

Chapter 1

A Review of Probability and Stochastic Processes

The seminal studies about probability go back to the 17th century with Blaise Pascal (1623–1662), Pierre de Fermat (1601–1665), Jacques Bernoulli (1654–1705) and Abraham de Moivre (1667–1754). Today, probability and random processes (or stochastic processes) are the basis for the study of many areas, including Electrical Engineering and, particularly, communications theory. That is the reason for including disciplines on the subject in the regular curriculum of such courses. Moreover, in communication systems, most of the signals are of random nature, demanding the use of probabilistic tools for their characterization and for supporting system design and assessment. In this chapter, mostly inspired in the book by Alberto Leon-Garcia [9], we review basic concepts about probability and random processes. The chapter is intended to allow for the reader to have prompt access to these concepts when studying specific subjects on digital transmission. We start by reviewing some concepts and properties of the set theory, aiming at using them to define probability and to help with the solutions of problems. In the sequel, basic combinatorial theory is presented as an additional tool for probability calculations based on counting methods. An important rule of probabilities, known as the Bayes rule is then presented. Random variables and averages of random variables are also discussed, followed by the central limit theorem. Confidence interval analysis is then studied. Its use is related to the analysis of the precision associated to averages obtained from sample data. Random number generation is also discussed as a means for computer simulating random phenomena that frequently appear in digital transmission systems. Random processes are defined and the analysis of averages and filtering of stochastic processes are presented. The chapter ends with the investigation of the thermal noise, one of the main random processes encountered in communication systems.

1.1 Set Theory Basics

Before starting the specific topic of this section, let us have a look at some terms that will appear frequently throughout the chapter. We do this by means of an example, where the terms we are interested in are in italics: let the *experiment* of tossing a coin and observing the *event* corresponding to the coin face shown. The possible

results will be the *sample points* head (H) and tail (T). These two possible results compose the *sample space*. Now, using the same *experiment* let us define another *event* corresponding to the occurrence of heads in the first two out of three consecutive tosses. The *sample space* is now composed by $2^3 = 8$ *results*: HHH, HHT, HTH, HTT, THH, THT, TTH, and TTT, from where we notice that the defined event occurs two times, corresponding to the *sample points* HHH and HHT.

From this simple example we can conclude that a given experiment can produce different results and have different sample spaces, depending on how the event we are interested in is defined.

Other two common terms in the study of probability and stochastic processes are *deterministic event* and *random event*. If the occurrence of an event is temporal, we say that it is deterministic if we have no doubt about its result in any time. A deterministic event can be described by an expression. For example, if $x(t) = A \cos(2\pi ft)$ describes the time evolution of a voltage waveform, the precise time when, for example, $x(t) = A/2$ can be determined. A random event, instead, can not be described by a closed expression, since always there is some degree of uncertainty about its occurrence. For example, the amount of rain in a given day of the year is a random event that can not be precisely estimated.

Fortunately, random events show some sort of regularity that allows for some degree of confidence about them. This is indeed the objective of the study of probabilistic models: to analyze the regularities or patterns of random events in order to characterize them and infer about them.

Somebody once said: “probability is a toll that permits the greatest possible degree of confidence about events that are inherently uncertain”.

1.1.1 The Venn Diagram and Basic Set Operations

Set theory deals with the pictorial representation, called *Venn diagram*, of experiments, events, sample spaces and results, and with the mathematical tools behind this representation. The main objective is to model real problems in a way that they can be easily solved.

We start by reviewing the above-mentioned pictorial representation and some basic set operations. Let the sample space of a given experiment be represented by the letter S , which corresponds to the *certain* event. Let the events A and B defined as the appearance of some specific set of results in S . In Fig. 1.1 we have the corresponding Venn diagram for S , A , B and for other events defined from operations involving S , A and B . The complement of A , or \bar{A} , is the set composed by all elements in S , except the elements in A , that is $\bar{A} = S - A$. The complement \bar{A} represents the event that A did not occur. The union of the events A and B is the set $A \cup B$ or $A + B$ and it is formed by the elements in A plus the elements in B . It represents the event that either A or B or both occurred. The intersection of the events A and B is the set $A \cap B$ or $A \cdot B$, or simply AB and it is formed by the

elements that belongs to both A and B . It represents the event that both A and B occurred. In the rightmost diagram in Fig. 1.1 we have *mutually exclusive* events for which $A \cap B = \emptyset$, where \emptyset represents the empty set or the impossible event.

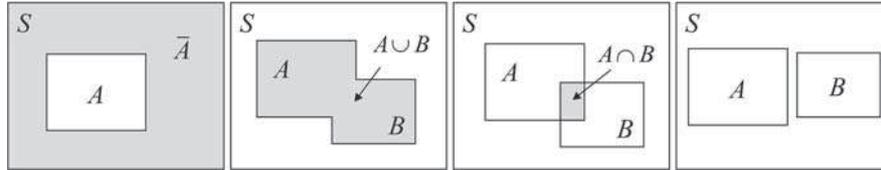


Fig. 1.1 Set representation and basic set operations

1.1.2 Other Set Operations and Properties

We list below the main operations involving sets. Most of them have specific names and proofs or basic explanatory statements. We omit this information, aiming only to review such operations. For a more applied and complete treatment on the subject, see [9–15].

1. $A \cup B = B \cup A$ and $A \cap B = B \cap A$
2. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
3. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
4. $\overline{\bigcap_i A_i} = \bigcup_i \overline{A_i}$ and $\overline{\bigcup_i A_i} = \bigcap_i \overline{A_i}$
5. If $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $\overline{A \cup (B \cap C)} = (\overline{A \cup B}) \cap (\overline{A \cup C})$ (1.1)
6. $A - B = A \cap \overline{B}$
7. If $A \subset B$, then $\overline{A} \supset \overline{B}$ or $\overline{B} \subset \overline{A}$
8. $A \cup \emptyset = A$ and $A \cap \emptyset = \emptyset$
9. $A \cup S = S$ and $A \cap S = A$
10. $A = (A \cap B) \cup (A \cap \overline{B})$,

where $A \subset B$ means that B contains A and $B \supset A$ means that A contains B .

1.2 Definitions and Axioms of Probability

Generally speaking, probability is a number $P[A]$ associated to the information on how likely an event A can occur, where $0 \leq P[A] \leq 1$. Three formal definitions of probability are the relative frequency, the axiomatic and the classic.

In the relative frequency approach, the probability of occurrence of an event A is given in the limit by

$$P[A] = \lim_{N \rightarrow \infty} \frac{n_A}{N}, \quad (1.2)$$

where n_A is the number of occurrences of event A and N is the number of trials or, equivalently, is the number of times that the experiment is repeated. It turns out that the greater the value of N , the more the result will approach the true probability of occurrence of A .

The axiomatic definition is based on the axioms: $0 \leq P[A]$; $P[S] = 1$; if $A \cap B \cap C \cap \dots = \emptyset$, then $P[A \cup B \cup C \cup \dots] = P[A] + P[B] + P[C] + \dots$. From these axioms are derived some important properties, as shown by:

1. $P[\bar{A}] = 1 - P[A]$
2. $P[A] \leq 1$
3. $P[\emptyset] = 0$
4. If $A \subset B$, then $P[A] \leq P[B]$
5. $P\left[\bigcup_{k=1}^n A_k\right] = \sum_{j=1}^n P[A_j] - \sum_{j < k} P[A_j \cap A_k] + \dots + (-1)^{n+1} P[A_1 \cap \dots \cap A_n]$.

The last property in (1.3) gives rise to an important axiom called *union bound*. If we do not know *a priori* if the events are or are not mutually exclusive, we can state that the probability of occurrence of the union of events will be less than or equal to the sum of the probabilities of occurrence of the individual events:

$$P\left[\bigcup_{k=1}^n A_k\right] \leq \sum_{k=1}^n P[A_k]. \quad (1.4)$$

In its classic definition, the probability of occurrence of an event A is determined without experimentation and is given by

$$P[A] = \frac{n_A}{N}, \quad (1.5)$$

where n_A is the number of favorable occurrences of the event A and N is the total number of possible results.

Example 1.1 – A cell in a cellular communication system has 5 channels that can be free or busy. The sample space is composed by $2^5 = 32$ combinations of the possible status for the 5 channels. Representing a free channel by a “0” and a busy channel by a “1”, we shall have the following sample space, where each five-element column is associated to one sample point:

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 1 1
0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1
0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1

Assume that the sample points are equally likely, that is, they have the same probability. Assume also that a conference call needs three free channels to be completed. Let us calculate the probability of a conference call be blocked due to busy channels. From the sample space we obtain that there are 16 favorable occurrences of three or more busy channels. Then, according to the classic definition, the probability of a conference call be blocked is given by $16/32 = 0.5$.

1.3 Counting Methods for Determining Probabilities

In many situations the sample space has a finite number of sample points and the probability of occurrence of a given event is determined according to the classic definition, i.e., by counting the number of favorable outcomes and dividing the result by the total number of possible outcomes in the sample space. In what follows some counting methods or combinatorial formulas are summarized. We consider *equally likely* outcomes in which the probability of a given sample point is $1/n$, where n is the total number of sample points in the sample space.

1.3.1 Distinct Ordered k -Tuples

The number of distinct ordered k -tuples with elements (x_1, x_2, \dots, x_k) obtained from sets having n_i distinct elements is given by

$$N_o = n_1 n_2 \cdots n_k. \quad (1.6)$$

The term “ordered” means that different orderings for the k -tuples form different and valid outcomes.

In other words, suppose that a number of choices are to be made and that there are n_1 possibilities for the first choice, n_2 for the second, \dots , and n_k for the k -th choice. If these choices can be combined without restriction, the total number of possibilities for the whole set of choices will be $n_1 n_2 \cdots n_k$. The following counting rules are based on (1.6).

1.3.2 Sampling with Replacement and with Ordering

If we choose k distinct objects from a set containing n distinct objects and repeat this procedure replacing the objects previously selected, the number of distinct ordered k -tuples is given by (1.6) with $n_1 = n_2 = \dots = n_k = n$, that is,

$$N_{\text{RO}} = n^k. \quad (1.7)$$

1.3.3 Sampling Without Replacement and with Ordering

If we choose k distinct objects from a set containing n distinct objects and repeat this procedure without replacing the objects previously selected, the number of distinct ordered k -tuples is given by

$$N_{\overline{\text{RO}}} = n(n-1) \cdots (n-k+1). \quad (1.8)$$

1.3.4 Permutation of Distinct Objects

If we select all $k = n$ distinct objects from a set containing n objects, the number of ways the n objects are selected, which is known as the number of *permutations*, is determined by (1.8) with $k = n$, yielding

$$N_{\text{P}} = n(n-1) \cdots (2)(1) = n!. \quad (1.9)$$

When n is large, the Stirling's formula can be used as an approximation for $n!$:

$$n! \cong \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}. \quad (1.10)$$

1.3.5 Sampling Without Replacement and Without Ordering

If we choose k distinct objects from a set containing n distinct objects and repeat this procedure without replacing the objects previously selected, the number of distinct k -tuples without considering the order, which is also known as the number of *combinations*, is given by

$$N_{\overline{R}O} = N_C = \frac{n!}{k!(n-k)!} = \binom{n}{k}, \quad (1.11)$$

where $\binom{n}{k}$ is the binomial coefficient¹.

1.3.6 Sampling with Replacement and Without Ordering

The number of different ways of choosing k distinct objects from a set with n distinct objects with replacement and without ordering is given by

$$N_{\overline{R}O} = \binom{n-1+k}{k} = \binom{n-1+k}{n-1}. \quad (1.12)$$

1.4 Conditional Probability and Bayes Rule

Let a combination of two experiments in which the probability of a *joint event* is given by $P[A, B]$. According to (1.6), the combined experiment produces a sample space that is formed by all distinct k -tuples from the n_i possible results from each experiment. Then, the joint probability $P[A, B]$ refers to the occurrence of one or more sample points out of $n_1 n_2$, depending on the event under interest. As an example, if two dice are thrown, the possible outcomes are formed by 36 distinct ordered pairs combining the numbers 1, 2, 3, 4, 5 and 6. Let A and B denote the events of throwing two dice and observing the number of points in each of them. The joint probability $P[A = 2, B \text{ odd}]$ corresponds to the occurrence of three sample points: (2, 1), (2, 3) and (2, 5). Then $P[A = 2, B \text{ odd}] = 3/36 = 1/12$.

Now suppose that the event B has occurred and, given this additional data, we are interested in knowing the probability of the event A . We call this probability a *conditional probability* and denote it as $P[A|B]$. It is shortly read as the probability of A , given B . The joint and the conditional probabilities are related through

$$P[A|B] = \frac{P[A, B]}{P[B]}. \quad (1.13)$$

Since $P[A, B] = P[B, A]$, using (1.13) we can write $P[B, A] = P[B|A]P[A] = P[A, B]$. With this result in (1.13) we have

$$P[A|B] = \frac{P[B|A]P[A]}{P[B]}. \quad (1.14)$$

¹ An experiment for computing the binomial coefficient using VisSim/Comm is given in the simulation file "CD drive:\Simulations\Probability\Bin.Coefficient.vsm". A block for computing the factorial operation is also provided in this experiment.

This important relation is a simplified form of the *Bayes rule*. Now, let the mutually exclusive events $A_i, i = 1, 2, \dots, n$. The general Bayes rule is given by

$$P[A_i|B] = \frac{P[A_i, B]}{P[B]} = \frac{P[B|A_i]P[A_i]}{\sum_{j=1}^n P[B|A_j]P[A_j]}, \quad (1.15)$$

where

$$P[B] = \sum_{j=1}^n P[B, A_j] = \sum_{j=1}^n P[B|A_j]P[A_j] \quad (1.16)$$

is the marginal probability of B , obtained from the joint probability $P[B, A_j]$. Equation (1.16) is referred to as the *theorem of total probability*.

In (1.14) we notice that if the knowledge of B does not modify the probability of A , we can say that A and B are *independent events*. In this case we have

$$P[A|B] = P[A] \text{ and } P[A, B] = P[A]P[B]. \quad (1.17)$$

Generally speaking, the joint probability of independent events is determined by the product of the probabilities of the isolated events. As an example, if we throw two fair coins, it is obvious that the result observed in one coin does not alter the result observed in the other one. This is a typical example of independent events. For example, the joint probability of heads in the two coins is $P[A = H, B = H] = P[H, H] = P[A = H]P[B = H] = P[H]P[H] = 0.5 \times 0.5 = 0.25$. In fact, the sample space is composed by four sample points: (H, H) , (H, T) , (T, H) and (T, T) and there is only one favorable event to heads in both tosses. Then, from the classical definition of probability, $P[H, H] = 1/4 = 0.25$.

Example 1.2 – In a digital communication system, the transmitter sends a bit “0” (event A_0) with probability $P[A_0] = p_0 = 0.6$ or a bit “1” (event A_1) with probability $P[A_1] = p_1 = (1 - p_0)$. The communication channel occasionally causes an error in a way that a transmitted “0” is converted into a received “1” and a transmitted “1” is converted into a received “0”. This channel is named *binary symmetric channel* (BSC) and it is represented by the diagram shown in Fig. 1.2.

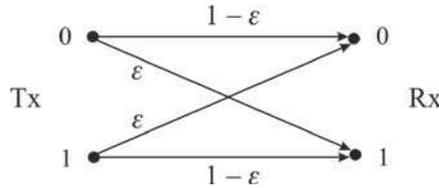


Fig. 1.2 Model for a binary symmetric channel (BSC)

Assume that the error probability is $\varepsilon = 0.1$, independent of the transmitted bit. Let the events B_0 and B_1 correspond to a received “0” and a received “1”, respectively. Let us compute the probabilities $P[B_0]$, $P[B_1]$, $P[B_1|A_0]$, $P[A_0|B_0]$, $P[A_0|B_1]$, $P[A_1|B_0]$ and $P[A_1|B_1]$:

1. From Fig. 1.2 we readily see that $P[B_0] = p_0(1 - \varepsilon) + p_1\varepsilon = 0.6(1 - 0.1) + 0.4 \times 0.1 = 0.58$. This result also comes from the total probability theorem given in (1.16): $P[B_0] = P[B_0|A_0]P[A_0] + P[B_0|A_1]P[A_1] = (1 - \varepsilon)p_0 + \varepsilon(1 - p_0)$.
 2. In a similar way, $P[B_1] = p_0\varepsilon + p_1(1 - \varepsilon) = 0.6 \times 0.1 + 0.4(1 - 0.1) = 0.42$.
 3. The probability $P[B_1|A_0] = P[B_0|A_1]$ is simply the probability of error. Then, $P[B_1|A_0] = \varepsilon = 0.1$.
 4. The probability $P[A_0|B_0]$ can be determined by applying Bayes rule: $P[A_0|B_0] = P[B_0|A_0]P[A_0]/P[B_0] = (1 - \varepsilon)p_0/0.58 = (1 - 0.1)0.6/0.58 \cong 0.931$.
 5. Similarly, the probability $P[A_0|B_1]$ can be determined by applying Bayes rule: $P[A_0|B_1] = P[B_1|A_0]P[A_0]/P[B_1] = \varepsilon p_0/0.42 = 0.1 \times 0.6/0.42 \cong 0.143$.
 6. The probability $P[A_1|B_0]$ can also be determined by applying Bayes rule: $P[A_1|B_0] = P[B_0|A_1]P[A_1]/P[B_0] = \varepsilon p_1/0.58 = 0.1(1 - 0.6)/0.58 \cong 0.069$.
 7. The last probability is given by $P[A_1|B_1] = P[B_1|A_1]P[A_1]/P[B_1] = (1 - \varepsilon)p_1/0.42 = (1 - 0.1)(1 - 0.6)/0.42 \cong 0.857$.
-

Simulation 1.1 – Conditional Probability



File – CD drive:\Simulations\Probability\Conditional.vsm

Default simulation settings: Frequency = 1 Hz; End = 2,000,000 seconds. Probability of a bit “0” in the random bits source: $p_0 = 0.6$. BSC error probability $\varepsilon = 0.1$.

This experiment complements Example 1.2, with emphasis on exploring the concepts of the conditional probability and the Bayes rule. It also aims at exploring the definition of probability by relative frequency.

