impulse, \( \delta(y - y_1) \) of area \( F_x(x_1) - F_x(x_0) \).

We shall take up a few examples.

**Example 2.6**

\[ Y = g(X) = X + a \], where \( a \) is a constant. Let us find \( f_y(y) \).

We have \( g'(x) = 1 \) and \( x = y - a \). For a given \( y \), there is a unique \( x \) satisfying the above transformation. Hence,

\[ f_r(y) \, dy = \frac{f_x(x)}{g'(x)} \, dy \quad \text{and as } g'(x) = 1, \text{ we have} \]

\[ f_r(y) = f_x(y - a) \]

Let \( f_x(x) = \begin{cases} 1 - |x|, & |x| \leq 1 \\ 0, & \text{elsewhere} \end{cases} \)

and \( a = -1 \)

Then, \( f_y(y) = 1 - |y + 1| \)

As \( y = x - 1 \), and \( x \) ranges from the interval \((-1, 1)\), we have the range of \( y \) as \((-2, 0)\).
Hence \( f_y(y) = \begin{cases} 1 - |y + 1|, & -2 \leq y \leq 0 \\ 0, & \text{elsewhere} \end{cases} \)

\( F_x(x) \) and \( F_y(y) \) are shown in Fig. 2.8.

![Fig. 2.8: \( F_x(x) \) and \( F_y(y) \) of example 2.6](image)

As can be seen, the transformation of adding a constant to the given variable simply results in the translation of its PDF.

**Example 2.7**

Let \( Y = bX \), where \( b \) is a constant. Let us find \( f_y(y) \).

Solving for \( X \), we have \( X = \frac{1}{b}Y \). Again for a given \( y \), there is a unique \( x \). As \( g'(x) = b \), we have \( f_y(y) = \frac{1}{|b|} f_x \left( \frac{y}{b} \right) \).

Let \( f_x(x) = \begin{cases} 1 - \frac{x}{2}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases} \)

and \( b = -2 \), then
\[ f_Y(y) = \begin{cases} \frac{1}{2} \left[ 1 + \frac{y}{4} \right], & -4 \leq y \leq 0 \\ 0, & \text{otherwise} \end{cases} \]

\( f_X(x) \) and \( f_Y(y) \) are sketched in Fig. 2.9.

\[ f_{x}(x) \] and \( f_{y}(y) \) of example 2.7

**Exercise 2.4**

Let \( Y = aX + b \), where \( a \) and \( b \) are constants. Show that

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

If \( f_X(x) \) is as shown in Fig. 2.8, compute and sketch \( f_Y(y) \) for \( a = -2 \), and \( b = 1 \).

**Example 2.8**

\( Y = aX^2 \), \( a > 0 \). Let us find \( f_Y(y) \).
\[ g'(x) = 2ax. \text{ If } y < 0, \text{ then the equation } y = ax^2 \text{ has no real solution.} \]

Hence \( f_Y(y) = 0 \) for \( y < 0 \). If \( y \geq 0 \), then it has two solutions, \( x_1 = \frac{\sqrt{y}}{\sqrt{a}} \) and \( x_2 = -\frac{\sqrt{y}}{\sqrt{a}} \), and Eq. 2.31 yields

\[
f_Y(y) = \begin{cases} 
\frac{1}{2a\sqrt{\frac{y}{a}}} \left[ f_X \left( \frac{\sqrt{y}}{\sqrt{a}} \right) + f_X \left( -\frac{\sqrt{y}}{\sqrt{a}} \right) \right], & y \geq 0 \\
0, & \text{otherwise}
\end{cases}
\]

Let \( a = 1 \), and \( f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), -\infty < x < \infty \)

(Note that \( \exp(\alpha) \) is the same as \( e^\alpha \))

Then \( f_Y(y) = \frac{1}{2\sqrt{y}\sqrt{2\pi}} \left[ \exp\left(-\frac{y}{2}\right) + \exp\left(-\frac{y}{2}\right) \right] \)

\[
= \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right), \quad y \geq 0
\]

\[
= 0, \quad \text{otherwise}
\]

Sketching of \( f_X(x) \) and \( f_Y(y) \) is left as an exercise.

---

**Example 2.9**

Consider the half wave rectifier transformation given by

\[
Y = \begin{cases} 
0, & X \leq 0 \\
X, & X > 0
\end{cases}
\]

a) Let us find the general expression for \( f_Y(y) \)

b) Let \( f_X(x) = \begin{cases} 
\frac{1}{2}, & -\frac{1}{2} < x < \frac{3}{2} \\
0, & \text{otherwise}
\end{cases} \)

We shall compute and sketch \( f_Y(y) \).
a) Note that $g(X)$ is a constant (equal to zero) for $X$ in the range of $(-\infty, 0)$. Hence, there is an impulse in $f_Y(y)$ at $y = 0$ whose area is equal to $F_X(0)$. As $Y$ is nonnegative, $f_Y(y) = 0$ for $y < 0$. As $Y = X$ for $x > 0$, we have $f_Y(y) = f_X(y)$ for $y > 0$. Hence

$$f_Y(y) = \begin{cases} f_X(y)w(y) + F_X(0)\delta(y) \\ 0, \text{ otherwise} \end{cases}$$

where $w(y) = \begin{cases} 1, y \geq 0 \\ 0, \text{ otherwise} \end{cases}$

b) Specifically, let $f_X(x) = \begin{cases} \frac{1}{2}, & -\frac{1}{2} < x < \frac{3}{2} \\ 0, & \text{ elsewhere} \end{cases}$

Then, $f_Y(y) = \begin{cases} \frac{1}{4}\delta(y), & y = 0 \\ \frac{1}{2}, & 0 < y \leq \frac{3}{2} \\ 0, \text{ otherwise} \end{cases}$

$f_X(x)$ and $f_Y(y)$ are sketched in Fig. 2.10.

![Fig. 2.10: $f_X(x)$ and $f_Y(y)$ for the example 2.9](image-url)
Note that $X$, a continuous RV is transformed into a mixed RV, $Y$.

**Example 2.10**

Let $Y = \begin{cases} -1, & X < 0 \\ +1, & X \geq 0 \end{cases}$

a) Let us find the general expression for $f_Y(y)$.

b) We shall compute and sketch $f_Y(x)$ assuming that $f_X(x)$ is the same as that of Example 2.9.

a) In this case, $Y$ assumes only two values, namely $\pm 1$. Hence the PDF of $Y$ has only two impulses. Let us write $f_Y(y)$ as

$$f_Y(y) = P_1 \delta(y-1) + P_{-1} \delta(y+1)$$

where

$$P_{-1} = P[X < 0] \text{ and } P_1 = P[X \geq 0]$$

b) Taking $f_X(x)$ of example 2.9, we have $P_1 = \frac{3}{4}$ and $P_{-1} = \frac{1}{4}$. Fig. 2.11 has the sketches $f_X(x)$ and $f_Y(y)$.

Note that this transformation has converted a continuous random variable $X$ into a discrete random variable $Y$. 

![Fig. 2.11: $f_X(x)$ and $f_Y(y)$ for the example 2.10](image-url)
Exercise 2.5

Let a random variable $X$ with the PDF shown in Fig. 2.12(a) be the input to a device with the input-output characteristic shown in Fig. 2.12(b). Compute and sketch $f_Y(y)$.

Fig. 2.12: (a) Input PDF for the transformation of exercise 2.5
(b) Input-output transformation

Exercise 2.6

The random variable $X$ of exercise 2.5 is applied as input to the $X - Y$ transformation shown in Fig. 2.13. Compute and sketch $f_Y(y)$.
We now assume that the random variable $X$ is of \textbf{discrete type} taking on the value $x_k$ with probability $P_k$. In this case, the RV, $Y = g(X)$ is also discrete, assuming the value $Y_k = g(x_k)$ with probability $P_k$.

If $y_k = g(x)$ for only one $x = x_k$, then $P[Y = y_k] = P[X = x_k] = P_k$. If however, $y_k = g(x)$ for $x = x_k$ and $x = x_m$, then $P[Y = y_k] = P_k + P_m$.

\textbf{Example 2.11}

Let $Y = X^2$.

a) If $f_X(x) = \frac{1}{6} \sum_{i=1}^{6} \delta(x - i)$, find $f_Y(y)$.

b) If $f_X(x) = \frac{1}{6} \sum_{i=-2}^{3} \delta(x - i)$, find $f_Y(y)$.

a) If $X$ takes the values $(1, 2, \ldots, 6)$ with probability of $\frac{1}{6}$, then $Y$ takes the values $1^2, 2^2, \ldots, 6^2$ with probability $\frac{1}{6}$. That is, $f_Y(y) = \frac{1}{6} \sum_{i=1}^{6} \delta(x - i^2)$.

b) If, however, $X$ takes the values $-2, -1, 0, 1, 2, 3$ with probability $\frac{1}{6}$, then $Y$ takes the values $0, 1, 4, 9$ with probabilities $\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}$ respectively.