A binary source generates random bits with prescribed probabilities of zeroes and ones. Specifically, the probability of a bit “0”, $p_0 = 1 - p_1$, can be configured. These random bits go through a binary symmetric channel (BSC) with configurable error probability ε . The input (A) and the output (B) of the BSC channel are analyzed and estimates of probabilities are made from them. These estimates use the concept of relative frequency, that is, the probability of occurrence of an event is computed by dividing the number of occurrences of the event by the total number of observations.

As an exercise, have a look inside the probability estimation blocks and try to understand how they were implemented. Pay special attention to the conditional probability estimation blocks, where the total number of observations reflects the

condition under which the probability is being estimated. In other words, note that the probability $P[B_1|A_0]$ is estimated by dividing the number of occurrences of a bit “1” at the BSC output B given that the input A was “0” by the total number of times that the input A was “0”.

Run the experiment and confirm the results obtained via Example 1.2. Change the configuration of the probabilities of ones and zeroes and the channel error probability. Repeat the computations of the probabilities considered in Example 1.2 and check your results by means of the simulation. An interesting configuration to be analyzed corresponds to equiprobable bits, i.e. $p_0 = p_1 = 1/2$.

As another exercise, reduce the simulation time and observe that the estimates deviate from the theoretical ones. The lower the simulation time, the more inaccurate the probability estimates are. Recall the concept of probability by relative frequency and justify this behavior.

1.5 Random Variables

Random variables are numbers that represent the results of an experiment. As an example, let the event A denote the number of heads in three tosses of a fair coin. We can create a random variable $X(A)$, or simply X , to represent the possible results of the experiment. In this example, X can assume the values 0, 1, 2 or 3.

The mapping of an event into a number is such that the probability of occurrence of an event is equal to the probability of occurrence of the corresponding value of the random variable. For example, the probability of two heads in the experiment above corresponds to the probability of occurrence of the number 2 for the random variable X .

1.5.1 The Cumulative Distribution Function

By definition, the cumulative distribution function (CDF) of a random variable X is given by

$$F_X(x) \triangleq P[X \leq x], \quad -\infty \leq x \leq \infty, \quad (1.18)$$

which means that $F_X(x)$ gives the probability that the random variable X assume a value in the interval $(-\infty, x]$. Figure 1.3 depicts the exemplifying CDF

$$F_X(x) = \begin{cases} 0, & x < -1 \\ \frac{x+1}{10}, & -1 \leq x < 2 \\ \frac{x}{5}, & 2 \leq x < 4 \\ 1, & x \geq 4 \end{cases}. \quad (1.19)$$

The main properties of the CDF are:

1. $0 \leq F_X(x) \leq 1$
 2. $\lim_{x \rightarrow \infty} F_X(x) = 1$
 3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$
 4. If $a \leq b \Rightarrow F_X(a) \leq F_X(b)$
 5. $P[a \leq X \leq b] = F_X(b) - F_X(a^-)$.
- (1.20)

In the last property given in (1.20), $F_X(a^-)$ is the value of $F_X(x)$ for $x \rightarrow a$ from the left. This property leads to the following additional property: if the CDF is continuous in $x = b$, then the probability of $X = b$ is zero, as shown by:

$$P[X = b] = F_X(b) - F_X(b^-) = 0. \quad (1.21)$$

Moreover, if $F_X(x)$ is continuous in $x = a$ and $x = b$, then

$$P[a < X < b] = P[a \leq X < b] = P[a < X \leq b] = P[a \leq X \leq b]. \quad (1.22)$$

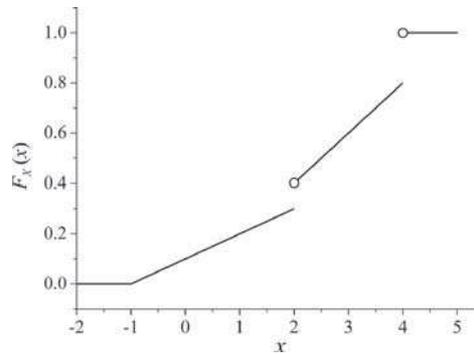


Fig. 1.3 Illustration of a cumulative distribution function

Example 1.3 – By using the CDF shown in Fig. 1.3, let us determine the probabilities $P[X < 2]$, $P[X = 2]$ and $P[X \geq 3]$.

1. $P[X < 2] = F_X(2^-) = 0.3$.
 2. $P[X = 2] = F_X(2) - F_X(2^-) = 0.4 - 0.3 = 0.1$.
 3. $P[X \geq 3] = 1 - P[X < 3] = 1 - F_X(3^-) = 1 - 0.6 = 0.4$.
-

1.5.2 Types of Random Variables

In this section we address the types of random variables and the corresponding mathematical functions that characterize them.

1.5.2.1 Discrete Random Variables

A discrete random variable is the one that assumes discrete countable values. These values can be in a finite or in an infinite range.

Discrete random variables are usually represented in terms of its *probability mass function* (PMF), defined by

$$p_X(x) \triangleq P[X = x], \quad x \text{ real.} \quad (1.23)$$

The cumulative distribution function of a discrete random variable is the sum of unit-step functions $u(x)$ located at each value of X , and weighted by each of the corresponding PMF values:

$$F_X(x) = \sum_k p_X(x_k) u(x - x_k). \quad (1.24)$$

The magnitude of the “jumps” in the CDF at $x = x_k$ is the value of $P[X = x_k]$.

1.5.2.2 Continuous Random Variables

A continuous random variable has a continuous cumulative distribution function for all of its values. In other words, a continuous random variable is the one that assumes continuous values in a finite or in an infinite range. According to (1.21), the probability of occurrence of a particular value for a continuous random variable is zero, that is, $P[X = x] = 0$, for any value of x .

Continuous random variables are usually characterized by a *probability density function* (PDF), defined as the first derivative of the CDF, as shown by:

$$f_X(x) \triangleq \frac{dF_X(x)}{dx}. \quad (1.25)$$

The PDF is called a *probability density* in the sense that it represents the probability of X be in a small interval in the neighborhood of x . More details about the PDF will be given later.

1.5.2.3 Mixed Random Variables

As the name suggests, mixed random variables are those represented by cumulative distribution functions having continuous and discontinuous portions. An example of such CDF was given in Fig. 1.3.

For mixed random variables the CDF can be written in terms of discrete and continuous parts, as shown below for one discrete component and one continuous component, that is,

$$F_X(x) = pF_{X_1}(x) + (1 - p)F_{X_2}(x), \quad (1.26)$$

where p is the probability of occurrence of the discrete event, $F_{X_1}(x)$ is the CDF of the discrete component random variable and $F_{X_2}(x)$ is the CDF of the continuous component random variable.

1.5.2.4 Probability Density Function of a Discrete Random Variable

The cumulative distribution function of a discrete random variable can be written in terms of the Dirac delta function $\delta(z)$ as [9, p. 151]

$$F_X(x) = \int_{-\infty}^x \sum_k p_X(x_k) \delta(z - x_k) dz, \quad (1.27)$$

which suggests that the probability density function of a discrete random variable is given by

$$f_X(x) = \frac{dF_X(x)}{dx} = \sum_k p_X(x_k) \delta(x - x_k). \quad (1.28)$$

We can notice the slight difference between this probability density function and the probability mass function defined in (1.23). Graphically, the representation of the PDF can be used for representing a PMF. However, the PDF is mathematically more adequate from the point of view of its relation with the CDF.

1.5.2.5 Properties of the Probability Density Function

The probability density function has the following main properties:

1. $f_X(x) \geq 0$
 2. $P[a \leq X \leq b] = \int_a^b f_X(x) dx$
 3. $F_X(x) = \int_{-\infty}^x f_X(t) dt$
 4. $\int_{-\infty}^{+\infty} f_X(t) dt = 1.$
- (1.29)

The second, third and fourth properties in (1.29), when applied to a discrete random variable, will lead to equivalent results produced by the following operations involving the probability mass function:

$$\begin{aligned}
1. \quad P[B] &= \sum_{x \in B} p_X(x) \\
2. \quad F_X(x) &= \sum_k p_X(x_k) u(x - x_k) \\
3. \quad \sum_{x \in S} p_X(x) &= 1.
\end{aligned} \tag{1.30}$$

1.5.2.6 The Normalized Histogram as an Approximation for the PDF

From [4, p. 348], a histogram is “a graph of vertical bars representing a frequency distribution in which the groups or classes of items are marked on the x -axis, and the number of items in each class is indicated by a horizontal line segment drawn above the x -axis at a height equal to the number of items in the class”.

Example 1.4 – This example aims at clarifying the above definition. Suppose you have collected the height of 100 graduate students, obtaining the following values in meters:

1.62 1.74 1.54 1.93 1.71 1.70 1.71 1.61 1.53 1.63
1.65 1.74 1.64 1.76 1.69 1.63 1.76 1.66 1.60 1.77
1.70 1.71 1.77 1.63 1.73 1.70 1.61 1.58 1.60 1.72
1.62 1.72 1.67 1.73 1.64 1.59 1.77 1.68 1.74 1.78
1.51 1.78 1.71 1.49 1.73 1.77 1.68 1.76 1.75 1.81
1.77 1.74 1.68 1.67 1.63 1.73 1.80 1.82 1.82 1.87
1.56 1.80 1.75 1.61 1.74 1.73 1.51 1.69 1.77 1.73
1.86 1.81 1.80 1.67 1.73 1.78 1.92 1.81 1.77 1.64
1.55 1.80 1.64 1.50 1.63 1.55 1.60 1.68 1.96 1.73
1.73 1.61 1.88 1.72 1.62 1.82 1.67 1.54 1.65 1.56

Following the definition above, you will obtain the histogram shown in Fig. 1.4. In this histogram, the range of heights was divided into 11 *classes* or *bins*. In the vertical axis are shown the number of students having their height in the corresponding class in the horizontal axis. For example, there are 6 students with heights between 1.50 and 1.55 m.

The information that there are 6 students with heights between 1.50 and 1.55 m can not be converted into an estimate of probability, unless, in the light of the relative frequency definition for probability, we have an infinite number of samples. However, if this number is high enough, a good probability estimate can be obtained. For example, if we had 10,000 samples in Example 1.4, the division of the number of students having heights between 1.50 and 1.55 m by the total area of the histogram would result in a good approximation of the probability of the student height be in this range.

The total area of the histogram can be easily calculated through

$$A = \sum_{c=1}^C n_c \Delta, \quad (1.31)$$

where C is the number of classes, n_c is the number of occurrences of the random variable in class c and Δ is the class width.

If we decrease the width of the classes in a histogram to a very small value and perform the normalization of the frequency axis in order to show the estimate of the probability corresponding to each class, the resultant graph will have the top of the thin bars very close to the probability density function of the random variable under analysis. Figure 1.5 illustrates the *normalized histogram* considering 10,000 samples and a class width of 0.005 m. Note that the histogram has a shape very similar to the real PDF of the corresponding random variable.

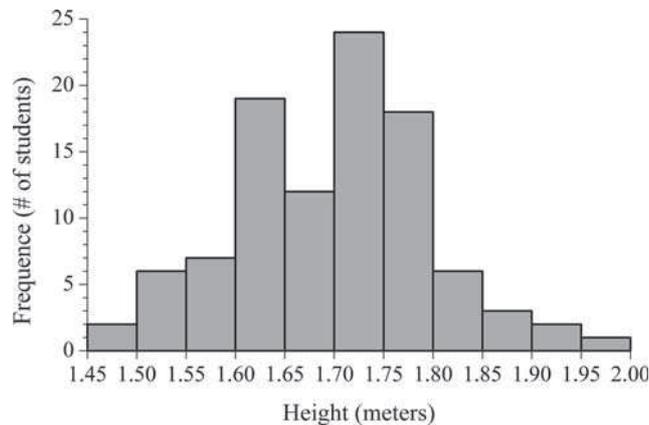


Fig. 1.4 Histogram for the Example 1.4, considering 100 samples

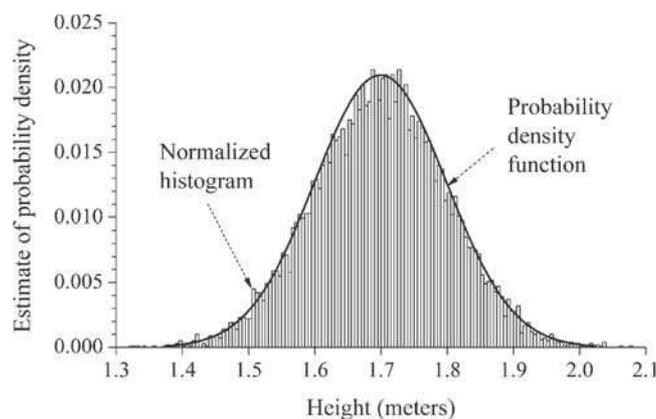


Fig. 1.5 Normalized histogram for the Example 1.4, considering 10,000 samples

In summary, the histogram is a very useful tool for estimating the shape of the probability density function of a random variable. The greater the sample size, the better the approximation between the normalized histogram and the PDF will be. The smaller the class width, the greater the histogram ability of recording short-term undulations in the estimated PDF will be. This is particularly useful in the case of multimodal PDFs, i.e. PDFs that exhibit two or more “peaks”.

1.6 Conditional Probability Densities and Mass Functions

In (1.13) we presented the definition of the conditional probability and, from this definition, in (1.14) and (1.16) we derived the Bayes rule and the total probability theorem, respectively. Now we are able to extend those concepts to conditional densities and mass functions. The concepts we are about to present in this section are of great relevance, since in practice we commonly encounter random variables that are not independent from each other.

We start by defining the conditional probability mass function for discrete random variables X and Y . The conditional PMF of X , given $Y = y$ is given by

$$p_X(x|y) = P[X = x|Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{p_{X,Y}(x, y)}{p_Y(y)}. \quad (1.32)$$

If X and Y are continuous random variables we have the following conditional PDF of X , given $Y = y$:

$$f_X(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_Y(y|x)f_X(x)}{f_Y(y)}. \quad (1.33)$$

1.6.1 The Bayes Rule Revisited

Suppose that we are interested in knowing the *a posteriori* probability of an event B , given $X = x$ has occurred, $P[B|X = x]$, from the knowledge of the conditional probability density function of x given B , $f_X(x|B)$, from the *a priori* probability of B , $P[B]$, and from the total probability density $f_X(x)$. The Bayes rule for probability density functions states that [15, p. 84]

$$P[B|X = x] = \frac{f_X(x|B)P[B]}{f_X(x)}, \quad f_X(x) \neq 0. \quad (1.34)$$

This is an important result, largely used in parameter estimation based on the *maximum a posteriori* (MAP) probability criterion. In Chaps. 4 and 5 we apply this version of Bayes rule for developing two important decision rules used in digital communication systems: the MAP itself and the *maximum likelihood* (ML).

1.6.2 The Total Probability Theorem Revisited

Multiplying both sides of (1.34) by $f_X(x)$ and integrating the result we obtain

$$P[B] = \int_{-\infty}^{\infty} P[B|X = x]f_X(x)dx. \quad (1.35)$$

This is also an important result and represents the *continuous version of the total probability theorem* given in (1.16). It states that if we know the probability of occurrence of an event, given the value of another random variable, the “average” probability of the event is obtained by averaging the conditional probability over all possible values of the random variable to which the probability is conditioned.

Although the definition of the total probability theorem given in (1.16) applies directly to discrete random variables, it is worth rewriting it in terms of the PMF of the condition discrete random variable, as follows:

$$P[B] = \sum_k P[B|X = x_k]p_X(x_k). \quad (1.36)$$

1.7 Statistical Averages of Random Variables

As pointed out at the beginning of this chapter, though random phenomena can not be precisely predicted, they exhibit regularities that permit some sort of inference and a given confidence on such inference. These regularities are given by statistical averages in a more formal and more generic sense than the intuitive concept about averages that we might have in mind.

An statistical average of a random variable or random process is usually denoted by the *expectation* or *expected value* operator $E[\cdot]$.

1.7.1 Mean of Random Variables

One of the most important statistical averages is the *mean*. For a discrete random variable X , the mean is given by

$$E[X] = \mu_X = \sum_k x_k p_X(x_k), \quad (1.37)$$

where $\{x_k\}$ are the possible values assumed by the random variable and $p_X(x_k)$ is the value of the probability mass function for $X = x_k$. The shorthand notation μ_X is sometimes preferred for the sake of notational simplicity.

For a continuous random variable X , the mean is given by

$$E[X] = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx, \quad (1.38)$$

where $f_X(x)$ is the probability density function of X .

If a random variable is defined in terms of another random variable, say, $Y = g(X)$, then the mean value of Y is given by

$$E[Y] = \sum_k g(x_k) p_X(x_k), \quad (1.39)$$

if the random variables are discrete, and

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{-\infty}^{\infty} g(x) f_X(x) dx, \quad (1.40)$$

if the random variables are continuous.

The concept of mean can be extended to the case in which a random variable Y is written in terms of multiple random variables, that is, $Y = g(X_1, X_2, \dots, X_n)$. For example, for multiple continuous random variables we have:

$$E[Y] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \quad (1.41)$$

The mean of the sum of random variables is the sum of the individual means:

$$E \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n E[X_i] \quad (1.42)$$

In Sect. 1.4 we discussed the concept of a joint event and the concept of independent events. We have seen that the joint probability of occurrence of independent events is the product of the probabilities of the individual events. This concept is analogous to the case of probability functions of random variables, that is, two discrete random variables are said to be independent random variables if the *joint probability mass function* is the product of the individual (marginal) probability mass functions of the discrete random variables:

$$p_{X,Y}(x_j, y_k) = p_X(x_j) p_Y(y_k), \quad (1.43)$$

for all j and k . Equivalently, two continuous random variables are said to be independent if the *joint probability density function* is the product of the marginal probability density functions of the continuous random variables:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y). \quad (1.44)$$

The mean of the product between independent random variables is the product of the individual means, that is,

$$E[XY] = E[X]E[Y]. \quad (1.45)$$

1.7.2 Moments of Random Variables

The n -th moment of a random variable X is defined as the expected value of the n -th power of X , as shown below for a continuous random variable:

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x) dx, \quad (1.46)$$

from where we can notice that the first moment of a random variable corresponds to its mean.

The second moment $E[X^2]$ of a random variable X is of great importance and is usually referred to as the *mean-square* value of the random variable. We shall see at the end of this chapter that the second moment of a random process corresponding to a voltage waveform is the total average power of the waveform.

We also have the definition of the n -th central moment as the expected value of the centralized random variable, where centralized indicates that the mean is taken out of the computation of the moment, as shown by:

$$E[(X - \mu_X)^n] = \int_{-\infty}^{\infty} (x - \mu_X)^n f_X(x) dx. \quad (1.47)$$

Similarly, for discrete random variables we have

$$E[(X - \mu_X)^n] = \sum_k (x_k - \mu_X)^n p_X(x_k). \quad (1.48)$$

Also of great importance is the second central moment, usually referred to as the *variance* of the random variable. We shall see at the end of this chapter that the second central moment of a random process corresponding to a voltage waveform is the total AC power² of the waveform. The shorthand notations for the variance of a random variable X are usually $\text{var}[X]$ or σ_X^2 , where σ_X is called *standard deviation* or *root-mean-square (rms)* value of X .

If a is a constant, then we have the following main properties of the variance:

² AC stands for *alternating current*. DC stands for *direct current*.

1. If $X = a$, then $\text{var}[X] = 0$
2. If $Y = X + a$, then $\text{var}[Y] = \text{var}[X]$
3. If $Y = aX$, then $\text{var}[Y] = a^2 \text{var}[X]$.

(1.49)

From (1.38) and (1.47) we can write the important relation:

$$\sigma_X^2 = E[X^2] - \mu_X^2. \quad (1.50)$$

Again, we shall see at the end of this chapter that for a random process corresponding to a voltage waveform, (1.50) states that the average AC power σ_X^2 is equal to the total average power $E[X^2]$ minus the average DC power $E^2[X]$, which is an intuitively satisfying interpretation.

The relation among the central moment, the non-centralized moment and the first moment of a random variable can be generalized as follows [15, p. 193]:

$$E[(X - \mu_X)^n] = \sum_{k=0}^n \binom{n}{k} (-1)^k \mu_X^k E[X^{n-k}], \quad (1.51)$$

where the binomial coefficient was defined in (1.11).

1.7.3 Joint Moments of Random Variables

The *joint moment* of two continuous random variables X and Y is defined by

$$E[X^i Y^n] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^i y^n f_{X,Y}(x, y) dx dy. \quad (1.52)$$

For discrete random variables we have

$$E[X^i Y^n] = \sum_j \sum_k x_j^i x_k^n p_{X,Y}(x_j, y_k), \quad (1.53)$$

where the *order* of the joint moment is given by $i + n$.

Similarly, the *joint central moment* of two random variables X and Y is $E[(X - \mu_X)^i (Y - \mu_Y)^n]$. For continuous random variables it is given by

$$E[(X - \mu_X)^i (Y - \mu_Y)^n] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^i (y - \mu_Y)^n f_{X,Y}(x, y) dx dy. \quad (1.54)$$

1.7.3.1 Correlation Between Random Variables

The second-order joint moment between two random variables is called *correlation* and it is obtained with $i = 1$ and $n = 1$ in (1.52), resulting in the second-order moment $E[XY]$.

Two random variables X and Y are said *orthogonal* if their correlation is zero.

If two orthogonal random variables are also *independent*, from (1.45) $E[XY] = E[X]E[Y] = 0$, which means that $E[X]$, $E[Y]$, or both are zero.

1.7.3.2 Covariance Between Random Variables

The second-order joint central moment between two random variables is called *covariance* and it is obtained with $i = 1$ and $n = 1$ in (1.54), yielding

$$\begin{aligned} \text{cov}[X, Y] &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - E[X]E[Y]. \end{aligned} \quad (1.55)$$

Two random variables X and Y are said to be *uncorrelated* if their covariance is zero. The degree of correlation is usually measured by a normalized covariance ranging from -1 to 1 and called *correlation coefficient*, which is defined by

$$\rho \triangleq \frac{\text{cov}[X, Y]}{\sqrt{\sigma_X^2 \sigma_Y^2}}. \quad (1.56)$$

Note that, with the condition (1.45) in (1.55), if two random variables are statistically independent, they are also uncorrelated. However, although two uncorrelated random variables satisfy $E[XY] = E[X]E[Y]$, they are also independent if and only if their joint probability density or mass function is the product of the marginal probability density or mass functions, according to (1.43) and (1.44).

Note also that if two random variables are orthogonal, then $E[XY] = 0$. With this condition in (1.55), these variables will be uncorrelated if and only if one of their expected values, or both, are zero.

1.7.4 The Characteristic and the Moment-Generating Functions

The characteristic functions (CF) of continuous and discrete random variables are defined by the statistical averages:

$$\psi_X(j\omega) \triangleq E[e^{j\omega X}] = \begin{cases} \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx & \text{if } X \text{ is continuous.} \\ \sum_k e^{j\omega x_k} p_X(x_k) & \text{if } X \text{ is discrete.} \end{cases} \quad (1.57)$$

Note that the characteristic function is similar to the Fourier transform (see Chap. 2), except by the sign of the exponential. Then, the probability density

function can be obtained from the characteristic function by means of the “inverse Fourier transformation”, leading to

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega x} \psi_X(j\omega) d\omega. \quad (1.58)$$

One of the major applications of the characteristic function is in the determination of the moments of a random variable. This is accomplished through the *moment theorem*, which is given by

$$E[X^n] = (-j)^n \left. \frac{d^n \psi_X(j\omega)}{d\omega^n} \right|_{\omega=0}. \quad (1.59)$$

Another application of the characteristic function is the estimation of the probability density function of a random variable from experimental measurements of the moments [15].

The *probability generating function* [9, p. 187] or *moment-generating function* (MGF) [10, p. 114; 15, p. 211; 18, p. 248] of a random variable is defined analogously to the characteristic function and has essentially the same applications. However, the characteristic function always exists, which is not the case of the moment-generating function [1, pp. 194, 350].

The moment-generating functions of continuous and discrete random variables are defined by

$$\phi_X(s) \triangleq E[e^{sX}] = \begin{cases} \int_{-\infty}^{\infty} e^{sx} f_X(x) dx & \text{if } X \text{ is continuous.} \\ \sum_k e^{sx_k} p_X(x_k) & \text{if } X \text{ is discrete.} \end{cases} \quad (1.60)$$

The MGF of a continuous random variable is similar to the Laplace transformation (see Chap. 2), the difference being that the MGF is defined for all s real. Then, the MGF and PMF or PDF form a transform pair.

The moment theorem also applies in the case of the MGF and is given by

$$E[X^n] = \left. \frac{d^n \phi_X(s)}{ds^n} \right|_{s=0}. \quad (1.61)$$

Other important properties of the moment-generating function are:

1. $\phi_X(s)|_{s=0} = 1$
 2. If $Y = aX + b$, then $\phi_Y(s) = e^{sb} \phi_X(as)$.
- (1.62)

In Sect. 1.10 we shall discuss other properties of the characteristic function that are also applicable to the moment-generating function, in the context of the sum of random variables. The summarized concepts considered up to this point suffice for the continuity of the presentation.

1.7.5 Conditional Expected Value

The *conditional expected value* $E[X|y]$ is the expected value of the random variable X , given a value of another random variable Y be equal to y . Mathematically, if X and Y are both continuous random variables,

$$E[X|y] = E[X|Y = y] = \int_{-\infty}^{\infty} x f_X(x|y) dx. \quad (1.63)$$

If X and Y are both discrete random variables, the conditional expectation is

$$E[X|y_j] = E[X|Y = y_j] = \sum_k x_k p_X(x_k|y_j). \quad (1.64)$$

From the *total expectation law*, $E[E[X|Y]] = E[X]$. As a consequence, we have the following unconditional expectations:

$$E[X] = \begin{cases} \int_{-\infty}^{\infty} E[X|y] f_Y(y) dy, & \text{if } X \text{ and } Y \text{ are continuous.} \\ \sum_j E[X|y_j] p_Y(y_j), & \text{if } X \text{ and } Y \text{ are discrete.} \end{cases} \quad (1.65)$$

1.8 Some Discrete Random Variables

In what follows we present some discrete random variables, examples of their use, their probability mass functions, main statistical averages and moment-generating functions. We omit further details for the sake of brevity. For more information, the interested reader is encouraged to consult the references listed at the end of the chapter. A great number of details about several distributions can also be found online at the *Wikipedia* encyclopedia [16] or at the *Wolfram MathWorld* [17]. Some of the simulations considered in this chapter will also help clarifying and exemplifying the concepts about these random variables.

1.8.1 Bernoulli Random Variable

This discrete random variable is largely used to model random phenomena that can be characterized by two states. For example: on/off, head/tail, bit “0” / bit “1”. These states are commonly associated to the terms *success* or *failure*. The attribution of these terms to a given state is somewhat arbitrary and depends on what we want for the meaning of the states in relation to the actual observed phenomenon.

A Bernoulli random variable can also be used to model any random phenomena for which an indicator function I_A is used to signalize the occurrence of an event A . The indicator function will be a Bernoulli random variable that assumes the state “1” (success) if the event under observation occurs and “0” (failure) otherwise. If

p is the probability of success and $q = (1 - p)$ is the probability of failure in this Bernoulli trial, then $p = P[A]$.

The probability mass function, mean, variance and moment-generating function of a Bernoulli random variable X with parameter p are, respectively:

$$\begin{aligned}
 1. \quad p_X(x) &= \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{otherwise} \end{cases} \\
 2. \quad E[X] &= p \\
 3. \quad \text{var}[X] &= p(1 - p) \\
 4. \quad \phi_X(s) &= 1 - p + pe^s.
 \end{aligned} \tag{1.66}$$

1.8.2 Binomial Random Variable

A binomial random variable X counts the number of successes in n Bernoulli trials or, equivalently, X is the sum of n Bernoulli random variables. Its applications arise, for example, in the analysis of the number of ones (or zeros) in a block of n bits at the output of a binary source. If we assign the probability of success p of a Bernoulli trial to the probability of bit errors in a digital communication system, the number of bits in error in a block of n bits is a binomial random variable.

The probability mass function, mean, variance and moment-generating function of a binomial random variable X with parameters n and p are, respectively:

$$\begin{aligned}
 1. \quad p_X(x) &= \binom{n}{x} p^x (1 - p)^{n-x} \\
 2. \quad E[X] &= np \\
 3. \quad \text{var}[X] &= np(1 - p) \\
 4. \quad \phi_X(s) &= (1 - p + pe^s)^n.
 \end{aligned} \tag{1.67}$$

1.8.3 Geometric Random Variable

The geometric random variable characterizes the number of failures before a success or between successes in a sequence of Bernoulli trials. Alternatively, a geometric random variable can characterize the number of Bernoulli trials needed to produce one success, no matter if a success has or has not occurred before. In what follows we consider the first case, for which the results are $0, 1, 2, \dots$, etc.

As an example, if we assign the probability of success p of a Bernoulli trial to the probability of bit errors in a digital communication system, the number of correct bits before a bit error occurs is a geometric random variable.

The probability mass function, mean, variance and moment-generating function of a geometric random variable X with parameter p are, respectively:

$$\begin{aligned}
 1. \quad p_X(x) &= \begin{cases} p(1-p)^x, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases} \\
 2. \quad E[X] &= \frac{1-p}{p} \\
 3. \quad \text{var}[X] &= \frac{1-p}{p^2} \\
 4. \quad \phi_X(s) &= \frac{p}{1-(1-p)e^s}.
 \end{aligned} \tag{1.68}$$

The geometric random variable is the only discrete *memoryless* random variable. This means that if you are about to repeat a Bernoulli trial, given that the first success has not yet occurred, the number of additional trials until a success shows up does not depend on how many failures have been observed till that moment. As an example, the die or the coin you throw in a given instant is not influenced by the failures or successes up to that instant.

1.8.4 Poisson Random Variable

A Poisson random variable is used to characterize the number of occurrences of a random, normally rare event in a given observation interval or space. Here the term “rare” is determined according to the observation interval or space. In other words, the event will occur a few times, on average, in the observed time or space.

If we define α as the average number of occurrences of the event in a specific observation interval or space, $\lambda = 1/\alpha$ can be regarded as the average rate of occurrence of the Poisson event.

The probability mass function, mean, variance and moment-generating function of a Poisson random variable X with parameter α are, respectively:

$$\begin{aligned}
 1. \quad p_X(x) &= \begin{cases} \frac{\alpha^x e^{-\alpha}}{x!}, & \alpha > 0, \quad x = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases} \\
 2. \quad E[X] &= \alpha \\
 3. \quad \text{var}[X] &= \alpha \\
 4. \quad \phi_X(s) &= e^{\alpha(e^s - 1)}.
 \end{aligned} \tag{1.69}$$

A binomial distribution with large n and small p can be approximated by a Poisson distribution with parameter $\alpha = np$, according to

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \cong \frac{np^x}{x!} e^{-np}. \tag{1.70}$$

Simulation 1.2 – Discrete Random Variables



File – CD drive:\Simulations\Probability\Discrete.RV.vsm

Default simulation settings: Frequency = 1 Hz; End = 10,000 seconds. Probability of success for the configurable Bernoulli sources: $p = 0.5$.

This experiment aims at helping the reader to understand some concepts related to discrete random variables. All discrete random variables considered in this section are generated: Bernoulli, binomial, Poisson and geometric. The variance and mean are estimated for each random variable. VisSim/Comm has a specific block to do these estimates, but in fact these blocks use *sample averages*, a topic covered later on in this chapter.

In what follows we shall analyze each of the above random variables. The Bernoulli source generates a random sequence of zeros and ones during the simulation interval. The probability of success of the Bernoulli trial is represented here by the probability of generating a binary “1”. Using the default simulation settings, run the experiment and observe the values of the mean and variance of the Bernoulli random variable. According to (1.66), these values should be $E[X] = p = 0.5$ and $\text{var}[X] = p(1 - p) = 0.25$. Look inside the block “histograms” and rerun the simulation. Observe that, for $p = 0.5$, in fact the frequencies of occurrence of ones and zeroes are roughly the same. Change the value of p and compare again the variance and mean of the Bernoulli random variable with the theoretical results. Analyze the histogram and observe the dependence of its shape on the value of p .

Still considering the first part of the experiment, now let us shift our attention to the binomial random variable. Observe that, according to the theoretical definition, it has been generated as the sum of $n = 10$ Bernoulli random variables. The simulation does this by summing-up 10 consecutive Bernoulli results. The updating of the sum is made by entering a new Bernoulli value to be added and discarding the tenth one. Run the simulation and observe the mean and variance of the binomial random variable. According to (1.67), $E[X] = np = 5$ and $\text{var}[X] = np(1 - p) = 2.5$. Observe the histogram and, with the help of (1.67), determine the PMF and compare its shape with the shape of the histogram. Change the value of p as you wish and repeat the calculations.

The second part of the experiment generates a geometric random variable and repeats in a different way the generation of a binomial random variable: now, the number of successes in each 10 Bernoulli trials is being counted, generating a binomial random variable.

According to the theory, the number of failures before each success of a Bernoulli trial has a geometric distribution. The number of failures before the first success is also a geometric random variable. Note that the number of failures before each success is updated several times during the simulation and that the number of failures before the first success is updated only in a simulation run basis.

Using the default settings, run the simulation and observe the mean and the variance of the geometric random variable. Compare the results with those obtained via

(1.68). Observe also the shape of the histogram, which corresponds to the shape of the probability mass function of the geometric distribution.

The third and last part of the experiment generates a Poisson random variable. Since a Poisson event occurs rarely, the corresponding random variable was generated by a binomial source with $p = 0.002$ and $n = 1,000$. According to the theory, a binomial random variable with such a small p and large n approximates a Poisson random variable. Using the default simulation settings, run the simulation and observe the “event display”. Note that, in fact, a Poisson event is relatively rare. You should observe something varying around $10np = 20$ Poisson events at the output of the “accumulated successes” block, where the multiplier 10 is the number of observation intervals of 1,000 seconds during the overall simulation time of 10,000 seconds.

Since the Poisson event is relatively rare, to have a better estimate of the random variable statistics, we must increase the simulation time to, say, 1,000,000 seconds. Run the simulation and observe the “accumulated successes”. Divide this number by the observation time of 1,000 seconds. You will find a value very close to 2, the actual mean of the Poisson random variable. Compare this value of mean and the variance with the expected theoretical results given by (1.69), where $\alpha = np = 1,000 \times 0.002 = 2$ is the expected number of events in 1,000 seconds, which by its turn is the mean and the variance of the Poisson random variable. Observe also the shape of the histogram, which corresponds to the shape of the probability mass function of the Poisson distribution.

Explore inside the individual blocks. Try to understand how they were implemented. Create and investigate for yourself new situations and configurations of the simulation parameters and try to reach your own conclusions.

1.9 Some Continuous Random Variables

In this section we present some continuous random variables, examples of their use, their probability density functions, main statistical averages and moment-generating functions. Since this is a review chapter, we omit further information for the sake of brevity. Details about several continuous distributions can also be found online at the *Wikipedia* encyclopedia homepage [16] or in the *Wolfram MathWorld* homepage [17]. Some of the simulations addressed throughout this chapter will also help clarifying and exemplifying the concepts about the continuous random variables considered here.

1.9.1 Uniform Random Variable

As the name suggests, a uniform random variable has its probability density function uniformly distributed in a given range. The probability density function, mean, variance and moment-generating function of a uniform random variable X with parameters a and b are, respectively:

$$\begin{aligned}
1. f_X(x) &= \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} \\
2. E[X] &= \frac{a+b}{2} \\
3. \text{var}[X] &= \frac{(b-a)^2}{12} \\
4. \phi_X(s) &= \frac{e^{bs} - e^{as}}{s(b-a)}.
\end{aligned} \tag{1.71}$$

Many physical phenomena are characterized by a uniform distribution. For example, the uniform quantization used as part of the digitalizing process of a signal generates an error that is uniformly distributed around zero. Moreover, when dealing with random number generation, a uniform random variable is the basis for the generation of random variables having other distributions. This last application will become clearer later on in this chapter, where the generation of random numbers is covered.