That is,

$$f_Y(y) = \frac{1}{6} \left[ \delta(y) + \delta(y - 9) \right] + \frac{1}{3} \left[ \delta(y - 1) + \delta(y - 4) \right]$$

\textbf{2.4.2 Functions of two random variables}

Given two random variables $X$ and $Y$ (assumed to be \textbf{continuous type}), two new random variables, $Z$ and $W$ are defined using the transformation $Z = g(X, Y)$ and $W = h(X, Y)$. 

2.37
Given the above transformation and \( f_{x,y}(x, y) \), we would like to obtain \( f_{z,w}(z, w) \). For this, we require the Jacobian of the transformation, denoted \( J \left( \frac{z, w}{x, y} \right) \) where

\[
J \left( \frac{z, w}{x, y} \right) = \begin{vmatrix}
\frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\
\frac{\partial w}{\partial x} & \frac{\partial w}{\partial y}
\end{vmatrix}
= \frac{\partial z}{\partial x} \frac{\partial w}{\partial y} - \frac{\partial z}{\partial y} \frac{\partial w}{\partial x}
\]

That is, the Jacobian is the determinant of the appropriate partial derivatives. We shall now state the theorem which relates \( f_{z,w}(z, w) \) and \( f_{x,y}(x, y) \).

We shall assume that the transformation is one-to-one. That is, given

\[
g(x, y) = z_1, \quad (2.33a)
\]
\[
h(x, y) = w_1, \quad (2.33b)
\]

then there is a unique set of numbers, \((x_1, y_1)\) satisfying Eq. 2.33.

**Theorem 2.2:** To obtain \( f_{z,w}(z, w) \), solve the system of equations

\[
g(x, y) = z_1,
\]
\[
h(x, y) = w_1,
\]

for \( x \) and \( y \). Let \((x_1, y_1)\) be the result of the solution. Then,

\[
f_{z,w}(z, w) = f_{x,y}(x_1, y_1) \left| J \left( \frac{z, w}{x_1, y_1} \right) \right| \quad (2.34)
\]

Proof of this theorem is given in appendix A2.1. For a more general version of this theorem, refer [1].

2.38
Example 2.12

$X$ and $Y$ are two independent RVs with the PDFs,

$$f_x(x) = \frac{1}{2}, \quad |x| \leq 1$$

$$f_y(y) = \frac{1}{2}, \quad |y| \leq 1$$

If $Z = X + Y$ and $W = X - Y$, let us find (a) $f_{z,w}(z, w)$ and (b) $f_z(z)$.

a) From the given transformations, we obtain $x = \frac{1}{2}(z + w)$ and $y = \frac{1}{2}(z - w)$. We see that the mapping is one-to-one. Fig. 2.14(a) depicts the (product) space $\mathcal{A}$ on which $f_{x,y}(x, y)$ is non-zero.

Fig. 2.14: (a) The space where $f_{x,y}$ is non-zero

(b) The space where $f_{z,w}$ is non-zero
We can obtain the space \( \mathcal{B} \) (on which \( f_{z,w}(z,w) \) is non-zero) as follows:

<table>
<thead>
<tr>
<th>( \mathcal{A} ) space</th>
<th>( \mathcal{B} ) space</th>
</tr>
</thead>
<tbody>
<tr>
<td>The line ( x = 1 ) ( \frac{1}{2}(z+w) = 1 \Rightarrow w = -z + 2 )</td>
<td></td>
</tr>
<tr>
<td>The line ( x = -1 ) ( \frac{1}{2}(z+w) = -1 \Rightarrow w = -(z+2) )</td>
<td></td>
</tr>
<tr>
<td>The line ( y = 1 ) ( \frac{1}{2}(z-w) = 1 \Rightarrow w = z - 2 )</td>
<td></td>
</tr>
<tr>
<td>The line ( y = -1 ) ( \frac{1}{2}(z-w) = -1 \Rightarrow w = z + 2 )</td>
<td></td>
</tr>
</tbody>
</table>

The space \( \mathcal{B} \) is shown in Fig. 2.14(b). The Jacobian of the transformation is

\[
J\left(\frac{z,w}{x,y}\right) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2 \quad \text{and} \quad |J(\quad)| = 2.
\]

Hence \( f_{z,w}(z,w) = \frac{1/4}{2} = \begin{cases} \frac{1}{8}, & z, w \in \mathcal{B} \\ 0, & \text{otherwise} \end{cases} \)

b) \( f_z(z) = \int_{-\infty}^{\infty} f_{z,w}(z,w) \, dw \)

From Fig. 2.14(b), we can see that, for a given \( z \) \( (z \geq 0) \), \( w \) can take values only in the \(-z\) to \( z\). Hence

\[
f_z(z) = \int_{-z}^{z} \frac{1}{8} \, dw = \frac{1}{4} z, \quad 0 \leq z \leq 2
\]

For \( z \) negative, we have

\[
f_z(z) = \int_{z}^{-z} \frac{1}{8} \, dw = -\frac{1}{4} z, \quad -2 \leq z \leq 0
\]
Hence $f_z(z) = \frac{1}{4} |z|, \ |z| \leq 2$

**Example 2.13**

Let the random variables $R$ and $\Phi$ be given by, $R = \sqrt{X^2 + Y^2}$ and $\Phi = \arctan\left(\frac{Y}{X}\right)$ where we assume $R \geq 0$ and $-\pi < \Phi < \pi$. It is given that

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left[-\left(\frac{x^2 + y^2}{2}\right)\right], \ -\infty < x, y < \infty.$$  

Let us find $f_{R, \Phi}(r, \varphi)$.

As the given transformation is from cartesian-to-polar coordinates, we can write $x = r \cos \varphi$ and $y = r \sin \varphi$, and the transformation is one-to-one;

$$J = \left| \begin{array}{cc}
\frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\
\frac{\partial \varphi}{\partial x} & \frac{\partial \varphi}{\partial y}
\end{array} \right| = \begin{vmatrix}
\cos \varphi & \sin \varphi \\
-\sin \varphi & \cos \varphi
\end{vmatrix} = \frac{1}{r}$$

Hence, $f_{R, \Phi}(r, \varphi) = \begin{cases} 
\frac{r}{2\pi} \exp\left(-\frac{r^2}{2}\right), & 0 \leq r \leq \infty, \ -\pi < \varphi < \pi \\
0, & \text{otherwise}
\end{cases}$

It is left as an exercise to find $f_R(r), f_\Phi(\varphi)$ and to show that $R$ and $\Phi$ are independent variables.

Theorem 2.2 can also be used when only one function $Z = g(X, Y)$ is specified and what is required is $f_z(z)$. To apply the theorem, a conveniently chosen auxiliary or dummy variable $W$ is introduced. Typically $W = X$ or $W = Y$; using the theorem $f_{Z,W}(z, w)$ is found from which $f_z(z)$ can be obtained.
Let \( Z = X + Y \) and we require \( f_z(z) \). Let us introduce a dummy variable \( W = Y \). Then, \( X = Z - W \), and \( Y = W \).

As \( J = 1 \),
\[
f_{z,w}(z, w) = f_{x,y}(z - w, w)
\]
and
\[
f_z(z) = \int_{-\infty}^{\infty} f_{x,y}(z - w, w) \, dw
\] (2.35)

If \( X \) and \( Y \) are independent, then Eq. 2.35 becomes
\[
f_z(z) = \int_{-\infty}^{\infty} f_x(z - w) f_y(w) \, dw
\] (2.36a)

That is, \( f_z = f_x \ast f_y \) (2.36b)

**Example 2.14**

Let \( X \) and \( Y \) be two independent random variables, with
\[
f_x(x) = \begin{cases} 
1/2, & -1 \leq x \leq 1 \\
0, & \text{otherwise}
\end{cases}
\]
\[
f_y(y) = \begin{cases} 
1/3, & -2 \leq y \leq 1 \\
0, & \text{otherwise}
\end{cases}
\]

If \( Z = X + Y \), let us find \( P[Z \leq -2] \).

From Eq. 2.36(b), \( f_z(z) \) is the convolution of \( f_x(z) \) and \( f_y(z) \). Carrying out the convolution, we obtain \( f_z(z) \) as shown in Fig. 2.15.
Fig. 2.15: PDF of $Z = X + Y$ of example 2.14

$P[Z \leq -2]$ is the shaded area $= \frac{1}{12}$.  

**Example 2.15**

Let $Z = \frac{X}{Y}$; let us find an expression for $f_z(z)$.

Introducing the dummy variable $W = Y$, we have

$X = Z W$

$Y = W$

As $J = \frac{1}{w}$, we have

$$f_z(z) = \int_{-\infty}^{\infty} |w| f_{x,y}(zw, w) \, dw$$

Let $f_{x,y}(x, y) = \begin{cases} \frac{1 + xy}{4}, & |x| \leq 1, \ |y| \leq 1 \\ 0, & \text{elsewhere} \end{cases}$

Then $f_z(z) = \int_{-\infty}^{\infty} |w| \frac{1 + zw^2}{4} \, dw$
$f_{x,y}(x,y)$ is non-zero if $(x,y) \in \mathcal{A}$, where $\mathcal{A}$ is the product space $|x| \leq 1$ and $|y| \leq 1$ (Fig. 2.16a). Let $f_{z,w}(z,w)$ be non-zero if $(z,w) \in \mathcal{B}$. Under the given transformation, $\mathcal{B}$ will be as shown in Fig. 2.16(b).