1.9.2 Exponential Random Variable

The probability density function, mean, variance and moment-generating function of an exponential random variable X with parameter λ are, respectively:

$$\begin{aligned}
1. f_X(x) &= \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \\
2. E[X] &= \frac{1}{\lambda} \\
3. \text{var}[X] &= \frac{1}{\lambda^2} \\
4. \phi_X(s) &= \frac{\lambda}{\lambda - s}.
\end{aligned} \tag{1.72}$$

The parameter λ is usually referred to as the *rate* parameter. The exponential distribution can be seen as the continuous counterpart of the geometric distribution and, as such, is the only continuous *memoryless* random variable. In terms of the conditional probability, this means that

$$P[X > x_1 + x_2 | X > x_1] = P[X > x_2], \quad x_1 \text{ and } x_2 \geq 0. \tag{1.73}$$

The exponential random variable has many applications and, among them we can mention: the time until the occurrence of the first event and between the events in

a Poisson random variable is exponentially-distributed. In this case, if the Poisson has α as the average number of occurrences of the events in a specific observation interval T_0 , $\lambda = \alpha/T_0$ is the average rate of occurrence of the Poisson events. The exponential distribution can also be used to characterize the time between beeps in a Geiger counter, the time between telephone calls or how long it takes for a teller to serve a client.

1.9.3 Gaussian Random Variable

The probability density function, mean, variance and moment-generating function of a Gaussian random variable X with mean μ and standard deviation σ are, respectively:

$$\begin{aligned} 1. f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \\ 2. E[X] &= \mu \\ 3. \text{var}[X] &= \sigma^2 \\ 4. \phi_X(s) &= e^{s\mu+s^2\sigma^2/2}. \end{aligned} \tag{1.74}$$

The Gaussian random variable has a major importance for the study of several areas, particularly for the study of digital transmission. It is used, for example, to model *thermal noise* generated by the random motion of electrons in a conductor. This noise is the omnipresent impairment whose influence must be considered in the design of any communication system. Moreover, a Gaussian distribution is used to characterize the sum of independent random variables having any distribution. The *central limit theorem*, a subject treated later on in this chapter, states that under a variety of conditions the PDF of the sum of n independent random variables converges to a Gaussian PDF for a sufficient large n .

Due to its importance for the study of communication systems, in what follows we shall devote a little bit more attention to the Gaussian random variable.

One of the problems that often arise in practice is related to the computation of the area under a Gaussian PDF. This computation is usually related to the estimate of probabilities of rare events, which are normally associated to the area of the tails of a Gaussian distribution. To elaborate more on this matter, consider the problem depicted in Fig. 1.6. Suppose that we want to calculate the probability of the underlying random variable be greater than a threshold γ . According to the second line in (1.29), this probability can be computed by

$$P[X > \gamma] = \int_{\gamma}^{\infty} f_X(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\gamma}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx. \tag{1.75}$$

Suppose that we also want to calculate the probability of the random variable X be in the interval $(\mu - 3\sigma, \mu + 3\sigma)$. This can be done by computing

$$P[(\mu - 3\sigma) < X < (\mu + 3\sigma)] = 1 - \frac{2}{\sigma\sqrt{2\pi}} \int_{\mu+3\sigma}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx. \quad (1.76)$$

When computing probabilities like in (1.75) and (1.76), it is usual to operate alternatively with the so-called normalized Gaussian random variable, which has $\mu = 0$ and $\sigma = 1$. This random variable is often referred to as a $\mathcal{N}(0, 1)$ or *normal* random variable. The reason is that we become able to compute probabilities from Gaussian random variables with any mean and any standard deviation from a single normal density. Moreover, areas under the $\mathcal{N}(0, 1)$ density function are tabulated in several references, facilitating even more our work. As an example, the probability in (1.75) can be computed from the normal Gaussian PDF according to

$$P[X > \gamma] = \frac{1}{\sqrt{2\pi}} \int_{\frac{\gamma-\mu}{\sigma}}^{\infty} e^{-x^2/2} dx. \quad (1.77)$$

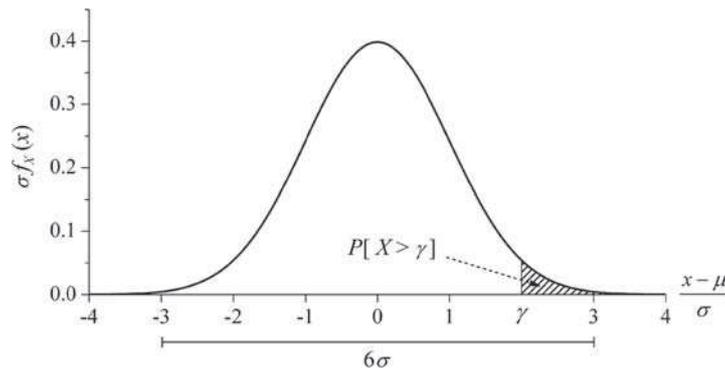


Fig. 1.6 Normalized Gaussian PDF and some probabilities obtained from it

The probability in (1.76) can be computed analogously.

Unfortunately, none of the preceding three integrals has a closed-form analytical solution and we must resort to numerical integration to solve them. The usual numerical solution makes use of the *complementary error function* or some of its variants (see Appendix A.2). The complementary error function is defined by

$$\operatorname{erfc}(u) \triangleq \frac{2}{\sqrt{\pi}} \int_u^{\infty} \exp[-z^2] dz. \quad (1.78)$$

It is apparent from (1.78) that $\operatorname{erfc}(u)$ computes two times the area on the right of u , under a Gaussian PDF of zero mean and variance $1/2$. It is then useful for the purpose of determining areas under a Gaussian PDF.

Applying (1.78) to the probability in (1.75) we have a very useful expression:

$$P[X > \gamma] = \frac{1}{2} \operatorname{erfc} \left(\frac{\gamma - \mu}{\sigma \sqrt{2}} \right), \quad (1.79)$$

from where it follows that the probability in (1.76) can be determined by

$$P[(\mu - 3\sigma) < X < (\mu + 3\sigma)] = 1 - \operatorname{erfc} \left(\frac{3}{\sqrt{2}} \right) \cong 0.9973. \quad (1.80)$$

This interesting result shows that 99.73% of the area under a Gaussian PDF is within the range $(\mu - 3\sigma, \mu + 3\sigma)$. For many practical purposes, these limits are considered the limits of excursion of the underlying Gaussian random variable.

The computation of the complementary error function is a feature of many mathematics software tools like *Mathcad*, *Matlab* and *Mathematica*,³ of spreadsheet software tools like *Excel*, or even in modern scientific calculators. The error function or the complementary error function is also often tabulated in books on probability, statistics and random processes and digital communications. Appendix A.2 provides such a table.

1.9.4 Rayleigh Random Variable

The probability density function, mean, variance, mean square value and moment-generating function of a Rayleigh random variable X with parameter $\Omega = 2\sigma^2$ are, respectively:

1. $f_X(x) = \begin{cases} \frac{2x}{\Omega} e^{-\frac{x^2}{\Omega}}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$
2. $E[X] = \sigma \sqrt{\frac{\pi}{2}}$
3. $\operatorname{var}[X] = \frac{4 - \pi}{2} \sigma^2$
4. $E[X^2] = 2\sigma^2 = \Omega$
5. $\phi_X(s) = 1 + \sigma s e^{\sigma^2 s^2 / 2} \sqrt{\frac{\pi}{2}} \left[\operatorname{erf} \left(\frac{\sigma s}{\sqrt{2}} \right) + 1 \right]$.

(1.81)

The Rayleigh random variable arises, for example, in *complex Gaussian random variables* whose real and imaginary parts are independent and identically distributed (i.i.d.) Gaussian random variables. In these cases the modulus of the complex Gaussian random variable is Rayleigh-distributed.

³ *Mathcad*, *Matlab*, *Mathematica* and *Excel* are trademarks of Mathsoft, Inc., Mathworks, Inc., Wolfram Research, Inc. and Microsoft Corporation, respectively.

As we shall see in Chap. 3, the Rayleigh distribution is often used to characterize the multipath fading phenomena in mobile communication channels.

Simulation 1.3 – Continuous Random Variables



File – CD drive:\Simulations\Probability\Continuous_RV.vsm

Default simulation settings: Frequency = 100 Hz; End = 100 seconds. Mean = 0 and standard deviation = 1 for the thermal (Gaussian) noise source.

Complementary to the previous simulation, this experiment aims at helping the reader to understand some concepts related to the continuous random variables: uniform, Laplace, exponential, Gaussian and Rayleigh.

The first part of the experiment generates a uniform random variable from the quantization error of a voice signal. The quantization process approximates the possibly infinite number of values of the input to a finite number of quantized levels. In this experiment we are simulating a 10-bit uniform quantizer which represents the input waveform by $2^{10} = 1,024$ levels with a step-size $q = 0.0156$ volts, resulting in a total dynamic range of approximately 16 volts.

The quantization error, defined as the subtraction of the quantized signal from the un-quantized signal, has a uniform distribution with parameters, $a = -q/2 = -0.0078$ V and $b = +q/2 = +0.0078$ V. Run the simulation and observe the values of the mean and the variance of the uniform quantization error. Compare them with the theoretical values obtained from (1.71). Observe also the corresponding histogram and its similarity with the brick-wall shape of the theoretical PDF.

The voice signal excerpt used in this experiment has an approximate Laplace distribution, which has probability density function, mean, variance and moment-generating function given respectively by:

$$\begin{aligned}
 1. \quad f_X(x) &= \frac{1}{2r} e^{-|x-\mu|/r}, \quad r > 0 \\
 2. \quad E[X] &= \mu \\
 3. \quad \text{var}[X] &= 2r^2 \\
 4. \quad \phi_X(s) &= \frac{e^{\mu s}}{1 - r^2 s^2}, \quad |s| < \frac{1}{r}.
 \end{aligned}
 \tag{1.82}$$

In this simulation, the scale parameter $r \cong 1.43$ and the mean $\mu = 0$. The variance of X is then $\text{var}[X] \cong 4.09$. Compare these results with those obtained in the simulation. Observe also the histogram of the voice signal and compare to the shape of the Laplace PDF obtained from (1.82).

Now go to the diagram in the second part of the experiment, which aims at analyzing the exponential random variable. You probably may have noticed the similarity between this diagram and the one used for analyzing the Poisson random variable in Simulation 1.2. In fact this is not a coincidence, since the time interval between

Poisson events or the time till the first Poisson event is an exponentially-distributed random variable. The Poisson source was implemented in the same way as in Simulation 1.2, that is, from a binomial source with $p = 0.002$ and $n = 1,000$. Note, however, that the observation interval now is $T_o = 10$ s and that the binomial source was derived from 100 Bernoulli trials per second, totalizing 1,000 trials in 10 s. Again, the mean and the variance of the Poisson random variable are $\alpha = np = 2$ and you will observe something varying around $10\alpha = 20$ Poisson events in the “accumulated successes”, where 10 is the number of observation intervals of 10 seconds during the overall simulation time of 100 seconds.

The rate parameter of the exponential random variable is $\lambda = \alpha/T_o = 2/10 = 0.2$. Then, the mean of this random variable is $1/\lambda = 5$ and its variance is $1/\lambda^2 = 25$.

Increase the simulation “end time” to 10,000 seconds and run the simulation. Compare the estimated mean and variance with the values just calculated. Now, while rerunning the simulation, look inside the block “histograms”. Compare the shape of the histograms with the theoretical Poisson PMF, determined via (1.69), and the exponential PDF, determined via (1.72). Finally, divide the “accumulated successes” by the total simulation time. You will find something around 0.2, which is indeed the rate parameter λ .

Let us move to the third part of the experiment, where a Gaussian random variable is generated and analyzed. The Gaussian source is simulating the thermal noise commonly present in the analysis of communication systems. For the sake of completeness, both mean and standard deviation of the Gaussian random variable can be configured. Using the default simulation settings, run the simulation and observe the estimated values of the mean and variance. Compare them with those determined by the Gaussian source configuration. While looking inside the block “histogram”, run the simulation and observe the similarity between the shape of the histogram and the bell-shaped Gaussian probability density function.

Note also that a block named “ $P[X > u]$ ” has a default value of 2 as its u input. According to (1.79), $P[X > 2] = 0.023$, a value that is estimated approximately in the simulation by using the relative frequency definition for the probability (analyze the construction of the block “ $P[X > u]$ ”). Just to stress this concept, change the mean and the standard deviation of the Gaussian source to 1 and 2, respectively. Now, according to (1.79), $P[X > 2] = 0.309$. Compare this result with the one obtained via simulation. Use a larger simulation “end time” if more accurate estimates are desired.

Finally, the last part of the experiment generates and analyzes a Rayleigh random variable. We shall see later on in this chapter that it is possible to determine the probability density function of a random variable defined as a function of one or more random variables. In this simulation, a Rayleigh random variable is generated as follows: let X and Y be two independent Gaussian random variables with zero mean and standard deviation σ . Define a new random variable R such that

$$R = \sqrt{X^2 + Y^2}. \quad (1.83)$$

This new random variable R is Rayleigh distributed. Moreover, let us define another random variable Θ such that

$$\Theta = \arctan\left(\frac{Y}{X}\right). \quad (1.84)$$

The random variable Θ is uniformly-distributed between $-\pi$ and $+\pi$.

The Gaussian sources in this part of the experiment have zero mean and standard deviation $\sigma = \sqrt{2}/2$. Then, according to (1.81), the Rayleigh random variable will have $E[R] \cong 0.886$, $\text{var}[R] \cong 0.215$ and $E[R^2] = \Omega = 2\sigma^2 = 1$. The uniform random variable will have zero mean and variance $q^2/12 = (2\pi)^2/12 \cong 3.29$. Compare these results with those produced by the simulation. Additionally, have a look at the histograms of the random variables generated and compare them to the probability density functions obtained from (1.71), for the uniform random variable, and from (1.81) for the Rayleigh random variable.

Explore inside the individual blocks. Try to understand how they were implemented. Create and investigate for yourself new situations and configurations of the simulation parameters and try to reach your own conclusions.

1.9.5 Multivariate Gaussian Random Variables

Among the many multivariate distributions, the multivariate Gaussian distribution arises in several practical situations. This is the reason for presenting a brief discussion on this distribution in what follows.

Let $X_1, X_2, \dots, X_i, \dots, X_n$ be a set of Gaussian random variables with means μ_i , variances σ_i^2 and covariances $K_{ij} = \text{cov}[X_i, X_j]$, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, n$. The joint PDF of the Gaussian random variables $\{X_i\}$ is given by

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^T \mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu}_X)\right]}{(2\pi)^{n/2}(\det \mathbf{M})^{1/2}}, \quad (1.85)$$

where \mathbf{M} is the $n \times n$ covariance matrix with elements K_{ij} , \mathbf{X} is the $n \times 1$ vector of random variables, $\boldsymbol{\mu}_X$ is the $n \times 1$ vector of means, T denotes a matrix transposition, \mathbf{M}^{-1} is the inverse of the matrix \mathbf{M} and $\det(\mathbf{M})$ is the determinant of \mathbf{M} .

For the special case of a bivariate or two-dimensional Gaussian random variable we have the following joint probability density function:

$$f_{X_1, X_2}(x_1, x_2) = \frac{\exp\left[-\frac{\sigma_1^2(x_2 - \mu_2)^2 + \sigma_2^2(x_1 - \mu_1)^2 - 2\rho\sigma_1\sigma_2(x_1 - \mu_1)(x_2 - \mu_2)}{2\sigma_1^2\sigma_2^2(1 - \rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}}, \quad (1.86)$$

where ρ is the correlation coefficient between X_1 and X_2 , given by

$$\rho = \frac{K_{12}}{\sigma_1 \sigma_2}. \quad (1.87)$$

Note from (1.86) that if two Gaussian random variables are uncorrelated, their correlation coefficient is zero and, as a consequence, their joint probability density function will be given by the product of the individual densities. Then, these two variables are also statistically independent. In fact, this condition applies to any set of Gaussian random variables, i.e. if Gaussian random variables are pair-wise uncorrelated, they are also independent and their joint probability density function is the product of the densities of the individual random variables in the set.

1.10 Sum of Random Variables and the Central Limit Theorem

Many problems encountered in the study of digital transmission are related to the combination of random phenomena, which commonly appear as the sum of random variables and random processes. In this section we review some of the main concepts related to the sum of random variables, leaving for later the analysis of random processes. Also in this section we present the *central limit theorem*, also in the context of the sum of random variables. This important theorem has applications in several areas and appears frequently in problems related to communication systems.

1.10.1 Mean and Variance for the Sum of Random Variables

Initially, let us define the random variable Y corresponding to the sum of n random variables:

$$Y = \sum_{i=1}^n X_i. \quad (1.88)$$

The expected value of this sum is the sum of the expected values, that is,

$$E[Y] = \sum_{i=1}^n E[X_i]. \quad (1.89)$$

The variance of the sum is given by

$$\text{var}[Y] = \sum_{i=1}^n \text{var}[X_i] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{cov}[X_i, X_j]. \quad (1.90)$$

If the random variables are independent, from (1.55) $\text{cov}[X_i, X_j] = 0$ for $i \neq j$. Then, the variance of the sum is the sum of the variances:

$$\text{var}[Y] = \sum_{i=1}^n \text{var}[X_i]. \quad (1.91)$$

1.10.2 Moments of the Sum of Random Variables

The characteristic function and the moment-generating function were defined in Sect. 1.7.4. The moment theorem defined in that point is also useful in the case of the sum of random variables, as we shall see below.

If the random variables $\{X_i\}$ are independent, the characteristic function of the sum defined in (1.88) is given by

$$\psi_Y(j\omega) = \prod_i \psi_{X_i}(j\omega). \quad (1.92)$$

Moreover, if the random variables have the same distribution we say that they are i.i.d. (independent and identically distributed). In this case we have

$$\psi_Y(j\omega) = [\psi_X(j\omega)]^n. \quad (1.93)$$

Then, by applying the moment theorem given by (1.59), the n -th moment of the sum of random variables can be determined.

1.10.3 PDF of the Sum of Random Variables

Let the random variable $Y = X_1 + X_2$. Its probability density function is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, y - x_1) dx_1 = \int_{-\infty}^{\infty} f_{X_1 X_2}(y - x_2, x_2) dx_2. \quad (1.94)$$

If X_1 and X_2 are independent, the PDF of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x_2) f_{X_2}(x_2) dx_2 = \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1, \quad (1.95)$$

which means that the PDF of the sum of random variables is the convolution of the individual PDFs. In general, the probability density function of the sum of independent random variables is the convolution among the individual densities. However, the computation of the resultant PDF can represent a hard and sometimes intractable mathematical task. This problem can be overcome by applying (1.92) or (1.93) and then the inverse transformation of the characteristic function according to (1.58).