![Graphs showing the non-zero regions for $f_{x,y}$ and $f_{z,w}$](image)

**Fig. 2.16:** (a) The space where $f_{x,y}$ is non-zero  
(b) The space where $f_{z,w}$ is non-zero

To obtain $f_z(z)$ from $f_{z,w}(z,w)$, we have to integrate out $w$ over the appropriate ranges.

i) Let $|z| < 1$; then

$$f_z(z) = \int_{-1}^{1} \frac{1+zw^2}{4} |w| dw = 2 \int_0^1 \frac{1+zw^2}{4} w dw$$

$$= \frac{1}{4} \left(1 + \frac{z}{2}\right)$$

ii) For $z > 1$, we have

$$f_z(z) = 2 \int_0^{\frac{1}{\sqrt{z}}} \frac{1+zw^2}{4} w dw = \frac{1}{4} \left(\frac{1}{z^2} + \frac{1}{2z^3}\right)$$

iii) For $z < -1$, we have

$$f_z(z) = 2 \int_0^{\frac{1}{\sqrt{|z|}}} \frac{1+zw^2}{4} w dw = \frac{1}{4} \left(\frac{1}{z^2} + \frac{1}{2z^3}\right)$$

2.44
Hence \( f_z(z) = \begin{cases} \frac{1}{4} \left(1 + \frac{z}{2}\right), & |z| \leq 1 \\ \frac{1}{4} \left(\frac{1}{z^2} + \frac{1}{2z^3}\right), & |z| > 1 \end{cases} \)

**Exercise 2.7**

Let \( Z = X + Y \) and \( W = Y \).

a) Show that
\[
\int_{-\infty}^{\infty} \frac{1}{|W|} f_{X,Y} \left( \frac{Z}{W}, W \right) \, dW
\]

b) \( X \) and \( Y \) be independent with
\[
f_X(x) = \frac{1}{\pi \sqrt{1 + x^2}}, \quad |x| \leq 1
\]

and
\[
f_Y(y) = \begin{cases} y e^{-y^2/2}, & y \geq 0 \\ 0, & \text{otherwise} \end{cases}
\]

Show that
\[
f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.
\]

In transformations involving two random variables, we may encounter a situation where one of the variables is continuous and the other is discrete; such cases are handled better, by making use of the **distribution function** approach, as illustrated below.

**Example 2.16**

The input to a noisy channel is a binary random variable with
\[
P[X = 0] = P[X = 1] = \frac{1}{2}.
\]

The output of the channel is given by \( Z = X + Y \).
where \( Y \) is the channel noise with 
\[
f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, -\infty < y < \infty.
\]
Find \( f_z(z) \).

Let us first compute the distribution function of \( Z \) from which the density function can be derived.

\[
P(Z \leq z) = P[Z \leq z | X = 0] P[X = 0] + P[Z \leq z | X = 1] P[X = 1]
\]
As \( Z = X + Y \), we have

\[
P[Z \leq z | X = 0] = F_Y(z)
\]

Similarly \( P[Z \leq z | X = 1] = F_Y(z - 1) \)

Hence 
\[
F_z(z) = \frac{1}{2} F_Y(z) + \frac{1}{2} F_Y(z - 1)
\]
As \( f_z(z) = \frac{d}{dz} F_z(z) \), we have

\[
f_z(z) = \frac{1}{2} \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z - 1)^2}{2}\right) \right]
\]

The distribution function method, as illustrated by means of example 2.16, is a very basic method and can be used in all situations. (In this method, if \( Y = g(X) \), we compute \( F_Y(y) \) and then obtain \( f_y(y) = \frac{d}{dy} F_Y(y) \). Similarly, for transformations involving more than one random variable. Of course computing the CDF may prove to be quite difficult in certain situations. The method of obtaining PDFs based on theorem 2.1 or 2.2 is called change of variable method.) We shall illustrate the distribution function method with another example.

**Example 2.17**

Let \( Z = X + Y \). Obtain \( f_z(z) \), given \( f_{x,y}(x, y) \).
\[ P[Z \leq z] = P[X + Y \leq z] = P[Y \leq z - x] \]

This probability is the probability of \((X, Y)\) lying in the shaded area shown in Fig. 2.17.

![Fig. 2.17: Shaded area is the \(P[Z \leq z]\)](image)

That is,

\[
F_z(z) = \int_{-\infty}^{\infty} dx \left[ \int_{-\infty}^{z} \int_{-\infty}^{z-x} f_{X,Y}(x, y) \, dy \right]
\]

\[
f_z(z) = \frac{\partial}{\partial z} \left[ \int_{-\infty}^{\infty} dx \left[ \int_{-\infty}^{z} \int_{-\infty}^{z-x} f_{X,Y}(x, y) \, dy \right] \right]
\]

\[
= \int_{-\infty}^{\infty} dx \left[ \frac{\partial}{\partial z} \int_{-\infty}^{z-x} f_{X,Y}(x, y) \, dy \right]
\]

\[
= \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) \, dx
\]  \hspace{1cm} (2.37a)

It is not too difficult to see the alternative form for \(f_z(z)\), namely,
\[ f_z(z) = \int_{-\infty}^{\infty} f_{x,y}(z - y, y) \, dy \] (2.37b)

If \( X \) and \( Y \) are independent, we have \( f_z(z) = f_x(z) \ast f_y(z) \). We note that Eq. 2.37(b) is the same as Eq. 2.35.

So far we have considered the transformations involving one or two variables. This can be generalized to the case of functions of \( n \) variables. Details can be found in [1, 2].

### 2.5 Statistical Averages

The PDF of a random variable provides a complete statistical characterization of the variable. However, we might be in a situation where the PDF is not available but are able to estimate (with reasonable accuracy) certain (statistical) averages of the random variable. Some of these averages do provide a simple and fairly adequate (though incomplete) description of the random variable. We now define a few of these averages and explore their significance.

The **mean value** (also called the **expected value**, **mathematical expectation** or simply **expectation**) of random variable \( X \) is defined as

\[
m_x = E[X] = \overline{X} = \int_{-\infty}^{\infty} x f_x(x) \, dx
\]

(2.38)

where \( E \) denotes the expectation operator. Note that \( m_x \) is a constant. Similarly, the expected value of a function of \( X \), \( g(X) \), is defined by

\[
E[g(X)] = \overline{g(X)} = \int_{-\infty}^{\infty} g(x) f_x(x) \, dx
\]

(2.39)

**Remarks:** The terminology expected value or expectation has its origin in games of chance. This can be illustrated as follows: Three small similar discs, numbered 1, 2 and 2 respectively are placed in bowl and are mixed. A player is to be
blindfolded and is to draw a disc from the bowl. If he draws the disc numbered 1, he will receive nine dollars; if he draws either disc numbered 2, he will receive 3 dollars. It seems reasonable to assume that the player has a '1/3 claim' on the 9 dollars and '2/3 claim' on three dollars. His total claim is \(9(1/3) + 3(2/3)\), or five dollars. If we take \(X\) to be (discrete) random variable with the PDF

\[
f_X(x) = \frac{1}{3} \delta(x - 1) + \frac{2}{3} \delta(x - 2) \text{ and } g(X) = 15 - 6X,
\]

then

\[
E[g(X)] = \int_{-\infty}^{\infty} (15 - 6x) f_X(x) \, dx = 5.
\]

That is, the mathematical expectation of \(g(X)\) is precisely the player's claim or expectation [3]. Note that \(g(x)\) is such that \(g(1) = 9\) and \(g(2) = 3\).

For the special case of \(g(X) = X^n\), we obtain the \(n\)-th moment of the probability distribution of the RV, \(X\); that is,

\[
E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x) \, dx \tag{2.40}
\]

The most widely used moments are the first moment \((n = 1, \text{ which results in the mean value of Eq. 2.38})\) and the second moment \((n = 2, \text{ resulting in the mean square value of } X)\).

\[
E[X^2] = \overline{X^2} = \int_{-\infty}^{\infty} x^2 f_X(x) \, dx \tag{2.41}
\]

If \(g(X) = (X - m_X)^n\), then \(E[g(X)]\) gives \(n\)-th central moment; that is,

\[
E[(X - m_X)^n] = \int_{-\infty}^{\infty} (x - m_X)^n f_X(x) \, dx \tag{2.42}
\]

We can extend the definition of expectation to the case of functions of \(k(k \geq 2)\) random variables. Consider a function of two random variables, \(g(X, Y)\). Then,

2.49
An important property of the expectation operator is linearity; that is, if \( Z = g(X, Y) = \alpha X + \beta Y \) where \( \alpha \) and \( \beta \) are constants, then \( \mathbb{E}[Z] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y] \).

This result can be established as follows. From Eq. 2.43, we have

\[
E[Z] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\alpha x + \beta y) f_{X,Y}(x, y) \, dx \, dy
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\alpha x) f_{X,Y}(x, y) \, dx \, dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\beta y) f_{X,Y}(x, y) \, dx \, dy
\]

Integrating out the variable \( y \) in the first term and the variable \( x \) in the second term, we have

\[
E[Z] = \int_{-\infty}^{\infty} \alpha x f_{X}(x) \, dx + \int_{-\infty}^{\infty} \beta y f_{Y}(y) \, dy
\]

\[
= \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]
\]

### 2.5.1 Variance

Coming back to the central moments, we have the first central moment being always zero because,

\[
E[(X - m_X)] = \int_{-\infty}^{\infty} (x - m_X) f_X(x) \, dx
\]

\[
= m_X - m_X = 0
\]

Consider the second central moment

\[
E[(X - m_X)^2] = E[X^2 - 2m_X X + m_X^2]
\]

From the linearity property of expectation,

\[
E[X^2 - 2m_X X + m_X^2] = E[X^2] - 2m_X E[X] + m_X^2
\]

\[
= E[X^2] - 2m_X^2 + m_X^2
\]

\[
= \mathbb{E}^2 - m_X^2 = \mathbb{E}^2 - \mathbb{E}^2
\]
The second central moment of a random variable is called the **variance** and its (positive) square root is called the **standard deviation**. The symbol $\sigma^2$ is generally used to denote the variance. (If necessary, we use a subscript on $\sigma^2$)

The variance provides a measure of the variable’s spread or randomness. Specifying the variance essentially constrains the effective width of the density function. This can be made more precise with the help of the **Chebyshev Inequality** which follows as a special case of the following theorem.

**Theorem 2.3**: Let $g(X)$ be a non-negative function of the random variable $X$. If $E[g(X)]$ exists, then for every positive constant $c$,$$
P[g(X) \geq c] \leq \frac{E[g(X)]}{c} \quad (2.44)
$$

**Proof**: Let $A = \{x : g(x) \geq c\}$ and $B$ denote the complement of $A$.

\[
E[g(X)] = \int_{-\infty}^{\infty} g(x) f_x(x) \, dx \\
= \int_{A} g(x) f_x(x) \, dx + \int_{B} g(x) f_x(x) \, dx
\]

Since each integral on the RHS above is non-negative, we can write

\[
E[g(X)] \geq \int_{A} g(x) f_x(x) \, dx
\]

If $x \in A$, then $g(x) \geq c$, hence

\[
E[g(X)] \geq c \int_{A} f_x(x) \, dx
\]

But $\int_{A} f_x(x) \, dx = P[x \in A] = P[g(X) \geq c]$

That is, $E[g(X)] \geq c \, P[g(X) \geq c]$, which is the desired result.

Note: The kind of manipulations used in proving the theorem 2.3 is useful in establishing similar inequalities involving random variables.
To see how innocuous (or weak, perhaps) the inequality 2.44 is, let \( g(X) \) represent the height of a randomly chosen human being with \( E[g(X)] = 1.6m \). Then Eq. 2.44 states that the probability of choosing a person over 16 m tall is at most \( \frac{1}{10} \)! (In a population of 1 billion, at most 100 million would be as tall as a full grown Palmyra tree!)

Chebyshev inequality can be stated in two equivalent forms:

\[
\begin{align*}
\text{i)} \quad & P[|X - m_X| \geq k \sigma_X] \leq \frac{1}{k^2}, \quad k > 0 \\
\text{ii)} \quad & P[|X - m_X| < k \sigma_X] > 1 - \frac{1}{k^2}
\end{align*}
\] (2.45a, 2.45b)

where \( \sigma_X \) is the standard deviation of \( X \).

To establish 2.45(a), let \( g(X) = (X - m_X)^2 \) and \( c = k^2 \sigma_X^2 \) in theorem 2.3. We then have,

\[
P[(X - m_X)^2 \geq k^2 \sigma_X^2] \leq \frac{1}{k^2}
\]

In other words,

\[
P[|X - m_X| \geq k \sigma_X] \leq \frac{1}{k^2}
\]

which is the desired result. Naturally, we would take the positive number \( k \) to be greater than one to have a meaningful result. Chebyshev inequality can be interpreted as: the probability of observing any RV outside \( \pm k \) standard deviations off its mean value is no larger than \( \frac{1}{k^2} \). With \( k = 2 \) for example, the probability of \( |X - m_X| \geq 2 \sigma_X \) does not exceed \( \frac{1}{4} \) or 25\%. By the same token, we expect \( X \) to occur within the range \( (m_X \pm 2 \sigma_X) \) for more than 75\% of the observations. That is, smaller the standard deviation, smaller is the width of the interval around \( m_X \), where the required probability is concentrated. Chebyshev
inequality thus enables us to give a quantitative interpretation to the statement 'variance is indicative of the spread of the random variable'.

Note that it is not necessary that variance exists for every PDF. For example, if
\[ f_x(x) = \frac{\alpha}{\alpha^2 + x^2}, \quad -\infty < x < \infty \text{ and } \alpha > 0, \]
then \( \overline{X} = 0 \) but \( \overline{X^2} \) is not finite.
(This is called Cauchy's PDF)

### 2.5.2 Covariance

An important joint expectation is the quantity called **covariance** which is obtained by letting \( g(X, Y) = (X - m_X)(Y - m_Y) \) in Eq. 2.43. We use the symbol \( \lambda \) to denote the covariance. That is,
\[
\lambda_{XY} = \text{cov}[X, Y] = E[(X - m_X)(Y - m_Y)] \tag{2.46a}
\]

Using the linearity property of expectation, we have
\[
\lambda_{XY} = E[X Y] - m_X m_Y \tag{2.46b}
\]

The \( \text{cov}[X, Y] \), normalized with respect to the product \( \sigma_X \sigma_Y \) is termed as the **correlation coefficient** and we denote it by \( \rho \). That is,
\[
\rho_{XY} = \frac{E[XY] - m_X m_Y}{\sigma_X \sigma_Y} \tag{2.47}
\]

The correlation coefficient is a measure of dependency between the variables.

Suppose \( X \) and \( Y \) are independent. Then,
\[
E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) \, dx \, dy
\]
\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y f_x(x) f_y(y) \, dx \, dy
\]
\[
= \int_{-\infty}^{\infty} x f_x(x) \, dx \int_{-\infty}^{\infty} y f_y(y) \, dy = m_X m_Y
\]

2.53
Thus, we have $\lambda_{XY}$ (and $\rho_{XY}$) being zero. Intuitively, this result is appealing. Assume $X$ and $Y$ to be independent. When the joint experiment is performed many times, and given $X = x_i$, then $Y$ would occur sometimes positive with respect to $m_y$, and sometimes negative with respect to $m_y$. In the course of many trials of the experiments and with the outcome $X = x_i$, the sum of the numbers $x_i(y - m_y)$ would be very small and the quantity, \[
 \frac{\text{sum}}{\text{number of trials}},
\]
tends to zero as the number of trials keep increasing.

On the other hand, let $X$ and $Y$ be dependent. Suppose for example, the outcome $y$ is conditioned on the outcome $x$ in such a manner that there is a greater possibility of $(y - m_y)$ being of the same sign as $(x - m_x)$. Then we expect $\rho_{XY}$ to be positive. Similarly if the probability of $(x - m_x)$ and $(y - m_y)$ being of the opposite sign is quite large, then we expect $\rho_{XY}$ to be negative.

Taking the extreme case of this dependency, let $X$ and $Y$ be so conditioned that, given $X = x_i$, then $Y = \pm \alpha x_i$, $\alpha$ being constant. Then $\rho_{XY} = \pm 1$. That is, for $X$ and $Y$ be independent, we have $\rho_{XY} = 0$ and for the totally dependent $(y = \pm \alpha x)$ case, $|\rho_{XY}| = 1$. If the variables are neither independent nor totally dependent, then $\rho$ will have a magnitude between 0 and 1.

Two random variables $X$ and $Y$ are said to be orthogonal if $E[XY] = 0$. If $\rho_{XY}$ (or $\lambda_{XY}$) = 0, the random variables $X$ and $Y$ are said to be uncorrelated. That is, if $X$ and $Y$ are uncorrelated, then $\overline{XY} = \overline{X}\overline{Y}$. When the random variables are independent, they are uncorrelated. However, the fact that they are uncorrelated does not ensure that they are independent. As an example, let
\[ f_x(x) = \begin{cases} \frac{1}{\alpha}, & -\frac{\alpha}{2} < x < \frac{\alpha}{2}, \\ 0, & \text{otherwise} \end{cases} \]

and \( Y = X^2 \). Then,

\[
\overline{XY} = \overline{X^3} = \int_{-\infty}^{\infty} x^3 \frac{1}{\alpha} dx = \frac{1}{\alpha} \alpha^{\frac{3}{2}} \int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} x^3 dx = 0
\]

As \( \overline{X} = 0 \), \( \overline{XY} = 0 \) which means \( \lambda_{XY} = \overline{XY} - \overline{X} \overline{Y} = 0 \). But \( X \) and \( Y \) are not independent because, if \( X = x_1 \), then \( Y = x_1^2 \)!