1.10.4 The Central Limit Theorem

The central limit theorem assumes several definitions and can be proved in several different ways. One of its forms can be stated as follows: let $X_1, X_2, \dots, X_i, \dots, X_n$ be a set of independent random variables, each of them having any probability distribution, mean μ_i and variance σ_i^2 . The normalized random variable

$$Z = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \quad (1.96)$$

has, in the limit of $n \rightarrow \infty$, a Normal $\mathcal{N}(0, 1)$ cumulative distribution function. If the random variables $\{X_i\}$ have the same mean μ_X and the same variance σ_X^2 , the random variable

$$Z = \frac{1}{n} \sum_{i=1}^n X_i \quad (1.97)$$

will be Gaussian in the limit, with mean $\mu_Z = \mu_X$ and variance $\sigma_Z^2 = \sigma_X^2/n$.

In other words, if we sum-up n random variables, no matter their individual distributions, the resultant random variable approximates a Gaussian random variable as n is increased.

One of the most widely uses of the central limit theorem is referred to as the Gaussian approximation for the binomial distribution. We know from previous sections that a binomial random variable with parameters p and n is the sum of n Bernoulli random variables having probability of success p . The resultant probability mass function, mean and variance are, respectively:

1. $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$
2. $E[X] = np$
3. $\text{var}[X] = np(1-p)$

(1.98)

From the central limit theorem, for a large n the PMF above approximates a Gaussian density with mean np and variance $np(1-p)$:

$$p_X(x) \cong \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left[-\frac{(x-np)^2}{2np(1-p)}\right]. \quad (1.99)$$

Figure 1.7 illustrates the Gaussian approximation for the binomial distribution, considering $n = 1,000$ and $p = 0.8$. Note that the asymmetry observed in the binomial distribution for $p \neq 0.5$ tend to become practically invisible for a large n .

The central limit theorem has several applications in the design and analysis of digital transmission systems. One of these applications will be illustrated with the help of an example at the end of the next section.

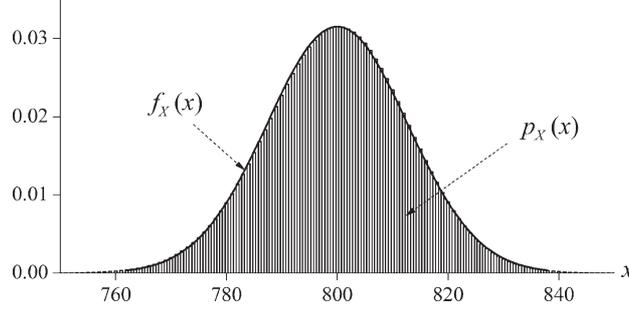


Fig. 1.7 Gaussian approximation for the binomial distribution ($n = 1,000$ and $p = 0.8$)

1.11 Transformations of Random Variables

A problem that frequently arises in practice is to find the probability density function of a random variable $Y = g(X)$, knowing the PDF of X and the mapping rule from one variable into another. Assuming that one value of X produces just one value of Y , the PDF of Y is given by

$$f_Y(y) = \frac{f_X(x)}{|g'(x)|} \Big|_{x=g^{-1}(y)}, \quad (1.100)$$

where the inverse function $g^{-1}(y)$ is simply the value of x written in terms of y .

For the cases where the mapping function $g(X)$ is such that one value of X produces more than one value of Y , firstly $g(X)$ must be divided into a number I of intervals, each of them having a one-to-one correspondence between X and Y . Then (1.100) is applied to each interval according to [10, p. 93]

$$f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \frac{f_X(x_2)}{|g'(x_2)|} + \dots + \frac{f_X(x_I)}{|g'(x_I)|}. \quad (1.101)$$

Generally, suppose we have n random variables arranged in a vector $\mathbf{X} = [X_1, X_2, \dots, X_n]$. Consider the mapping functions $g_i(\mathbf{X})$, $i = 1, 2, \dots, n$ so that we form the vector of random variables $\mathbf{Y} = [Y_1, Y_2, \dots, Y_n]$ through the transformations:

$$Y_1 = g_1(\mathbf{X}), Y_2 = g_2(\mathbf{X}), \dots, Y_n = g_n(\mathbf{X}). \quad (1.102)$$

The joint (or *multivariate*) probability density function of the random variables in \mathbf{Y} is given by

$$f_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n) = \frac{f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)}{|J(x_1, x_2, \dots, x_n)|}, \quad (1.103)$$

where $J(\cdot)$ is referred to as the *Jacobian of the transformation* and is given by the determinant

$$J(x_1, x_2, \dots, x_n) = \det \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_n}{\partial x_1} & \dots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}. \quad (1.104)$$

Example 1.5 – This long example aims at applying several of the concepts presented so far. At the same time, it is intended to illustrate how a typical problem encountered during the analysis of a communication system can be handled with the tools present in this chapter.

Figure 1.8 shows the digital communication system model used for the present analysis. Data bits enter the transmitter, which is responsible for adapting the information to the communication channel. In this example the transmitted signal is a radio-frequency signal and the channel is the so-called *multipath-fading channel*. The behavior of this channel will be discussed in detail in Chap. 3, but for now it suffices to know that the multipath-fading channel causes the received signal to vary its magnitude with time. These variations are normally called *fading*.

To combat the effect of fading, the receiver can be equipped with a technique called *spatial diversity*. In this technique, if the S receiving antennas are sufficiently spaced apart, the signals at their outputs will be uncorrelated and statistically independent. Then, the probability of a deep fading in the S signals will be the product of the probability of a deep fading in one of the signals. As a consequence, by combining the signals at the output of the antennas a signal processing at the receiver will be able to make more reliable decisions about the transmitted data bits. One of the commonly used combining techniques is the MRC (maximal ratio combining), whose some details are presented throughout this example. Our aim here is to analyze some statistical properties of the signal at the output of the MRC combiner.

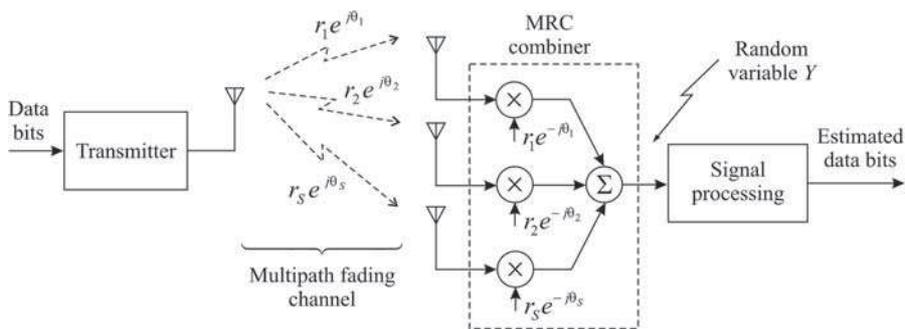


Fig. 1.8 System model for the Example 1.5

The multipath-fading channel behavior can be represented by a multiplicative effect, which means that the signal received by each antenna can be written as

$$Z_i(t) = G_i(t)U(t), \quad i = 1, 2, \dots, S, \quad (1.105)$$

where $G_i(t)$ represents the multiplicative effect of the fading from the transmit antenna to the i -th receive antenna and $U(t)$ is the transmitted signal. Generally speaking, $Z_i(t)$, $G_i(t)$ and $U(t)$ can be complex and in fact are *random processes*, a subject treated in the second half part of this chapter. However, without loss of generality we will assume that these signals can be treated as random variables. Furthermore, since we are interested in analyzing the effect of the channel and of the receiver combiner, we shall assume that $|U(t)| = 1$ and $\arg[U(t)] = 0$. Then we can rewrite (1.105) in terms of random variables, yielding

$$Z_i = G_i. \quad (1.106)$$

In a multipath-fading channel, the received signal comes from an infinite set of propagation paths, which justify the term “multipath”. An example of this behavior happens when you are using your cell-phone in a dense urban area, where its antenna receives signals coming from virtually all directions due to the echoes of the transmitted signal in buildings, trees, etc.

Assuming that the transmitted signal is represented by a complex function, from the central limit theorem we can state that the real and the imaginary parts of the received signal are approximately Gaussian-distributed. Then we conclude that G_i is a complex Gaussian random variable, which is given by

$$G_i = W_i + jZ_i, \quad (1.107)$$

where W_i and Z_i are zero-mean, independent Gaussian random variables with the same variance σ^2 . Omitting the subscript i for notational simplicity, the joint probability density function of W_i and Z_i is then

$$\begin{aligned} f_{WZ}(w, z) &= f_w(w)f_z(z) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{w^2}{2\sigma^2}\right] \times \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{z^2}{2\sigma^2}\right] \\ &= \frac{1}{2\pi\sigma^2} \exp\left[-\frac{w^2 + z^2}{2\sigma^2}\right], \end{aligned} \quad (1.108)$$

which is a special case of (1.86), with $X_1 = W$, $X_2 = Z$, $\sigma_1 = \sigma_2 = \sigma$, $\mu_1 = \mu_2 = \mu$ and correlation coefficient $\rho = 0$. A bivariate PDF like this is depicted in Fig. 1.9, just for illustration purposes.

Now let us determine the probability density function of the magnitude (modulus) and the phase (argument) of the received signal in each antenna. In terms of the components of the complex Gaussian random variable we have

$$\begin{aligned}
|G_i| = R_i &= \sqrt{W_i^2 + Z_i^2}, \quad R_i \geq 0 \quad \text{and} \\
\arg(G_i) = \Theta_i &= \arctan\left(\frac{Z_i}{W_i}\right), \quad -\pi \leq \Theta_i \leq \pi,
\end{aligned} \tag{1.109}$$

from where we can additionally write the relations:

$$\begin{aligned}
1. \quad W_i &= R_i \cos \Theta_i \Rightarrow w_i = r_i \cos \theta_i \\
2. \quad Z_i &= R_i \sin \Theta_i \Rightarrow z_i = r_i \sin \theta_i \\
3. \quad R_i^2 &= W_i^2 + Z_i^2 \Rightarrow r_i^2 = w_i^2 + z_i^2
\end{aligned} \tag{1.110}$$

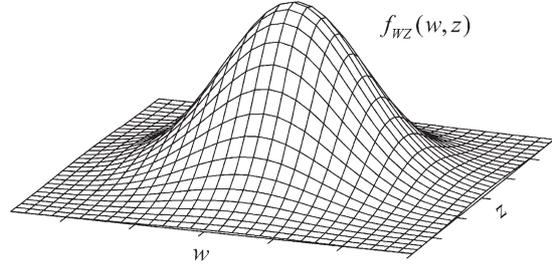


Fig. 1.9 Bivariate Gaussian probability density function

Dropping again the subscript i for notational simplicity, from (1.103) we obtain the transformation rule

$$f_{WZ}(w, z) = \frac{f_{R\Theta}(r, \theta)}{|J(r, \theta)|} \Rightarrow f_{R\Theta}(r, \theta) = f_{WZ}(w, z)|J(r, \theta)|, \tag{1.111}$$

where the Jacobian of the transformation, defined in (1.104), is

$$\begin{aligned}
J(r, \theta) &= \begin{vmatrix} \frac{\partial(r \cos \theta)}{\partial r} & \frac{\partial(r \cos \theta)}{\partial \theta} \\ \frac{\partial(r \sin \theta)}{\partial r} & \frac{\partial(r \sin \theta)}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} \\
&= r \cos^2 \theta + r \sin^2 \theta = r.
\end{aligned} \tag{1.112}$$

Applying (1.108) and (1.112) in (1.111), also using the last relation given in (1.110), we obtain the probability density function of the random variable representing the magnitude of the received signal, which in this case corresponds to the magnitude of the channel fading affecting the signal received by each antenna:

$$f_{R\Theta}(r, \theta) = \frac{r}{2\pi\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2}\right], \quad r \geq 0, \quad |\theta| \leq \pi. \tag{1.113}$$

The marginal PDF $f_R(r)$ is obtained by integrating $f_{R\Theta}(r, \theta)$ in θ , yielding

$$f_R(r) = \int_{-\pi}^{\pi} f_{R\Theta}(r, \theta) d\theta = \frac{r}{\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2}\right], \quad (1.114)$$

from where we conclude that R and Θ are also independent, since we have

$$f_{R\Theta}(r, \theta) = f_R(r)f_{\Theta}(\theta) = \frac{r}{\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2}\right] \times \frac{1}{2\pi}. \quad (1.115)$$

Then, the fading magnitude affecting the signal received by each antenna is Rayleigh-distributed and the phase is uniformly-distributed in $(-\pi, \pi]$. These densities are plotted in Fig. 1.10, considering $\sigma = \sqrt{2}/2$. Since $E[R_i^2] = \Omega = 2\sigma^2$, the mean square value of the Rayleigh random variable is $\Omega = 1$. This is usually adopted to simplify communication systems analysis, since $E[R_i^2] = 1$ means that the fading only affect the instantaneous received power, not affecting its average.

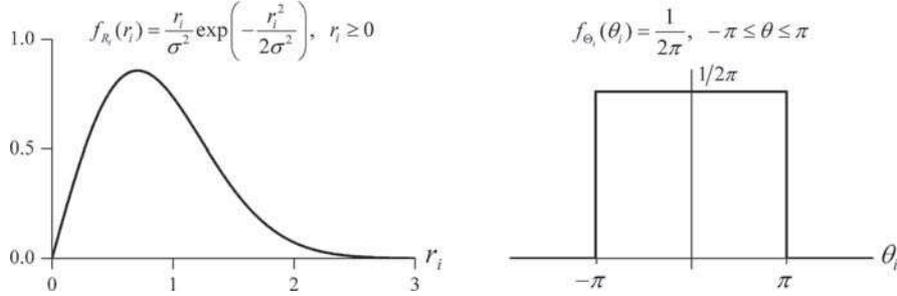


Fig. 1.10 Rayleigh magnitude and uniform phase of a typical multipath-fading

Before added-up to form the random variable Y from which the receiver signal processing will extract the estimated data bits, the signal at the output of each receiving antenna is multiplied by the complex-conjugate of the fading affecting the corresponding signal. This intrinsic characteristic of the MRC combiner will cancel the phase rotation introduced by the channel, attenuate the weak signals and reinforce the stronger ones. Note that this process of amplifying (or not attenuating) strong signals and making the weak signals even weaker seems to be contrary to our intuition.

After the complex multiplication, at the input of the summing process we shall have the random variables $X_i = R_i^2$, $i = 1, 2, \dots, S$, from where we have $dx_i/dr_i = 2r_i$. Then, the probability density function of these new random variables X_i can be determined through the transformation in (1.100), as follows:

$$\begin{aligned} f_X(x) &= \frac{f_R(r)}{|g'(r)|} \Big|_{x=g^{-1}(y)} = \frac{f_R(r)}{|dx/dr|} \Big|_{x=g^{-1}(y)} = \frac{f_R(\sqrt{x})}{|2\sqrt{x}|} \\ &= \frac{1}{2\sigma^2} \exp\left[-\frac{x}{2\sigma^2}\right] = \frac{1}{\Omega} \exp\left[-\frac{x}{\Omega}\right], \quad x \geq 0, \end{aligned} \quad (1.116)$$

where, again, we have dropped the subscripts for simplicity. The PDF in (1.116) is a *Chi-square* (χ^2) PDF with two degrees of freedom [13, p. 41], which corresponds to an exponential PDF with rate parameter $1/\Omega$.

The mean square value of the random variable X can be obtained from the second derivative of its moment-generating function $\phi_X(s)$, according to (1.61). Applying this operation to the $\phi_X(s)$ given in (1.72) we obtain

$$E[X^2] = \left. \frac{d^2 \phi_X(s)}{ds^2} \right|_{s=0} = \left. \frac{d^2}{ds^2} \left(\frac{1/\Omega}{1/\Omega - s} \right) \right|_{s=0} = 2\Omega^2. \quad (1.117)$$

The random variable at the output of the MRC combiner will be

$$Y = \sum_{i=1}^S X_i. \quad (1.118)$$

At the beginning of this example we have assumed that the receiving antennas are sufficiently apart from each other to guarantee that the signals at their outputs are independent and, as a consequence, uncorrelated. Then, the PDF of Y can be determined by the convolution among the PDFs of X_i . Since this represents a hard mathematical problem, we resort to the property given in (1.92), which states that the characteristic function of a sum of independent random variables is the product of the corresponding characteristic functions. Moreover, if the random variables are also identically-distributed, according to (1.93) the characteristic function of the sum of S random variables is the S -th power of the characteristic function of one of the random variables considered in the sum:

$$\psi_Y(j\omega) = [\psi_X(j\omega)]^S. \quad (1.119)$$

According to (1.58), the probability density function of the random variable at the output of the MRC combiner can be determined through the inverse transform

$$f_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega y} [\psi_X(j\omega)]^S d\omega, \quad (1.120)$$

where

$$\psi_X(j\omega) = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx = \int_0^{\infty} \frac{1}{\Omega} \exp\left(-\frac{x}{\Omega}\right) e^{j\omega x} dx. \quad (1.121)$$

From Chap. 2, Table 2.4, we have the Fourier transform pair

$$\exp(-\alpha t)u(t), \alpha > 0 \Leftrightarrow \frac{1}{\alpha + j2\pi f}. \quad (1.122)$$

By analogy to (1.122) and recalling the fact that (1.121) has the opposite signal of the true Fourier transform, we obtain

$$\psi_X(j\omega) = \frac{1}{\Omega} \int_{-\infty}^{\infty} e^{-x/\Omega} u(x) e^{j\omega x} dx = \frac{1/\Omega}{1/\Omega - j\omega} = \frac{1}{1 - j\omega\Omega}. \quad (1.123)$$

Then, applying (1.119),

$$\psi_Y(j\omega) = [\psi_X(j\omega)]^S = \frac{1}{(1 - j\omega\Omega)^S}. \quad (1.124)$$

With this result in (1.120) we have

$$f_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega y} (1 - j\omega\Omega)^{-S} d\omega. \quad (1.125)$$

From [9, p. 170] and [10, p. 116], the probability density function and the characteristic function of a gamma random variable are respectively given by

$$\begin{aligned} f_Y(y) &= \frac{\lambda(\lambda y)^{\alpha-1} e^{-\lambda y}}{\Gamma(\alpha)}, \quad y \geq 0, \text{ and} \\ \psi_Y(j\omega) &= \frac{\lambda^\alpha}{(\lambda - j\omega)^\alpha}. \end{aligned} \quad (1.126)$$

Comparing the characteristic function in (1.126) with (1.124) we can find that $\lambda = 1/\Omega$ and $\alpha = S$ in the gamma distribution. Then, with these results in (1.126), the PDF of the random variable at the output of the MRC is finally obtained:

$$f_Y(y) = \frac{y^{S-1} e^{-y/\Omega}}{\Omega^S (S-1)!}, \quad y \geq 0, \quad (1.127)$$

where we have used the relation $\Gamma(S) = (S-1)!$, valid for $S > 0$ and integer. The PDF in (1.127) is a special form of the gamma PDF and is very similar to a *Chi-square* (χ^2) distribution with $2S$ degrees of freedom.