Let \( Y \) be the linear combination of the two random variables \( X_1 \) and \( X_2 \). That is \( Y = k_1 X_1 + k_2 X_2 \) where \( k_1 \) and \( k_2 \) are constants. Let \( E[X_i] = m_i \), \( \sigma_{X_i}^2 = \sigma_i^2 \), \( i = 1, 2 \). Let \( \rho_{12} \) be the correlation coefficient between \( X_1 \) and \( X_2 \).

We will now relate \( \sigma_Y^2 \) to the known quantities.

\[
\sigma_Y^2 = E[Y^2] - [E[Y]]^2
\]

\[
E[Y] = k_1 m_1 + k_2 m_2
\]

\[
E[Y^2] = E[k_1 X_1^2 + k_2 X_2^2 + 2k_1 k_2 X_1 X_2]
\]

With a little manipulation, we can show that

\[
\sigma_Y^2 = k_1 \sigma_1^2 + k_2 \sigma_2^2 + 2 \rho_{12} k_1 k_2 \sigma_1 \sigma_2
\] (2.48a)

If \( X_1 \) and \( X_2 \) are uncorrelated, then

\[
\sigma_Y^2 = k_1 \sigma_1^2 + k_2 \sigma_2^2
\] (2.48b)

Note: Let \( X_1 \) and \( X_2 \) be uncorrelated, and let \( Z = X_1 + X_2 \) and \( W = X_1 - X_2 \). Then \( \sigma_Z^2 = \sigma_W^2 = \sigma_1^2 + \sigma_2^2 \). That is, the sum as well as the difference random variables have the same variance which is larger than \( \sigma_1^2 \) or \( \sigma_2^2 \)!

The above result can be generalized to the case of linear combination of \( n \) variables. That is, if
\[ Y = \sum_{i=1}^{n} k_i X_i, \text{ then} \]

\[ \sigma_Y^2 = \sum_{i=1}^{n} k_i^2 \sigma_i^2 + 2 \sum_{i<j} \sum_{j} k_i k_j \rho_{ij} \sigma_i \sigma_j \]

(2.49)

where the meaning of various symbols on the RHS of Eq. 2.49 is quite obvious.

We shall now give a few examples based on the theory covered so far in this section.

**Example 2.18**

Let a random variable \( X \) have the CDF:

\[ F_X(x) = \begin{cases} 
0, & x < 0 \\
\frac{x}{8}, & 0 \leq x \leq 2 \\
\frac{x^2}{16}, & 2 \leq x \leq 4 \\
1, & 4 \leq x 
\end{cases} \]

We shall find a) \( \mu_X \) and b) \( \sigma_X^2 \)

a) The given CDF implies the PDF:

\[ f_X(x) = \begin{cases} 
0, & x < 0 \\
\frac{1}{8}, & 0 \leq x \leq 2 \\
\frac{1}{8}x, & 2 \leq x \leq 4 \\
0, & 4 \leq x 
\end{cases} \]

Therefore,

\[ E[X] = \int_{0}^{2} \frac{1}{8}x \, dx + \int_{2}^{4} \frac{1}{8}x^2 \, dx \]

2.56
\[ \sigma^2_x = E[X^2] - [E[X]]^2 \]

\[ E[X^2] = \int_0^2 \frac{1}{8} x^2 \, dx + \int_2^4 \frac{1}{8} x^3 \, dx \]

\[ = \frac{1}{3} + \frac{15}{2} = \frac{47}{6} \]

\[ \sigma^2_x = \frac{47}{6} - \left( \frac{31}{12} \right)^2 = \frac{167}{144} \]

Example 2.19

Let \( Y = \cos \pi X \), where

\[ f_X(x) = \begin{cases} 1, & -\frac{1}{2} < x < \frac{1}{2} \\ 0, & \text{otherwise} \end{cases} \]

Let us find \( E[Y] \) and \( \sigma_Y^2 \).

From Eq. 2.38 we have,

\[ E[Y] = \frac{1}{\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos(\pi x) \, dx = \frac{2}{\pi} = 0.0636 \]

\[ E[Y^2] = \frac{1}{\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos^2(\pi x) \, dx = \frac{1}{2} = 0.5 \]

Hence \( \sigma_Y^2 = \frac{1}{2} - \frac{4}{\pi^2} = 0.96 \)

Example 2.20

Let \( X \) and \( Y \) have the joint PDF

\[ f_{X,Y}(x, y) = \begin{cases} x + y, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases} \]
Let us find  
\[
\begin{align*}
\text{a) } & E\left[X^{2}Y^{2}\right] \quad \text{and} \quad \text{b) } \rho_{XY}
\end{align*}
\]

\text{a) From Eq. 2.43, we have}
\[
E\left[X^{2}Y^{2}\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^{2}y^{2} f_{X,Y}(x, y) \, dx \, dy
\]
\[
= \int_{0}^{1} \int_{0}^{1} x^{2}(x + y) \, dx \, dy = \frac{17}{72}
\]

\text{b) To find } \rho_{XY}, \text{ we require } E[X], E[Y], \sigma_{x} \text{ and } \sigma_{y}. \text{ We can easily show that}
\[
E[X] = E[Y] = \frac{7}{12}, \quad \sigma_{x} = \sigma_{y} = \frac{11}{144} \quad \text{and} \quad E[X \cdot Y] = \frac{48}{144}
\]
\text{Hence } \rho_{XY} = -\frac{1}{11}

Another statistical average that will be found useful in the study of communication theory is the \textbf{conditional expectation}. The quantity,
\[
E\left[g(X) \mid Y = y\right] = \int_{-\infty}^{\infty} g(x) f_{X\mid Y}(x \mid y) \, dx
\]
(2.50)
is called the conditional expectation of \(g(X)\), given \(Y = y\). If \(g(X) = X\), then we have the \textbf{conditional mean}, namely, \(E[X \mid Y]\).
\[
E[X \mid Y = y] = E[X \mid Y] = \int_{-\infty}^{\infty} x f_{X\mid Y}(x \mid y) \, dx
\]
(2.51)

Similarly, we can define the conditional variance etc. We shall illustrate the calculation of conditional mean with the help of an example.

**Example 2.21**

Let the joint PDF of the random variables \(X\) and \(Y\) be
\[ f_{x,y}(x, y) = \begin{cases} \frac{1}{x}, & 0 < x < 1, \ 0 < y < x, \\ 0, & \text{outside} \end{cases} \]

Let us compute \( E[X \mid Y] \).

To find \( E[X \mid Y] \), we require the conditional PDF, 
\[ f_{x\mid y}(x \mid y) = \frac{f_{x,y}(x, y)}{f_r(y)} \]

\[ f_r(y) = \int_{y}^{1} \frac{1}{x} \, dx = -\ln y, \ 0 < y < 1 \]

\[ f_{x\mid y}(x \mid y) = \frac{\frac{1}{x}}{-\ln y} = -\frac{1}{x \ln y}, \ y < x < 1 \]

Hence \( E(X \mid Y) = \int_{y}^{1} x \left[ -\frac{1}{x \ln y} \right] \, dx \)

\[ = \frac{y - 1}{\ln y} \]

Note that \( E[X \mid Y = y] \) is a function of \( y \).

### 2.6 Some Useful Probability Models

In the concluding section of this chapter, we shall discuss certain probability distributions which are encountered quite often in the study of communication theory. We will begin our discussion with discrete random variables.

#### 2.6.1. Discrete random variables

i) Binomial:

Consider a random experiment in which our interest is in the occurrence or non-occurrence of an event \( A \). That is, we are interested in the event \( A \) or its complement, say, \( B \). Let the experiment be repeated \( n \) times and let \( p \) be the
probability of occurrence of \( A \) on each trial and the trials are independent. Let \( X \) denote random variable, ‘number of occurrences of \( A \) in \( n \) trials’. \( X \) can be equal to 0, 1, 2, \ldots, \( n \). If we can compute \( P[X = k], k = 0, 1, \ldots, n \), then we can write \( f_X(x) \).

Taking a special case, let \( n = 5 \) and \( k = 3 \). The sample space (representing the outcomes of these five repetitions) has 32 sample points, say, \( s_1, \ldots, s_{32} \). The sample point \( s_1 \) could represent the sequence \( ABBBB \). The sample points such as \( ABAAB, AAABB \) etc. will map into real number 3 as shown in the Fig. 2.18. (Each sample point is actually an element of the five dimensional Cartesian product space).

\[ P(ABAAB) = P(A)P(B)P(A)P(A)P(B), \text{ as trials are independent.} \]

\[ = p(1 - p) p^2 (1 - p) = p^3 (1 - p)^2 \]

There are \( \binom{5}{3} = 10 \) sample points for which \( X(s) = 3 \).