As a complementary analysis, let us determine the mean and the mean square value of the random variable Y . They are given respectively by:

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} y f_Y(y) dy = \frac{1}{\Omega^S (S-1)!} \int_0^{\infty} y^S e^{-y/\Omega} dy \quad \text{and} \\ E[Y^2] &= \int_{-\infty}^{\infty} y^2 f_Y(y) dy = \frac{1}{\Omega^S (S-1)!} \int_0^{\infty} y^{S+1} e^{-y/\Omega} dy. \end{aligned} \quad (1.128)$$

From Appendix A.4 we have the definite integral

$$\int_0^{\infty} y^n e^{-ay} dy = \frac{n!}{a^{n+1}}, \quad a > 0, \quad n = 0, 1, 2, \dots \quad (1.129)$$

Applying this result in (1.128) we obtain, after some simplifications,

$$\begin{aligned} E[Y] &= S\Omega \text{ and} \\ E[Y^2] &= S(S+1)\Omega^2. \end{aligned} \quad (1.130)$$

As a final analysis, let us determine the ratio between the average signal power and the average noise power at the MRC combiner output. We start this analysis by defining the magnitude of a new random variable at each combiner input as aR_i , where a is the random variable corresponding to the magnitude of the transmitted signal. We also denote the normalized (over a 1 ohm load) instantaneous signal power at the combiner's output as P and the normalized average noise power in each combiner input as η . Then, the instantaneous power is

$$P = \left[\sum_{i=1}^S (aR_i)R_i \right]^2 = a^2 \left(\sum_{i=1}^S R_i^2 \right)^2. \quad (1.131)$$

The normalized instantaneous noise power at the combiner's output will be

$$N = \sum_{i=1}^S \eta R_i^2 = \eta \sum_{i=1}^S R_i^2. \quad (1.132)$$

The average signal-to-noise ratio (SNR) is then

$$\begin{aligned} \langle \text{SNR} \rangle &= E \left[\frac{P}{N} \right] = E \left[\frac{a^2 \left(\sum_{i=1}^S R_i^2 \right)^2}{\eta \sum_{i=1}^S R_i^2} \right] = E \left[\frac{a^2 \sum_{i=1}^S R_i^2}{\eta} \right] \\ &= \sum_{i=1}^S E[(aR_i)^2/\eta] = \sum_{i=1}^S \langle \text{SNR}_i \rangle, \end{aligned} \quad (1.133)$$

where $E[(aR_i)^2/\eta]$ is the average signal-to-noise ratio in each of the combiner's input. Then, with the result given in (1.133) we conclude that the average SNR at the MRC combiner's output is the sum of the signal-to-noise ratios in the combiner's inputs. We now are able to understand why the use of a combiner can bring lower probabilities of decision errors by the receiver's signal processing block. We accomplish this with the help of the next computer simulation.

Simulation 1.4 – A Communication System Analysis



File – CD drive:\Simulations\Probability\MRC.vsm

Default simulation settings: Frequency = 100 Hz; End = 1,000 seconds. Second moment of the Rayleigh random variables $E[R^2] = 1.5$.

This experiment aims at simulating the scenario described in Example 1.5. The system shown in Fig. 1.8 is partially implemented using VisSim/Comm. The transmitted signals were simulated by constants $a_i = a = 1$, permitting that the variations observed in the received signal are attributed to the channel fading R_i . Three received signals are being generated to simulate the use of $S = 3$ receiving antennas. Each of these signals is affected by a multiplicative Rayleigh fading with configurable mean square value $E[R^2] = \Omega = 2\sigma^2$. The phase rotation component of the fading did not need to be considered, since we are assuming that the MRC combiner would cancel it off anyway. The received signals enter the MRC combiner to form the random variable Y at the combiner's output.

Histograms and moments estimates are generated for one of the Rayleigh fading random variables (R_i), for one of the summing inputs inside the MRC combiner (X_i) and for the MRC output random variable (Y).

Average signal-to-noise ratio (SNR) estimates are also computed and that is the reason for the repeated implementation of the combiner in the lower part of the simulation worksheet, where we have only Gaussian noise sources as the MRC input. In other words, we need two systems to analyze signal and noise separately.

The upper SNR measurement corresponds to the output of the combiner and the average SNR is estimated according to

$$\langle \text{SNR} \rangle = E \left[a^2 \left(\sum_{i=1}^S R_i^2 \right)^2 / \eta \sum_{i=1}^S R_i^2 \right], \quad (1.134)$$

where $\eta = 1$ is the average noise power in each of the MRC inputs. Observe the simulation diagram and note that the MRC output corresponds exactly to the square-root of the numerator in (1.134). Note also that, since $a = 1$, the value at the MRC output times the average noise power in one of its inputs gives the instantaneous noise power at the combiner's output. Have a look inside the corresponding SNR estimation block and try to associate the operations with (1.134).

The expected value estimation blocks in this simulation are applying two concepts not yet discussed in this chapter. For now it suffices to know that in some cases statistical averages can be realized through time averages. We shall present this concept in more detail in the study of *ergodic random processes*. Moreover, these time averages are being estimated here through *sample averages*, a concept also treated in more detail later on in this chapter.

Continuing the analysis of the simulation diagram, the intermediate SNR measurement is estimating the average signal-to-noise ratio at one of the combiners' inputs. It is performing the operation

$$\langle \text{SNR}_i \rangle = E [(aR_i)^2 / \eta]. \quad (1.135)$$

Analyze the operations inside this block and confirm the realization of (1.135).

The SNR estimation at the bottom of the worksheet is a *wrong* SNR measurement. It is dividing the average signal power by the average noise power. Although this measurement seems to be intuitively satisfying, we must not forget that a given instantaneous fading magnitude affects simultaneously the signal and the noise components. Separate averages do not take this fact into account and, as a consequence, produce incorrect SNR estimates.

Run the simulation using the default settings. After it is finished, observe the estimated SNR at one of the combiner's inputs and compare the result with the SNR at the combiner's output. Note that the output SNR is approximately three times the input SNR, since $S = 3$. Compare this result with the theoretical expected value given in (1.133). Observe also that the *wrong* SNR estimation has indeed produced an incorrect result.

Still using the default settings open the "histogram and moments" block. Run the simulation and observe the shapes of the histograms: the upper one resembles the Rayleigh PDF of R_i given in (1.114). The intermediate one resembles the exponential PDF of X given in (1.116), and the histogram at the bottom shows the shape of the PDF given in (1.127), which corresponds to the gamma random variable Y at the output of the combiner. The mean and second moment for each of these random variables are expected to be:

$$\begin{aligned} 1. E[R_i] &= \sigma \sqrt{\pi/2} = \sqrt{\Omega\pi/4} = \sqrt{1.5\pi/4} \cong 1.085 \\ 2. E[R_i^2] &= 2\sigma^2 = \Omega = 1.5 \\ 3. E[X_i] &= 1/\lambda = \Omega = 1.5 \\ 4. E[X_i^2] &= 2\Omega^2 = 2 \times 1.5^2 = 4.5 \\ 5. E[Y_i] &= S\Omega = 3 \times 1.5 = 4.5 \\ 6. E[Y_i^2] &= S(S+1)\Omega^2 = 3(3+1)1.5^2 = 27. \end{aligned} \quad (1.136)$$

Compare the above theoretical results with those estimated in the simulation. Explore inside the individual blocks. Try to understand how they were implemented. Create and investigate for yourself new situations and configurations of the simulation parameters and try to reach your own conclusions.

1.12 Parameter Estimation via Sample Mean

When dealing with random data, it is common to have no a priori information about the statistical properties of the samples we have. The analysis of such data would

be impossible if the samples did not show some average behavior when more and more data are analyzed.

In this section we present the concept of sample mean. It is a tool that permits the average behavior of the data to be unveiled, allowing for the estimation of some statistical properties about the samples. Additionally, this estimation is always attached to some degree of uncertainty that can be controlled through the analysis of confidence intervals, a topic also covered in this section.

1.12.1 The Sample Mean

Let the random variable Y correspond to the sum of n independent and identically distributed random variables $X_1, X_2, \dots, X_i, \dots, X_n$. The *sample mean* is another random variable defined according to

$$M_n(X) = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.137)$$

The mean and the variance of the sample mean are given respectively by

$$\begin{aligned} E[M_n(X)] &= E[X] \text{ and} \\ \text{var}[M_n(X)] &= \frac{\text{var}[X]}{n}. \end{aligned} \quad (1.138)$$

From the mean in (1.138) we conclude that the sample mean is an unbiased⁴ estimator for $E[X]$. From the variance in (1.138) we conclude that the value of the sample mean converges to $E[X]$ as n goes to infinity.

Note that there is a connection between the sample mean and the central limit theorem: as n increases, $M_n(X)$ approaches a Gaussian random variable with mean $E[X]$ and variance going towards zero.

The sample mean is useful not only for estimating the first moment of a random variable. For this reason it is sometimes called *sample average*. For example, if we compute the sample average $M_n(X^2) = (X_1^2 + X_2^2 + \dots + X_i^2 + \dots + X_n^2)/n$ we are estimating the second moment of the random variable X . This gives to the sample mean a broader use in practical applications, as we shall see below with the help of an example.

Example 1.6 – By observing the quotation of the coffee bag from the stock exchange in different periods of 100 days we shall obtain values that can be associated to the random variables $X_1, X_2, \dots, X_i, \dots, X_{100}$. Similarly, by observing the

⁴ An estimate \hat{R} of the parameter r is said to be *unbiased* if $E[\hat{R}] = r$. If $E[\hat{R}] \neq r$ the estimator is said to be *biased* [18, p. 280].

negotiated volume of points in the New York Stock Exchange we shall obtain values that can be associated to the random variables $Y_1, Y_2, \dots, Y_i, \dots, Y_{100}$. Suppose you were given 100 values for the quotation of the coffee bag, in US Dollars, and 100 values of the negotiated volume in the New York Stock Exchange, in points. These values are shown in the tables below, so the reader can reproduce the results of this example if desired. Note that these are hypothetical values, not necessarily corresponding to actual data.

Values for X

103	100	102	108	99	99	119	98	94	92
98	85	104	85	108	83	86	98	88	88
97	104	115	96	120	109	109	90	109	94
102	95	114	89	97	95	118	98	92	105
98	98	101	88	103	90	107	93	101	90
107	96	85	118	104	112	96	82	92	101
88	125	102	92	96	89	98	105	107	115
88	105	93	111	88	106	106	92	90	102
100	112	108	90	100	101	104	116	95	101
99	91	119	105	99	92	102	99	103	104

Values for Y

64080	50460	51858	39682	48308	52654	52846	57635	53347	51904
44796	53025	66087	50035	51642	45370	51301	51892	47067	42859
46069	45692	49684	45753	50100	51099	53043	45920	46629	57238
51098	46240	52957	41834	50989	50382	50692	55374	56924	42253
44355	57226	48398	46604	52469	55984	52124	65964	48246	49163
44727	41056	50876	48261	52639	50935	60812	51555	49420	54409
43682	50468	48305	46321	51802	44650	45323	55115	46890	51920
52418	48200	50848	45794	43510	52829	48762	54778	40825	47765
57259	45158	49284	50983	49550	49833	53183	45777	52950	46808
51945	50495	45480	43774	59169	55923	55935	45682	48602	50394

In what follows we give some statistical averages estimated via sample mean.

$$\text{Mean: } \mu_X = \frac{1}{n} \sum_{i=1}^n X_i \cong 99.85, \quad \mu_Y = \frac{1}{n} \sum_{i=1}^n Y_i \cong 50,204$$

$$\text{Variance: } \sigma_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu_X)^2 \cong 85.907, \quad \sigma_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \mu_Y)^2 \cong 2.543 \times 10^7$$

$$\text{Second moment: } E[X^2] = \frac{1}{n} \sum_{i=1}^n X_i^2 \cong 1.006 \times 10^4, \quad E[Y^2] = \frac{1}{n} \sum_{i=1}^n Y_i^2 \cong 2.546 \times 10^9$$

$$\text{Correlation: } E[XY] = \frac{1}{n} \sum_{i=1}^n X_i Y_i \cong 5.013 \times 10^6$$

$$\text{Covariance: } \text{cov}[X, Y] = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y) \cong 44.086$$

$$\text{Correlation coefficient: } \rho = \frac{\text{cov}[X, Y]}{\sqrt{\sigma_X^2} \sqrt{\sigma_Y^2}} \cong 9.53 \times 10^{-4}$$

Comments on the Above Results:

- Note that we have made the division by $n - 1$ instead of n in the estimation of the variance and of the covariance, since the division by n would produce biased estimators [18, p. 284]. However, for large n the difference between the biased and the unbiased estimators is small and the use of $n - 1$ or n is indifferent. Moreover, if we know the true values of mean inside the variance and covariance computations, the division by n produces unbiased estimators.
- We are not able to precisely estimate the shapes of the probability density functions of the random variables X and Y through histograms, since the sample size is small (see Example 1.4 for more details on this subject).
- Suppose that we want to use the New York Stock Exchange data as a reference for predicting the variations in the price of a coffee bag. From the averages obtained we notice that, despite of the high correlation and apparently high covariance, the correlation coefficient is approximately zero, indicating that X and Y are uncorrelated. We would have inferred erroneously if we have said that an increase in the New York Stock Exchange quotation would probably have been accompanied by an increase in the price of the coffee bag.

1.12.2 The Confidence Interval on the Sample Mean

In this subsection we shall quantify the precision of the parameter estimation made via sample mean. This quantification will be made through the *confidence interval* and the *confidence coefficient*.

The analysis of confidence interval is not restricted to those cases mentioned in this section. The reader is encouraged to consult the book edited by H. Pham [12] for a broader treatment on this and many other related subjects.

The Chebyshev inequality [9, p. 181] states that, given a random variable X with expected value $E[X]$ and variance $\text{var}[X]$, for any positive number c ,

$$P[|X - E[X]| \geq c] \leq \frac{\text{var}[X]}{c^2}. \quad (1.139)$$

Applying this inequality to the definition of the sample mean we obtain

$$P[|M_n(X) - E[X]| \geq c] \leq \frac{\text{var}[M_n(X)]}{c^2} = \frac{\text{var}[X]}{nc^2} = \alpha. \quad (1.140)$$

This inequality states that the probability that the sample mean is distant from the true mean more than a value c is less than or equal to α . In other words, at least $100 \times (1 - \alpha)\%$ of the values of the sample mean will be in the range $(E[X] - c, E[X] + c)$. Yet another interpretation, the value of the sample mean will be in the range $(E[X] - c, E[X] + c)$ with probability $(1 - \alpha)$.

The length $2c$ is called *confidence interval* and the value $(1 - \alpha)$ is called *confidence coefficient*. Then, the size of the confidence interval gives the precision we want on the sample mean and the confidence coefficient gives us an idea about how many values of the sample mean will be in the confidence interval. The sample size, n , governs the overall desired precision.

The confidence interval obtained via the Chebyshev inequality is somewhat loose or pessimistic in many cases. This means that, for most of the applications, more than $100 \times (1 - \alpha)\%$ of the values of the sample mean will be within the confidence interval $2c$. However, the Chebyshev inequality approach is useful when we do not know *a priori* the distribution of the data.

Fortunately, most of the cases related to communication systems can resort to the central limit theorem and consider that the sample mean is Gaussian-distributed. This is often the case in Monte Carlo simulations, where a large number of random values are computer-generated according to the random phenomena of interest and the desired parameter is estimated via sample mean. The following example aims at clarifying these statements.

Example 1.7 – In this example we shall analyze the confidence interval on the estimate of the bit error probability P_b in a digital communication system. We start by recalling that P_b can be computed through the relative frequency interpretation of probability, that is,

$$P_b = \lim_{n \rightarrow \infty} \frac{n_E}{n}, \quad (1.141)$$

where n_E is the number of bits in error and n is the total number of transmitted bits. The estimate of P_b is made in practice through a finite sample size and, in this case, P_b is usually referred to as the *bit error rate* (BER), which is given by

$$\text{BER} = \frac{n_E}{n}. \quad (1.142)$$

A well-known rule-of-thumb states that we shall have a “good” estimate of the BER if we wait for the occurrence of 100 bit errors. The aim of this example is to obtain a better judgment on how good is the BER estimation, in the light of the confidence interval analysis.

A usual measurement of the precision of an estimate made via sample mean is known as the *normalized standard error* *standard error*, which can be defined by

$$S_\mu = \frac{\sqrt{\sigma_e^2}}{\mu}, \quad (1.143)$$

where the numerator corresponds to the square root of the estimator error variance, and is called the *standard error* of this estimator. For the present context, the normalized standard error can be written as

$$S_{\mu_X} = \frac{\sqrt{\text{var}[M_n(X)]}}{\mu_X} = \frac{\sigma_X}{\sqrt{n}\mu_X}. \quad (1.144)$$

We know that the number of errors in a block of n bits has a binomial distribution with parameters n and $p = P_b$. Then, the sample mean $M_n(X)$ corresponds to the desired BER estimate. In this case we can rewrite the normalized standard error, yielding

$$S_p = \frac{\sqrt{p(1-p)/n}}{p} = \sqrt{\frac{1-p}{np}}. \quad (1.145)$$

Since the probability of error is a small quantity (compared to 1) in real communication systems, we can approximate (1.145) according to

$$S_p \cong \frac{1}{\sqrt{np}}. \quad (1.146)$$

If we expect 100 errors in a block of n bits and the probability of error is p , then the mean of the corresponding binomial random variable is $np = 100$. With this value in (1.146) we have $S_p = 0.1$. In light of (1.144), $S_p = 0.1$ means that the standard deviation of the sample mean is $0.1p$.

Considering a confidence coefficient of 0.9, in what follows we determine the confidence interval, assuming that, if $np = 100$ and p is small, n will be sufficiently large, allowing us to approximate the binomial sample mean to a Gaussian random variable. First we recall the calculation of the area under a Gaussian tail given in (1.79), repeated here for convenience:

$$P[X \geq \gamma] = \frac{1}{2} \text{erfc} \left(\frac{\gamma - \mu}{\sigma\sqrt{2}} \right). \quad (1.147)$$

Additionally we observe that

$$P[|M_n(X) - E[X]| \geq c] = \alpha = 2P[M_n(X) \geq (E[X] + c)]. \quad (1.148)$$

Applying this relation in (1.147) we obtain

$$\alpha = \text{erfc} \left(\frac{c}{\sqrt{\text{var}[M_n(X)]}\sqrt{2}} \right). \quad (1.149)$$

Using the confidence coefficient $(1 - \alpha) = 0.9$ and the standard deviation of the sample mean of $0.1p$ we have

$$0.1 = \operatorname{erfc}\left(\frac{c}{0.141p}\right). \quad (1.150)$$

From the computation of the complementary error function we obtain

$$\frac{c}{0.141p} \cong 1.16309 \Rightarrow c = 0.164p. \quad (1.151)$$

From the above result we conclude that, since the confidence coefficient is 0.9, around 90% of the values of the estimated BER will be in the range $(P_b - 0.164P_b, P_b + 0.164P_b)$. Since normally we do not know the value of P_b , by waiting for 100 bit errors we expect that around 90% of the estimated BER values will be distant from the true value of P_b in a maximum of $\pm 16.4\%$. This is indeed a quite good estimate, as subjectively stated at the beginning of this example.

Simulation 1.5 – Confidence Interval on BER Estimations



File – CD drive:\Simulations\Probability\Confidence_BER.vsm

Default simulation settings: Frequency = 1 Hz; End = 500,000 seconds. Number of bit errors: 100. Theoretical probability of error: 0.02. Confidence coefficient: 0.9.

In this experiment we consider the bit error rate (BER) estimation problem given in Example 1.7, and investigate the convergence of the estimation using a confidence interval analysis. Open the simulation file indicated in the header and follow the description given below.

Data bits are randomly generated at 1 bit per second. These bits are transmitted through a binary symmetric channel whose behavior was already described in Example 1.2. The crossover probability of the BSC channel can be configured to simulate different values for the bit error probability P_b . At the output of the BSC, the received bits are compared with the transmitted ones and an estimate of the BER is made according to (1.142), where the number of expected bits in error per blocks, n_E , can be configured and n is the number of transmitted bits, calculated according to $n = n_E/P_b$.

A histogram of the estimated BER is computed and a time plot registers the confidence interval limits and the estimated BER values. These limits are computed according to (1.149), given the configurable confidence coefficient $(1 - \alpha)$.

Using the default settings, open the “time plot and BER histogram” block and run the simulation. The points in blue in the time plot represent the values of the estimated BER in each block of $n = n_E/P_b$ transmitted bits. The straight lines in black crosses correspond to the limits of the confidence interval. Observe that most of the BER estimates are in-between these limits.

Above the time plot you can see the estimated probability of points inside the limits of the confidence interval, which corresponds to the confidence coefficient. Below the time plot you can see a histogram of the BER estimates. In order to see more precise results, increase the simulation end time to 5,000,000 seconds and run the simulation. After the simulation ends, compare the estimated confidence coefficient with the configured one. Observe the shape of the histogram and note that it resembles a Gaussian distribution.

Now, reset the simulation to its default configuration and then reduce the number of expected bits in error per block from 100 to 10. Open the “time plot and BER histogram” plot and run the simulation. Note that the histogram now resembles a discrete probability mass function. This is consistent with the fact that the adoption of 10 expected bits in error has reduced the number of transmitted bits per block to $n = n_E/P_b = 10/0.02 = 500$, causing the non convergence of the binomial distribution of the errors to a Gaussian density. Note also that the confidence interval has been increased to accommodate $100 \times (1 - \alpha)\% = 90\%$ of the BER estimates. In other words, we now have a lower precision, since the variance of the estimates has been increased.

Analyze the block “computation of the confidence interval” and compare its implementation with the theoretical expressions. Compare also the computed values of $(p - c)$ and $(p + c)$ with the theoretical ones.

Recall that we have fixed the number of transmitted bits per block, n , according to an expected number of bit errors per block, n_E . It is worth mentioning that if the simulation were implemented in a way that n will vary to guarantee a fixed n_E , the distribution of the estimated BER would not be Gaussian anymore. The investigation about this behavior and its consequences on the confidence interval analysis is left to an exercise for the reader. If you have a full version of VisSim/Comm, modify the experiment so that n_E is fixed and n is varied accordingly. *Hint*: modify the components of the “BER estimation” block.

Explore inside the individual blocks. Try to understand how they were implemented. Create and investigate for yourself new situations and configurations of the simulation parameters and try to reach your own conclusions.

1.12.3 The Law of Large Numbers

The law of large numbers is put into focus whenever we want to give the notion that even rare events can show up if a sufficiently large number of chances are given. In statistical terms, with a large sample size it is more likely that the statistical characteristics of the population are unveiled, and with greater precision.

The *weak law of large numbers* is mathematically stated as:

$$\begin{aligned} 1. \lim_{n \rightarrow \infty} P[|M_n(X) - \mu_X| \geq c] &= 0 \\ 2. \lim_{n \rightarrow \infty} P[|M_n(X) - \mu_X| < c] &= 1. \end{aligned} \quad (1.152)$$

The first expression in (1.152) states that $M_n(X) - \mu_X$ tend to be small with high probability as n is increased. It does not state that $M_n(X) = \mu_X$ with high probability. It also does not state that $P[|M_n(X) - \mu_X| \geq c] = 0$. Both a non-zero value for c and for the probability are necessary here, despite of the fact that by increasing n it is possible to make both as small as we want. The second expression in (1.152) states that $M_n(X)$ converges to μ_X when n is increased. This is known as *probabilistic* or *stochastic convergence*.

The *strong law of large numbers* can be written as

$$P \left[\lim_{n \rightarrow \infty} M_n(X) = \mu_X \right] = 1. \quad (1.153)$$

It means that $M_n(X)$ converges to μ_X with probability 1. It also means that the sample mean does not converge to any other value different from μ_X .

1.13 Generation of Random Numbers

Currently, research and development activities in engineering, especially in communication systems engineering, make extensive use of computer simulations. Moreover, since most of the systems created or analyzed through engineering tools are affected by some random phenomena, it is of major importance that the computer simulation of such systems is able to reproduce these random phenomena with accuracy and, preferably, in a short time frame.

Usually, simulations are based on Monte Carlo methods in which random (actually pseudo-random) numbers generation is the most crucial part [2]. In essence, in Monte Carlo methods a large amount of these numbers is generated to simulate the random variables within the system and to permit the estimation of system parameters through sample averages.

In this section we present some algorithms for generating random numbers. These algorithms will be studied without a detailed analysis about the *goodness-of-fit* of the results. For a deeper treatment on the subject, the interested reader is referred to the book by J. E. Gentle [3], where special attention is given to the assessment of the quality of random numbers generators. This assessment can be made necessary since, for example, if the shape of a histogram resembles the probability distribution of a desired random variable, it is not guaranteed that all statistical properties of the random variable are being reproduced by the generated random numbers.

1.13.1 Congruential Methods

In a congruential method, each generated number is determined by its predecessor by means of a function (not necessarily linear) followed by a modular operation. For example, in the *multiplicative congruential method* we have

$$Z_k = \alpha Z_{k-1} \bmod M, \quad Z_k = 1, 2, \dots, M - 1, \quad (1.154)$$

where α is called the multiplier, M is the modulus and $Z_0 \neq 0$ is the *seed* of the generator. The congruence function ($x \bmod y$) returns the remainder on dividing x by y .

The congruential method is able to generate a uniformly-distributed sequence of pseudo-random numbers between 1 and $M - 1$. We say that the sequence will be pseudo-random since it will repeat itself after its period. This period depends on the choice of α and M and its maximum value is $M - 1$.

If M is very large and we divide the resultant numbers by $M - 1$, we will have the generation of a uniform random variable in the interval $(0, 1]$. The greater the value of M , the closer to zero the smallest number will be.

A commonly used modulus is the *Mersenne prime* $2^{31} - 1$, and a commonly used multiplier is 7^5 [3, p. 13]. The resultant sequence will have a period of 2,147,483,646 numbers, a value that is sufficiently large for most applications. In other words, the sequence will not repeat itself during most of the simulations of practical interest. With such a large value of M , if we divide the generated numbers by $M - 1$, the smallest resultant number will be indeed very close to 0.

The importance of generating random numbers with uniform distribution is that random numbers having other distributions are usually generated from uniformly-distributed numbers, as we shall see in the sequel.

1.13.2 Inverse CDF Method

The inverse cumulative distribution function (CDF) method is one of the simplest manners for generating continuous or discrete random variables. It is better explained through an illustration, as shown in Fig. 1.11. Uniformly-distributed random numbers U in the range $(0, 1]$ are generated and the random numbers X with the desired distribution are obtained from the inverse of its cumulative distribution function for a given value of U .

For discrete random variables, the method is even more straightforward. The probabilities of the values of X determine comparison thresholds and a given value of X is generated if U is within a specific range of values. For example, the value 3 for the random variable X in Fig. 1.11(b) is generated whenever a value of U is generated in the range $[0.4, 0.7)$. Note that a value of U in this range will happen with probability 0.3, which will correspond to the probability $P[X = 3]$.

Other continuous distributions with CDF $F_Y(y)$ can be used instead of the uniform, though uniform random numbers are used in the majority of cases.

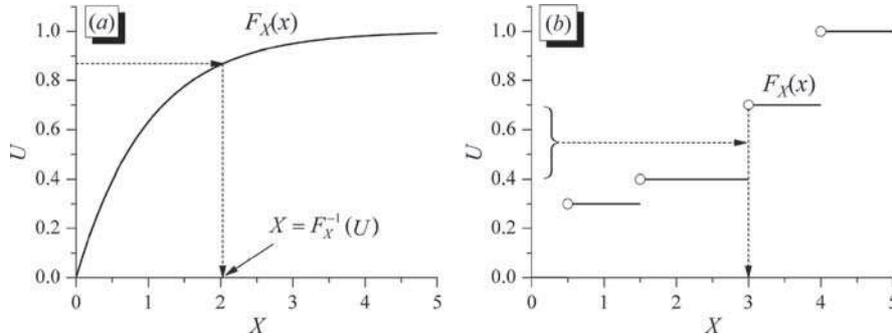


Fig. 1.11 The inverse CDF method for continuous (a), and discrete (b) random variables

We can generalize the inverse CDF method and write that

$$X = F_X^{-1}[F_Y(Y)]. \quad (1.155)$$

The inverse CDF method is simple in all cases where it is simple to find the inverse CDF of the desired random variable. The method is summarized below:

1. Generate a random variable U having a uniform distribution in $(0, 1)$.
2. Make $X = F_X^{-1}(U)$, where $F_X^{-1}(U)$ is the value of x when $F_X(x) = U$.

1.13.3 Acceptance/Rejection Methods

An acceptance/rejection method generates a random variable X with probability density $f_X(x)$ based on the realizations of another random variable Y with PDF $f_Y(y)$, preferably similar to $f_X(x)$.

The following algorithm corresponds to one of the acceptance/rejection methods, which is used to operate from uniformly-distributed random numbers (Fig. 1.12 illustrates examples of the PDFs involved):

1. Choose a scaled PDF $cf_Y(y)$ in a way that the total area of the desired PDF $f_X(x)$ lies inside the area of $cf_Y(y)$.
2. Generate a random number Y with density $f_Y(y)$.
3. Generate a uniformly-distributed random number U in the interval $(0, cf_Y(Y))$.
4. If $U \leq f_X(Y)$, accept Y as a valid realization of X , that is $X \leftarrow Y$.
5. If $U > f_X(Y)$, reject Y and go to step 2.

The idea behind an acceptance/rejection method is that sub-samples properly selected from samples having a given distribution are stochastically equivalent to samples from a different distribution.

To reduce the number of computations in the acceptance/rejection method, one must ensure that the proportion of acceptances is high, which means that the area in-between the desired and the scaled PDFs must be minimized.

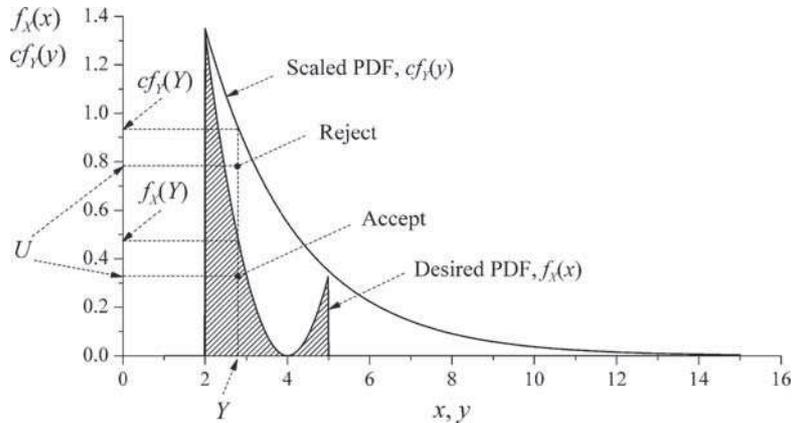


Fig. 1.12 Illustration of the acceptance/rejection method to convert a uniform random number

The acceptance/rejection method is very versatile and, unlike the inverse CDF method, it also applies to multivariate random variables.

1.13.4 Box-Muller Method

The Box-Muller method is used to generate Gaussian random variables X and Y from two independent and uniformly-distributed random numbers U_1 and U_2 , through the following relations:

$$\begin{aligned} 1. X &= \cos(2\pi U_1) \sqrt{-2\sigma_X^2 \ln(U_2)} + \mu_X \\ 2. Y &= \sin(2\pi U_1) \sqrt{-2\sigma_Y^2 \ln(U_2)} + \mu_Y, \end{aligned} \quad (1.156)$$

where U_1 and U_2 are in the interval $(0, 1]$, σ_X^2 and σ_Y^2 are the variances of the desired Gaussian random variables and μ_X and μ_Y are their expected values.

Simulation 1.6 – Random Numbers Generation



File – CD drive: \Simulations\Probability\Random_numbers.vsm

Default simulation settings: Frequency = 100 Hz; End = 1,000 seconds. Seed value in the uniform random variable (RV) generator: 5. Variance in the exponential RV generator: 2. Probability of success of the Bernoulli trials in the binomial RV generator: 0.5. Variance and mean in the Gaussian RV generator: 2 and 1, respectively.

This experiment is intended to show examples of random numbers generation using the methods described in this section.

A total of 100,000 numbers are generated for each random variable, using the default simulation “end” time. The histograms and estimates of the variance and mean are shown for each generated random variable.

In the sequel we present brief descriptions of the generators. Follow these descriptions while analyzing inside each generator through a double click over it.

Generator A implements the multiplicative congruential method to produce uniformly-distributed random variables in the interval $(0, 1]$. Specifically, it implements the expression

$$Z_k = (2^{31} - 2)^{-1}(7^5)Z_{k-1} \bmod(2^{31} - 1), \quad (1.157)$$

where the initial seed Z_{k-1} is configurable and $k = 1, 2, \dots, 100,000$.

Generator B implements the inverse CDF method to generate an exponential random variable with configurable variance $1/\lambda^2$. For this distribution we know that the CDF is $F_X(x) = 1 - e^{-\lambda x}$. Applying the inverse CDF method we have:

$$\begin{aligned} u &= 1 - e^{-\lambda x} \Rightarrow \ln(1 - u) = -\lambda x \\ \Rightarrow X &= -\frac{1}{\lambda} \ln(1 - U) \equiv -\frac{1}{\lambda} \ln(U) \end{aligned} \quad (1.158)$$

where U is a uniformly-distributed random variable in the interval $(0, 1]$.

Generator C also implements the inverse CDF method, but this time it produces a binomial random variable with parameters $n = 5$ and configurable p . The block diagram of this generator is shown in Fig. 1.13, where the uniform random number generation is realized by a built-in block of VisSim/Comm and the comparison thresholds were computed according to a binomial CDF, as follows:

$$T_x = \sum_{k=0}^4 \binom{5}{x} p^x (1-p)^{5-x}, \quad x = 0, 1, 2, 3, 4. \quad (1.159)$$

Generator D implements the acceptance/rejection method to produce random numbers with the “custom” desired distribution $f_X(x)$ illustrated in Fig. 1.12. This PDF and the scaled PDF $cf_Y(y)$, which is also shown in Fig. 1.12, are given by

$$\begin{aligned} f_X(x) &= \begin{cases} (4-x)^2/3, & 2 < x < 5 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \\ cf_Y(y) &= \begin{cases} c\lambda e^{-\lambda(y-2)}, & y \geq 2, c = 3, \lambda = 0.45 \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (1.160)$$

The algorithm for realizing the acceptance/rejection method was implemented via *Mathcad* and the numbers generated were exported to VisSim/Comm via a “.dat” file. In this algorithm, the uniform random variable U was generated by a built-in

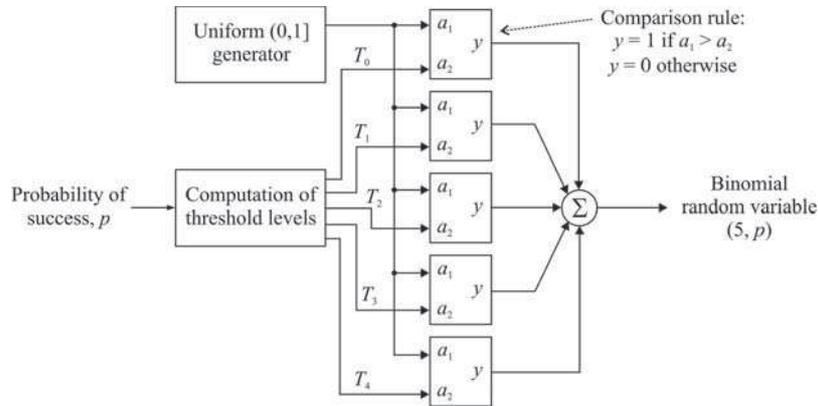


Fig. 1.13 Binomial random number generator: $n = 5$ and configurable p

function of *Mathcad* and the exponential random variable was generated via inverse CDF method, by applying the mapping equation:

$$X = -\frac{1}{\lambda} \ln(1 - U) + 2. \quad (1.161)$$

Finally, *Generator E* is implementing the Box-Muller method described by the equations in (1.156) to generate a Gaussian random variable with configurable variance σ_X^2 and mean μ_X . The uniform random variables U_1 and U_2 are generated by a built-in block of VisSim/Comm.