In other words, for \( n = 5, k = 3 \),

\[ P[X = 3] = \binom{5}{3} p^3 (1 - p)^2 \]

Generalizing this to arbitrary \( n \) and \( k \), we have the binomial density, given by
\[ f_X(x) = \sum_{i=0}^{n} P_i \delta(x - i) \quad (2.52) \]

where \( P_i = \binom{n}{i} p^i (1-p)^{n-i} \)

As can be seen, \( f_X(x) \geq 0 \) and

\[
\int_{-\infty}^{\infty} f_X(x) \, dx = \sum_{i=0}^{n} P_i = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \]

\[
= \left[ (1-p) + p \right]^n = 1
\]

It is left as an exercise to show that \( E[X] = np \) and \( \sigma_X^2 = np(1-p) \). (Though the formulae for the mean and the variance of a binomial PDF are simple, the algebra to derive them is laborious).

We write \( X \) is \( b(n, p) \) to indicate \( X \) has a binomial PDF with parameters \( n \) and \( p \) defined above.

The following example illustrates the use of binomial PDF in a communication problem.

**Example 2.22**

A digital communication system transmits binary digits over a noisy channel in blocks of 16 digits. Assume that the probability of a binary digit being in error is 0.01 and that errors in various digit positions within a block are statistically independent.

i) Find the expected number of errors per block

ii) Find the probability that the number of errors per block is greater than or equal to 3.

Let \( X \) be the random variable representing the number of errors per block. Then \( X \) is \( b(16, 0.01) \).

i) \( E[X] = np = 16 \times 0.01 = 0.16 \);
ii) \[ P(X \geq 3) = 1 - P[X \leq 2] \]
\[ = 1 - \sum_{i=0}^{2} \left( \frac{16}{i!} \right) (0.1)^i (1-p)^{16-i} \]
\[ = 0.002 \]

Exercise 2.8

Show that for the example 2.22, Chebyshev inequality results in \( P[|X| \geq 3] \leq 0.0196 = 0.02 \). Note that Chebyshev inequality is not very tight.

ii) Poisson:

A random variable \( X \) which takes on only integer values is Poisson distributed, if

\[ f_X(x) = \sum_{m=0}^{\infty} \delta(x-m) \frac{\lambda^m e^{-\lambda}}{m!} \quad (2.53) \]

where \( \lambda \) is a positive constant.

Evidently \( f_X(x) \geq 0 \) and \( \int_{-\infty}^{\infty} f_X(x) \, dx = 1 \) because \( \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^\lambda \).

We will now show that

\[ E[X] = \lambda = \sigma_X^2 \]

Since,

\[ e^\lambda = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}, \quad \text{we have} \]

\[ \frac{d(e^\lambda)}{d\lambda} = e^\lambda = \sum_{m=0}^{\infty} \frac{m \lambda^{m-1}}{m!} = \frac{1}{\lambda} \sum_{m=1}^{\infty} m \frac{\lambda^m}{m!} \]

\[ E[X] = \sum_{m=1}^{\infty} \frac{m \lambda^m}{m!} e^{-\lambda} = \lambda e^\lambda e^{-\lambda} = \lambda \]

Differentiating the series again, we obtain,
\[ E[X^2] = \lambda^2 + \lambda. \text{ Hence } \sigma_X^2 = \lambda. \]

### 2.6.2 Continuous random variables

i) **Uniform:**

A random variable \( X \) is said to be uniformly distributed in the interval \( a \leq x \leq b \) if,

\[
f_X(x) = \begin{cases} 
  \frac{1}{b-a}, & a \leq x \leq b \\
  0, & \text{elsewhere} 
\end{cases}
\]

A plot of \( f_X(x) \) is shown in Fig.2.19.

![Uniform PDF](image)

Fig.2.19: Uniform PDF

It is easy show that

\[ E[X] = \frac{a + b}{2} \text{ and } \sigma_X^2 = \frac{(b-a)^2}{12} \]

Note that the variance of the uniform PDF depends only on the width of the interval \((b-a)\). Therefore, whether \( X \) is uniform in \((-1, 1)\) or \((2, 4)\), it has the same variance, namely \( \frac{1}{3} \).

ii) **Rayleigh:**

An RV \( X \) is said to be Rayleigh distributed if,
\[ f_x(x) = \begin{cases} \frac{x}{b} \exp \left( -\frac{x^2}{2b} \right), & x \geq 0 \\ 0, & \text{elsewhere} \end{cases} \] (2.55)

where \( b \) is a positive constant,

A typical sketch of the Rayleigh PDF is given in Fig.2.20. (\( f_r(r) \) of example 2.12 is Rayleigh PDF.)

![Rayleigh PDF](image)

Rayleigh PDF frequently arises in radar and communication problems. We will encounter it later in the study of narrow-band noise processes.
iii) Gaussian

By far the most widely used PDF, in the context of communication theory is the Gaussian (also called normal) density, specified by

$$f_x(x) = \frac{1}{\sqrt{2\pi \sigma_x}} \exp\left[ -\frac{(x - m_x)^2}{2\sigma_x^2} \right], \quad -\infty < x < \infty$$  \hspace{1cm} (2.56)

where $m_x$ is the mean value and $\sigma_x^2$ the variance. That is, the Gaussian PDF is completely specified by the two parameters, $m_x$ and $\sigma_x^2$. We use the symbol $N(m_x, \sigma_x^2)$ to denote the Gaussian density\(^1\). In appendix A2.3, we show that $f_x(x)$ as given by Eq. 2.56 is a valid PDF.

As can be seen from the Fig. 2.21, The Gaussian PDF is symmetrical with respect to $m_x$.

---

\(^1\) In this notation, $N(0, 1)$ denotes the Gaussian PDF with zero mean and unit variance. Note that if $X$ is $N(m_x, \sigma_x^2)$, then $Y = \left( \frac{X - m_x}{\sigma_x} \right)$ is $N(0, 1)$.
Fig. 2.21: Gaussian PDF

Hence \( F_X(m_X) = \int_{-\infty}^{m_X} f_X(x) \, dx = 0.5 \)

Consider \( P[X \geq a] \). We have,

\[
P[X \geq a] = \int_{a}^{\infty} \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left[-\frac{(x - m_X)^2}{2\sigma_X^2}\right] \, dx
\]

This integral cannot be evaluated in closed form. By making a change of variable

\[ z = \left(\frac{x - m_X}{\sigma_X}\right), \]

we have

\[
P[X \geq a] = \int_{a - m_X / \sigma_X}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz
\]

\[ = Q\left(\frac{a - m_X}{\sigma_X}\right) \]

where \( Q(y) = \int_{y}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \, dx \) \hspace{1cm} (2.57)

Note that the integrand on the RHS of Eq. 2.57 is \( N(0, 1) \).

\( Q(\ ) \) function table is available in most of the text books on communication theory as well as in standard mathematical tables. A small list is given in appendix A2.2 at the end of the chapter.

2.66
The importance of Gaussian density in communication theory is due to a theorem called central limit theorem which essentially states that:

If the RV \( X \) is the weighted sum of \( N \) independent random components, where each component makes only a small contribution to the sum, then \( F_x(x) \) approaches Gaussian as \( N \) becomes large, regardless of the distribution of the individual components.

For a more precise statement and a thorough discussion of this theorem, you may refer [1-3]. The electrical noise in a communication system is often due to the cumulative effects of a large number of randomly moving charged particles, each particle making an independent contribution of the same amount, to the total. Hence the instantaneous value of the noise can be fairly adequately modeled as a Gaussian variable. We shall develop Gaussian random processes in detail in Chapter 3 and, in Chapter 7, we shall make use of this theory in our studies on the noise performance of various modulation schemes.

Example 2.23

A random variable \( Y \) is said to have a log-normal PDF if \( X = \ln Y \) has a Gaussian (normal) PDF. Let \( Y \) have the PDF, \( f_y(y) \) given by,

\[
f_y(y) = \begin{cases} \frac{1}{\sqrt{2\pi} y \beta} \exp \left( -\frac{(\ln y - \alpha)^2}{2 \beta^2} \right), & y \geq 0 \\ 0, & \text{otherwise} \end{cases}
\]

where \( \alpha \) and \( \beta \) are given constants.

a) Show that \( Y \) is log-normal
b) Find \( E(Y) \)
c) If \( m \) is such that \( F_Y(m) = 0.5 \), find \( m \).

a) Let \( X = \ln Y \) or \( x = \ln y \) (Note that the transformation is one-to-one)
\[
\frac{d x}{d y} = \frac{1}{y} \rightarrow |J| = \frac{1}{y}
\]

Also as \( y \rightarrow 0 \), \( x \rightarrow -\infty \) and as \( y \rightarrow \infty \), \( x \rightarrow \infty \)

Hence \( f_x(x) = \frac{1}{\sqrt{2\pi} \beta} \exp \left[ - \frac{(x - \alpha)^2}{2 \beta^2} \right], \quad -\infty < x < \infty \)

Note that \( X \) is \( N(\alpha, \beta^2) \)

b) \( \bar{Y} = E[e^X] = \frac{1}{\sqrt{2\pi} \beta} \int_{-\infty}^{\infty} e^x \left[ e^{-\frac{(x - \alpha)^2}{2 \beta^2}} \right] dx \)

\[
= e^{\alpha + \frac{\beta^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \beta} e^{-\frac{(x - (\alpha + \beta^2))^2}{2 \beta^2}} dx
\]

As the bracketed quantity being the integral of a Gaussian PDF between the limits \((-\infty, \infty)\) is 1, we have

\( \bar{Y} = e^{\alpha + \frac{\beta^2}{2}} \)

c) \( P[Y \leq m] = P[X \leq \ln m] \)

Hence if \( P[Y \leq m] = 0.5 \), then \( P[X \leq \ln m] = 0.5 \)

That is, \( \ln m = \alpha \) or \( m = e^{\alpha} \)

iv) Bivariate Gaussian

As an example of a two dimensional density, we will consider the bivariate Gaussian PDF, \( f_{x,y}(x, y), -\infty < x, \ y < \infty \) given by,

\[
f_{x,y}(x, y) = \frac{1}{k_1} \exp \left\{ - \frac{1}{k_2} \left[ \frac{(x - m_x)^2}{\sigma_x^2} + \frac{(y - m_y)^2}{\sigma_y^2} - 2 \rho \frac{(x - m_x)(y - m_y)}{\sigma_x \sigma_y} \right] \right\} \quad (2.58)
\]

where,

\[
k_1 = 2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}
\]

2.68
\[ k_2 = 2(1 - \rho^2) \]

\[ \rho = \text{Correlation coefficient between } X \text{ and } Y \]

The following properties of the bivariate Gaussian density can be verified:

**P1)** If \( X \) and \( Y \) are jointly Gaussian, then the marginal density of \( X \) or \( Y \) is Gaussian; that is, \( X \sim N(m_X, \sigma_X^2) \) and \( Y \sim N(m_Y, \sigma_Y^2) \).

**P2)** \( f_{X,Y}(x, y) = f_X(x) f_Y(y) \) iff \( \rho = 0 \)

That is, if the Gaussian variables are uncorrelated, then they are independent. That is not true, in general, with respect to non-Gaussian variables (we have already seen an example of this in Sec. 2.5.2).

**P3)** If \( Z = \alpha X + \beta Y \) where \( \alpha \) and \( \beta \) are constants and \( X \) and \( Y \) are jointly Gaussian, then \( Z \) is Gaussian. Therefore \( f_Z(z) \) can be written after computing \( m_Z \) and \( \sigma_Z^2 \) with the help of the formulae given in section 2.5.

Figure 2.22 gives the plot of a bivariate Gaussian PDF for the case of \( \rho = 0 \) and \( \sigma_X = \sigma_Y \).

\[ \text{Note that the converse is not necessarily true. Let } f_X \text{ and } f_Y \text{ be obtained from } f_{X,Y} \text{ and let } f_X \text{ and } f_Y \text{ be Gaussian. This does not imply } f_{X,Y} \text{ is jointly Gaussian, unless } X \text{ and } Y \text{ are independent. We can construct examples of a joint PDF } f_{X,Y}, \text{ which is not Gaussian but results in } f_X \text{ and } f_Y \text{ that are Gaussian.} \]
Fig. 2.22: Bivariate Gaussian PDF ($\sigma_X = \sigma_Y$ and $\rho = 0$)

For $\rho = 0$ and $\sigma_X = \sigma_Y$, $f_{X,Y}$ resembles a (temple) bell, with, of course, the striker missing! For $\rho \neq 0$, we have two cases (i) $\rho$, positive and (ii) $\rho$, negative. If $\rho > 0$, imagine the bell being compressed along the $X = -Y$ axis so that it elongates along the $X = Y$ axis. Similarly for $\rho < 0$.

**Example 2.24**

Let $X$ and $Y$ be jointly Gaussian with $\overline{X} = \overline{Y} = 1$, $\sigma^2_X = \sigma^2_Y = 1$ and $\rho_{X,Y} = -\frac{1}{2}$. Let us find the probability of $(X, Y)$ lying in the shaded region $\mathcal{D}$ shown in Fig. 2.23.
Let $\mathcal{A}$ be the shaded region shown in Fig. 2.24(a) and $\mathcal{B}$ be the shaded region in Fig. 2.24(b).

The required probability = $P[(x, y) \in \mathcal{A}] - P[(x, y) \in \mathcal{B}]$

For the region $\mathcal{A}$, we have $y \geq -\frac{1}{2}x + 1$ and for the region $\mathcal{B}$, we have $y \geq -\frac{1}{2}x + 2$. Hence the required probability is, 2.71
\[
P\left[ \frac{Y + X}{2} \geq 1 \right] - P\left[ \frac{Y + X}{2} \geq 2 \right]
\]

Let \( Z = Y + \frac{X}{2} \)

Then \( Z \) is Gaussian with the parameters,

\[
Z = \bar{Y} + \frac{1}{2} \bar{X} = -\frac{1}{2}
\]

\[
\sigma^2_Z = \frac{1}{4} \sigma^2_X + \sigma^2_Y + 2 \cdot \frac{1}{2} \rho_{XY}
\]

\[
\sigma^2_Z = \frac{1}{4} + 2 \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{4}
\]

That is, \( Z \) is \( N\left( -\frac{1}{2}, \frac{3}{4} \right) \). Then \( W = \frac{Z + 1/2}{\sqrt{\frac{3}{4}}} \) is \( N(0, 1) \)

\[
P[Z \geq 1] = P[W \geq \sqrt{3}]
\]

\[
P[Z \geq 2] = P[W \geq \frac{5}{\sqrt{3}}]
\]

Hence the required probability \( = Q\left(\sqrt{3}\right) - Q\left(\frac{5}{\sqrt{3}}\right) = (0.04 - 0.001) = 0.039 \)
Exercise 2.10

X and Y are independent, identically distributed (iid) random variables, each being $N(0, 1)$. Find the probability of $X, Y$ lying in the region $\mathcal{A}$ shown in Fig. 2.25.

![Region A of Exercise 2.10](image)

Note: It would be easier to calculate this kind of probability, if the space is a product space. From example 2.12, we feel that if we transform $(X, Y)$ into $(Z, W)$ such that $Z = X + Y$, $W = X - Y$, then the transformed space $\mathcal{B}$ would be square. Find $f_{Z,W}(z, w)$ and compute the probability $(Z, W) \in B$. 

2.73
Exercise 2.11

Two random variables $X$ and $Y$ are obtained by means of the transformation given below.

$$X = (-2 \log_e U_1)^{1/2} \cos (2\pi U_2)$$ \hspace{1cm} (2.59a)

$$Y = (-2 \log_e U_1)^{1/2} \sin (2\pi U_2)$$ \hspace{1cm} (2.59b)

$U_1$ and $U_2$ are independent random variables, uniformly distributed in the range $0 < u_1, u_2 < 1$. Show that $X$ and $Y$ are independent and each is $N(0, 1)$.

Hint: Let $X_1 = -2 \log_e U_1$ and $Y_1 = \sqrt{X_1}$

Show that $Y_1$ is Rayleigh. Find $f_{X,Y}(x, y)$ using $X = Y_1 \cos \Theta$ and $Y = Y_1 \sin \Theta$, where $\Theta = 2\pi U_2$.

Note: The transformation given by Eq. 2.59 is called the Box-Muller transformation and can be used to generate two Gaussian random number sequences from two independent uniformly distributed (in the range 0 to 1) sequences.
Appendix A2.1

Proof of Eq. 2.34

The proof of Eq. 2.34 depends on establishing a relationship between the
differential area \(dz\,dw\) in the \(z-w\) plane and the differential area \(dx\,dy\) in
the \(x-y\) plane. We know that

\[
f_{z,w}(z, w) \, dz \, dw = P[z < Z \leq z + d z, w < W \leq w + d w]
\]

If we can find \(dx\,dy\) such that

\[
f_{z,w}(z, w) \, dz \, dw = f_{x,y}(x, y) \, dx \, dy,
\]
then \(f_{z,w}\) can be found. (Note that the variables \(x\) and \(y\) can be replaced by their inverse transformation
quantities, namely, \(x = g^{-1}(z, w)\) and \(y = h^{-1}(z, w)\))

Let the transformation be one-to-one. (This can be generalized to the case of
one-to-many.) Consider the mapping shown in Fig. A2.1.

Infinitesimal rectangle \(ABCD\) in the \(z-w\) plane is mapped into the
parallelogram in the \(x-y\) plane. (We may assume that the vertex \(A\) transforms
to \(A'\), \(B\) to \(B'\) etc.) We shall now find the relation between the differential area of
the rectangle and the differential area of the parallelogram.

Fig. A2.1: A typical transformation between the \((x-y)\) plane and \((z-w)\) plane
Consider the parallelogram shown in Fig. A2.2, with vertices \( P_1, P_2, P_3 \) and \( P_4 \).

![Fig. A2.2: Typical parallelogram](image)

Let \((x, y)\) be the co-ordinates of \( P_1 \). Then the \( P_2 \) and \( P_3 \) are given by

\[
P_2 = \left( x + \frac{\partial g^{-1}}{\partial z} \, dz, \, y + \frac{\partial h^{-1}}{\partial z} \, dz \right)
\]

\[
= \left( x + \frac{\partial x}{\partial z} \, dz, \, y + \frac{\partial y}{\partial z} \, dz \right)
\]

\[
P_3 = \left( x + \frac{\partial x}{\partial w} \, dw, \, y + \frac{\partial y}{\partial w} \, dw \right)
\]

Consider the vectors \( \mathbf{V}_1 \) and \( \mathbf{V}_2 \) shown in the Fig. A2.2 where

\[
\mathbf{V}_1 = (P_2 - P_1) \quad \text{and} \quad \mathbf{V}_2 = (P_3 - P_1).
\]

That is,

\[
\mathbf{V}_1 = \frac{\partial x}{\partial z} \, dz \, \mathbf{i} + \frac{\partial y}{\partial z} \, dz \, \mathbf{j}
\]

\[
\mathbf{V}_2 = \frac{\partial x}{\partial w} \, dw \, \mathbf{i} + \frac{\partial y}{\partial w} \, dw \, \mathbf{j}
\]

where \( \mathbf{i} \) and \( \mathbf{j} \) are the unit vectors in the appropriate directions. Then, the area \( A \) of the parallelogram is,

\[
A = |\mathbf{V}_1 \times \mathbf{V}_2|
\]

2.76
As \( \mathbf{i} \times \mathbf{i} = 0, \mathbf{j} \times \mathbf{j} = 0, \) and \( \mathbf{i} \times \mathbf{j} = -(\mathbf{j} \times \mathbf{i}) = \mathbf{k} \) where \( \mathbf{k} \) is the unit vector perpendicular to both \( \mathbf{i} \) and \( \mathbf{j} \), we have

\[
|\mathbf{V}_1 \times \mathbf{V}_2| = \left| \frac{\partial x}{\partial z} \frac{\partial y}{\partial w} - \frac{\partial y}{\partial z} \frac{\partial x}{\partial w} \right| dz \, dw
\]

\[
A = \left| J \left( \frac{x, y}{z, w} \right) \right| dz \, dw
\]

That is,

\[
f_{Z,W}(z, w) = f_{X,Y}(x, y) \left| J \left( \frac{x, y}{z, w} \right) \right|
\]

\[
= \frac{f_{X,Y}(x, y)}{\left| J \left( \frac{z, w}{x, y} \right) \right|}.
\]
Appendix A2.2

\(Q(\alpha)\) Function Table

\[
Q(\alpha) = \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx
\]

It is sufficient if we know \(Q(\alpha)\) for \(\alpha \geq 0\), because \(Q(-\alpha) = 1 - Q(\alpha)\). Note that \(Q(0) = 0.5\).

<table>
<thead>
<tr>
<th>(y)</th>
<th>(Q(y))</th>
<th>(y)</th>
<th>(Q(y))</th>
<th>(y)</th>
<th>(Q(y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.4801</td>
<td>1.05</td>
<td>0.1469</td>
<td>2.10</td>
<td>0.0179</td>
</tr>
<tr>
<td>0.10</td>
<td>0.4602</td>
<td>1.10</td>
<td>0.1357</td>
<td>2.20</td>
<td>0.0139</td>
</tr>
<tr>
<td>0.15</td>
<td>0.4405</td>
<td>1.15</td>
<td>0.1251</td>
<td>2.30</td>
<td>0.0107</td>
</tr>
<tr>
<td>0.20</td>
<td>0.4207</td>
<td>1.20</td>
<td>0.1151</td>
<td>2.40</td>
<td>0.0082</td>
</tr>
<tr>
<td>0.25</td>
<td>0.4013</td>
<td>1.25</td>
<td>0.0156</td>
<td>2.50</td>
<td>0.0062</td>
</tr>
<tr>
<td>0.30</td>
<td>0.3821</td>
<td>1.30</td>
<td>0.0968</td>
<td>2.60</td>
<td>0.0047</td>
</tr>
<tr>
<td>0.35</td>
<td>0.3632</td>
<td>1.35</td>
<td>0.0885</td>
<td>2.70</td>
<td>0.0035</td>
</tr>
<tr>
<td>0.40</td>
<td>0.3446</td>
<td>1.40</td>
<td>0.0808</td>
<td>2.80</td>
<td>0.0026</td>
</tr>
<tr>
<td>0.45</td>
<td>0.3264</td>
<td>1.45</td>
<td>0.0735</td>
<td>2.90</td>
<td>0.0019</td>
</tr>
<tr>
<td>0.50</td>
<td>0.3085</td>
<td>1.50</td>
<td>0.0668</td>
<td>3.00</td>
<td>0.0013</td>
</tr>
<tr>
<td>0.55</td>
<td>0.2912</td>
<td>1.55</td>
<td>0.0606</td>
<td>3.10</td>
<td>0.0010</td>
</tr>
<tr>
<td>0.60</td>
<td>0.2743</td>
<td>1.60</td>
<td>0.0548</td>
<td>3.20</td>
<td>0.00069</td>
</tr>
<tr>
<td>0.65</td>
<td>0.2578</td>
<td>1.65</td>
<td>0.0495</td>
<td>3.30</td>
<td>0.00048</td>
</tr>
<tr>
<td>0.70</td>
<td>0.2420</td>
<td>1.70</td>
<td>0.0446</td>
<td>3.40</td>
<td>0.00034</td>
</tr>
<tr>
<td>0.75</td>
<td>0.2266</td>
<td>1.75</td>
<td>0.0401</td>
<td>3.50</td>
<td>0.00023</td>
</tr>
<tr>
<td>0.80</td>
<td>0.2119</td>
<td>1.80</td>
<td>0.0359</td>
<td>3.60</td>
<td>0.00016</td>
</tr>
<tr>
<td>0.85</td>
<td>0.1977</td>
<td>1.85</td>
<td>0.0322</td>
<td>3.70</td>
<td>0.00010</td>
</tr>
<tr>
<td>0.90</td>
<td>0.1841</td>
<td>1.90</td>
<td>0.0287</td>
<td>3.80</td>
<td>0.00007</td>
</tr>
<tr>
<td>0.95</td>
<td>0.1711</td>
<td>1.95</td>
<td>0.0256</td>
<td>3.90</td>
<td>0.00005</td>
</tr>
<tr>
<td>1.00</td>
<td>0.1587</td>
<td>2.00</td>
<td>0.0228</td>
<td>4.00</td>
<td>0.00003</td>
</tr>
</tbody>
</table>

\(y\) | \(Q(y)\) | \(y\) | \(Q(y)\) | \(y\) | \(Q(y)\) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(10^{-3})</td>
<td>3.10</td>
<td>(10^{-3})</td>
<td>3.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>(10^{-4})</td>
<td>3.70</td>
<td>(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10^{-5})</td>
<td>3.90</td>
<td>(10^{-6})</td>
<td>4.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Note that some authors use $\text{erfc}(\cdot)$, the *complementary error function* which is given by

$$\text{erfc}(\alpha) = 1 - \text{erf}(\alpha) = \frac{2}{\sqrt{\pi}} \int_{\alpha}^{\infty} e^{-\beta^2} \, d\beta$$

and the *error function*, $\text{erf}(\alpha) = \frac{2}{\sqrt{\pi}} \int_{0}^{\alpha} e^{-\beta^2} \, d\beta$

Hence $Q(\alpha) = \frac{1}{2} \text{erfc} \left( \frac{\alpha}{\sqrt{2}} \right)$. 

2.79
Appendix A2.3

Proof that \( N(\mu, \sigma^2) \) is a valid PDF

We will show that \( f_x(x) \) as given by Eq. 2.56, is a valid PDF by establishing \( \int_{-\infty}^{\infty} f_x(x) \, dx = 1 \). (Note that \( f_x(x) \geq 0 \) for \(-\infty < x < \infty\)).

Let \( l = \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} \, dv = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \, dy \).

Then, \( l^2 = \left[ \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} \, dv \right] \left[ \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \, dy \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{v^2+y^2}{2}} \, dv \, dy \).

Let \( v = r \cos \theta \) and \( y = r \sin \theta \). Then, \( r = \sqrt{v^2+y^2} \) and \( \theta = \tan^{-1}\left(\frac{y}{v}\right) \), and \( dx \, dy = r \, dr \, d\theta \). (Cartesian to Polar coordinate transformation).

\[
\begin{align*}
l^2 &= \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} \, r \, dr \, d\theta \\
&= 2\pi \quad \text{or} \quad l = \sqrt{2\pi}
\end{align*}
\]

That is, \( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} \, dv = 1 \) \hspace{1cm} (A2.3.1)

Let \( v = \frac{x - \mu}{\sigma_x} \) \hspace{1cm} (A2.3.2)

Then, \( dv = \frac{dx}{\sigma_x} \) \hspace{1cm} (A2.3.3)

Using Eq. A2.3.2 and Eq. A2.3.3 in Eq. A2.3.1, we have the required result.
References