Now, run the simulation and compare the mean and the variance of each random variable with the theoretical results presented in Sects. 1.8 and 1.9. It is left as an exercise for the reader to show that the mean and the variance for the random variable X with “custom” distribution $f_X(x)$ are 2.75 and 0.638, respectively.

Open the histogram of each random variable and run the simulation. Observe the resultant shape of the histogram and compare it with the shape of the theoretical probability density or mass function.

Explore inside the individual blocks. Try to understand how they were implemented. Create and investigate for yourself new situations and configurations of the simulation parameters and try to reach your own conclusions.

1.14 Random Processes

When random variables manifest in time or space we say that we have a *random process*. In other words, when the realizations of the random phenomenon of interest can be represented through waveforms with the x -axis being time or space, we say that we have a random process.

Let us now develop a more formal definition of a random process by observing the illustration in Fig. 1.14. The random process is denoted by $X(t)$. It is composed by a set (not necessarily finite) of *sample-functions* $X(t, \xi_i)$, $i = 1, 2, \dots$ etc. These sample-functions can be thought as realizations of the random process $X(t)$ during some observation interval. This definition is analogous to the definition of a sample-point, which represents a realization of a random variable. As an example, think of $X(t, \xi_i)$ as the waveforms associated to the amount of rainfall measured by a standard rain gauge during Decembers of consecutive years.

If we collect samples of a random process at some instants, say, t_1 and t_2 , we shall obtain values that can be associated to the random variables $X(t_1)$ and $X(t_2)$, which are simply denoted by X_1 and X_2 . Notice that X_1 and X_2 are obtained *across* the random process, that is, across the sample-functions.

An important concept associated to a random process is that its statistics are obtained from the random variables derived from the sample-functions in the way just described. However, the precise estimation of a statistical parameter of a random process demands the observation of an infinite number of sample-functions, if this number is unlimited.

Depending on its nature, a random process can be classified in: (1) continuous in time and continuous in its values (or simply continuous-time, continuous-valued random process), (2) discrete-time, continuous-valued, (3) continuous-time, discrete-valued or (4) discrete-time, discrete-valued. Additionally, the sample space can be limited, if the number of sample-functions is finite, and unlimited, if the number of sample-functions is infinite.

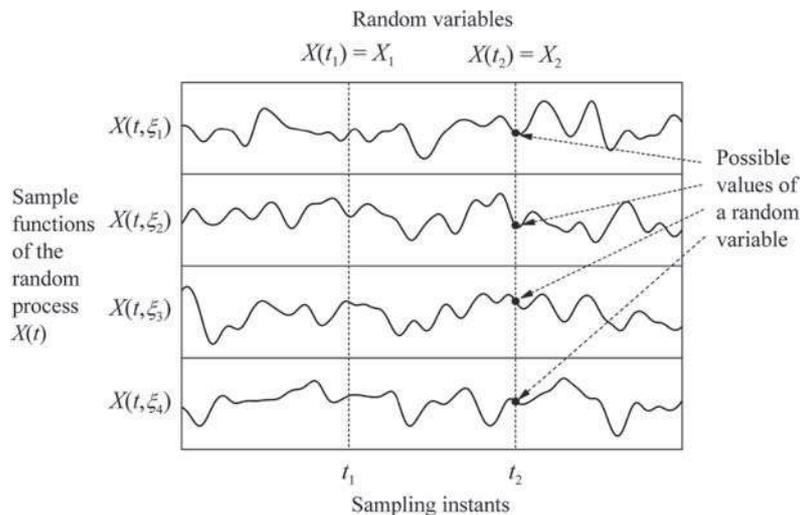


Fig. 1.14 The concept of a continuous-time, continuous-valued random process $X(t)$

A random process can also be classified as real or complex. Unless otherwise mentioned, in this chapter we consider only real random processes. For an applied and condensed treatment of complex random processes, see [13, p. 66].

Simulation 1.7 – The Concept of Random Process



File – CD drive:\Simulations\Probability\Random_process.vsm

Default simulation settings: Frequency = 100 Hz; End = 3 seconds. Auto Restart and Retain State enabled. Number of sample functions: 10. Export to file blocks disabled. Upper (continuous-time, continuous-valued) process selected.

This experiment aims at helping the reader to understand the concept of random processes. It also permits that samples of the analyzed random processes are exported to specific files. It can also be used to export data from other random processes created by the user in substitution of the ones already provided. This feature allows the reader to implement his own statistical data analysis from the files exported, using the concepts presented in this chapter and well-known mathematics or spreadsheet software tools.

Open the file indicated in the header and note that two continuous-time random processes are being generated. The upper one is also a continuous-valued process corresponding to a filtered noise signal. The lower process is a discrete-valued random sequence of equally-likely rectangular pulses of amplitude ± 1 and duration 1 second. The user can select the process to be analyzed by means of the “selector” block.

Although the total number of sample-functions generated for each process can be configured, a time plot shows the first 10 sample-functions of the selected random process. Each sample-function lasts for 3 seconds and the complete set is generated by an automatic restart of the simulation.

Each sample-function of the selected random process is sampled at the instants 1 and 1.4 seconds and the sample values are exported to the files “c:\X1.dat” and “c:\X2.dat”. When the exporting option is enabled in the corresponding block, each time you run the simulation n samples are augmented to the samples previously stored in the file, where n is the number of sample-functions generated. If the exporting options are not enabled, the files “c:\X1.dat” and “c:\X2.dat” are created anyway, but no data are stored on them.

Note that the sampling process described above is being realized *across* the random process, according to the theoretical concepts. These samples can be used to estimate the statistics of a random process through *ensemble averages*.

A third sampling process is being realized, but *along* one of the sample-functions. The resultant samples can also be used to estimate the statistics of the random processes, but now by means of *time averages*. In this case, samples are taken in 0.1 second intervals during 3 seconds and the resultant values are exported to the file “c:\X3.dat”. Whenever you run the simulation, a new set of $30n$ values are augmented to the previously stored ones. By using the default settings you will have 300 samples per run.

Shortcuts are also given in the workplace to speed-up the access to the exported files, so that you can browse, copy, delete or modify the stored data as desired.

Using the default simulation settings, open the time plot of the sample-functions and run the simulation. Observe the similarity of the sample-functions displayed with those shown in Fig. 1.14. Rerun the simulation and observe that you will be presented a new set of sample-functions, as expected from the definition of a random process. If you prefer, clear the overplot to view a new set without superimposing the previous ones. Select the other process and repeat the instructions in this paragraph. When finished, reset the simulation parameters to the default configuration (if you prefer, close the file without saving and reopen it).

Now enable the file exporting process in each export block and run the simulation. Through the shortcuts given, open each of the files just to get familiar with the data displacement. Run the simulation again and notice that the number of samples has been doubled.

This simulation will be addressed again later on in this chapter, so you can implement statistical analysis of the random processes on your own.

Explore inside the individual blocks. Try to understand how they were implemented. Create and investigate for yourself new situations and configurations of the simulation parameters and try to reach your own conclusions.

1.14.1 Stationarity Classes for Random Processes

Random processes can be classified as *strict-sense stationary*, *wide-sense stationary* or *ergodic*. A random process is said to be strict-sense stationary if the joint distribution of any set of samples does not depend on the time origin, that is,

$$f_{X(t_1), X(t_2), \dots, X(t_n)}(x_1, x_2, \dots, x_n) = f_{X(t_1+\tau), X(t_2+\tau), \dots, X(t_n+\tau)}(x_1, x_2, \dots, x_n). \quad (1.162)$$

In other words, all the statistical parameters of a strict-sense stationary random process do not vary with any shift on the time axis.

A wide-sense random process has a constant mean and an autocovariance (or autocorrelation) that is dependent only on the time difference $t_2 - t_1$, not on the absolute values of t_1 and t_2 . In mathematical terms,

$$\begin{aligned} \mu_X(t) &= \mu_X, \text{ for all } t, \text{ and} \\ \text{cov}[X(t_1), X(t_2)] &= \text{cov}[X(t_2 - t_1)] = K_X(\tau), \text{ for all } t_1 \text{ and } t_2. \end{aligned} \quad (1.163)$$

Note that strict-sense stationarity is more restrictive than wide-sense. In what follows, unless otherwise mentioned, we shall refer to a wide-sense stationary random process simply as a *stationary process*. This class of processes is of major importance for the study of communication systems theory, since most of the processes encountered in these systems can be regarded as stationary, at least for a sufficient long period of time. Unless otherwise mentioned, we shall consider only stationary random processes for the rest of the presentation of this chapter.

Another important classifications correspond to the *mean-ergodic* and the *autocovariance-ergodic* random processes, which correspond to sub-classes of wide-sense random processes. For these random processes, *time averages can substitute ensemble averages*, since a single realization of an ergodic process is statistically representative of the whole process. Putting in a different way, the mean and the correlation (or covariance) function of an ergodic random process can be obtained from the temporal analysis of a single sample-function, instead of the analysis across the sample-functions. Specifically,

$$\begin{aligned}\hat{\mu}_X &= \frac{1}{2T} \int_{-T}^T x(t) dt \quad \text{and} \\ \hat{K}_X(\tau) &= \frac{1}{2T} \int_{-T}^T [x(t+\tau) - \mu_X][x(t) - \mu_X] dt\end{aligned}\tag{1.164}$$

converge, in a mean square sense, to the mean μ_X and to the autocovariance $K_X(\tau)$ as $T \rightarrow \infty$, respectively. Here $x(t)$ is a realization or sample-function of $X(t)$.

Many random processes encountered in communications systems analysis can be considered ergodic, which facilitates enormously the estimation of the mean and of the correlation or covariance functions, as we shall see in several instances throughout the rest of this chapter.

One last classification refers to *cyclostationary* or *periodically stationary* random processes [9, p. 525; 10, p. 373; 13, p. 75; 15, p. 458]. These are processes in which statistical averages are periodic. These processes also arise in many situations related to communications systems analysis, for example when dealing with linear-modulated signals.

1.14.2 Averages of Stationary Random Processes

Similar to the definition of moments of random variables, the n -th moment of a random process $X(t)$ is defined by

$$E[X^n(t)] = \int_{-\infty}^{\infty} x^n f_{X(t)}(x) dx.\tag{1.165}$$

In this section we study the main moments and other statistical averages of stationary random processes, with emphasis on those averages that involves one or two processes. The analysis for multiple random processes follows by analogy to multiple random variables and will not be presented here.

1.14.2.1 Mean and Variance

The mean and the variance of a stationary random process $X(t)$ sampled at any instant t are given respectively by

$$\begin{aligned}\mu_X &= E[X(t)] = \int_{-\infty}^{\infty} x f_{X(t)}(x) dx \text{ and} \\ \sigma_X^2 &= E[(X(t) - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_{X(t)}(x) dx.\end{aligned}\tag{1.166}$$

From the law of large numbers, if we collect samples from a sufficiently large number of sample-functions, the mean and the variance can be estimated through the ensemble sample averages

$$\begin{aligned}\hat{\mu}_X &= \frac{1}{n} \sum_{j=1}^n X(t, \xi_j) \text{ and} \\ \hat{\sigma}_X^2 &= \frac{1}{n-1} \sum_{j=1}^n [X(t, \xi_j) - \hat{\mu}_X]^2.\end{aligned}\tag{1.167}$$

If the process is ergodic, the first expression in (1.164) can be applied to the computation of the mean. Moreover, if the process is band-limited [13, p. 72] such that it has no or almost zero frequency content for $|f| > W$, the following discrete-time averages can be applied:

$$\begin{aligned}\hat{\mu}_X &= \frac{1}{n} \sum_{j=1}^n x(t_j) \text{ and} \\ \sigma_X^2 &= \frac{1}{n-1} \sum_{j=1}^n [x(t_j) - \hat{\mu}_X]^2,\end{aligned}\tag{1.168}$$

where the values of $x(t_j)$ are obtained by sampling a single sample-function of the random process $X(t)$ at a minimum rate of $2W$ samples per second.

1.14.2.2 Autocovariance and Autocorrelation Functions

The autocovariance and the autocorrelation functions of a random process are defined respectively by

$$\begin{aligned}\text{cov}[X_1, X_2] &= E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_{X_1})(x_2 - \mu_{X_2}) f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \text{ and} \\ R_X(t_1, t_2) &= E[X_1 X_2] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2,\end{aligned}\tag{1.169}$$

where we have used the shorthand notation $X_1 = X(t_1)$ and $X_2 = X(t_2)$.

Now we consider only the autocorrelation function for stationary random processes, since the autocovariance can be easily obtained by considering the adequate subtraction of the mean of the random process, according to (1.169).

The autocorrelation of a stationary random process can be estimated through the ensemble sample average

$$\hat{R}_X(t_1, t_2) = \hat{R}_X(t_2 - t_1) = \hat{R}_X(\tau) = \frac{1}{n} \sum_{j=1}^n X(t, \xi_j)X(t + \tau, \xi_j). \quad (1.170)$$

If the process is ergodic, we have

$$\hat{R}_X(\tau) = \frac{1}{2T} \int_{-T}^T x(t + \tau)x(t)dt. \quad (1.171)$$

If the process is ergodic and band-limited, the following discrete-time average can be applied:

$$\hat{R}_X(\tau) = \frac{1}{n} \sum_{j=1}^n x(t_j)x(t_j + \tau). \quad (1.172)$$

The autocorrelation function has some important properties, as shown by:

1. $R_X(\tau) = R_X(-\tau)$
 2. $|R_X(\tau)| \leq R_X(0)$
 3. $E[X^2(t)] = R_X(0)$.
- (1.173)

The autocovariance and the autocorrelation functions of a stationary random process follow the relation

$$\text{cov}[X(t), X(t + \tau)] = R_X(\tau) - \mu_X^2. \quad (1.174)$$

If $\text{cov}[X(t), X(t + \tau)] = 0$, the random variables obtained by sampling the random process at instants t and $t + \tau$ are said to be *uncorrelated*. If $R_X(\tau) = 0$, the random variables obtained by sampling the random process at instants t and $t + \tau$ are said to be *orthogonal*.

Example 1.8 – In Fig. 1.15 it is plotted a bipolar binary random waveform corresponding to a sample-function $x(t)$ of a stationary stochastic process $X(t)$. The pulses of duration T are assumed to be equally-likely and the initial instant of an entire pulse can start after t_d seconds, where t_d is a uniformly-distributed random variable in $(0, T]$. The objective of this example is to find $R_X(\tau)$, the autocorrelation function of the random process $X(t)$.

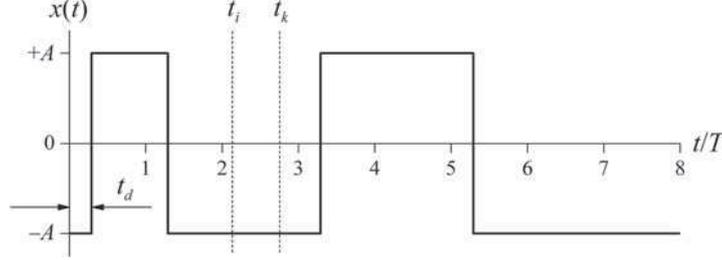


Fig. 1.15 Sample-function of a bipolar binary random wave

Since the pulses are equally-likely, $E[X(t)] = 0$. Moreover, the polarity of a given pulse is independent of the polarity of the previous one.

In Fig. 1.15 two sampling instants are indicated, t_i and t_k . First let $|t_k - t_i| \geq T$. In this case $X(t_k)$ and $X(t_i)$ occur in different pulses, so that

$$E[X(t_k)X(t_i)] = E[X(t_k)]E[X(t_i)] = 0, \quad |t_k - t_i| \geq T. \quad (1.175)$$

For $|t_k - t_i| < T$, with $t_i < t_k$, $X(t_k)$ and $X(t_i)$ will occur in the same pulse if $t_d < (T - |t_k - t_i|)$. In this case the correlation between $X(t_k)$ and $X(t_i)$ will be conditioned to the value of t_d . Then we have

$$E[X(t_k)X(t_i)|t_d] = \begin{cases} A^2, & t_d < T - |t_k - t_i| \\ 0, & \text{otherwise.} \end{cases} \quad (1.176)$$

To eliminate the conditioning in (1.176) we resort to the *total expectance* law:

$$E[E[X|Y]] = E[X]. \quad (1.177)$$

Then, applying (1.177) in (1.176) we obtain

$$\begin{aligned} E[X(t_k)X(t_i)] &= \int_0^{T-|t_k-t_i|} A^2 f_{T_d}(t_d) dt_d \\ &= \int_0^{T-|t_k-t_i|} \frac{A^2}{T} dt_d = A^2 \left(1 - \frac{|t_k - t_i|}{T}\right). \end{aligned} \quad (1.178)$$

Letting $\tau = t_k - t_i$ in (1.178), we finally obtain the autocorrelation function of the bipolar random binary sequence:

$$R_X(\tau) = \begin{cases} A^2 \left(1 - \frac{|\tau|}{T}\right), & |\tau| < T \\ 0, & \text{otherwise.} \end{cases} \quad (1.179)$$

This autocorrelation function is plotted in Fig. 1.16.

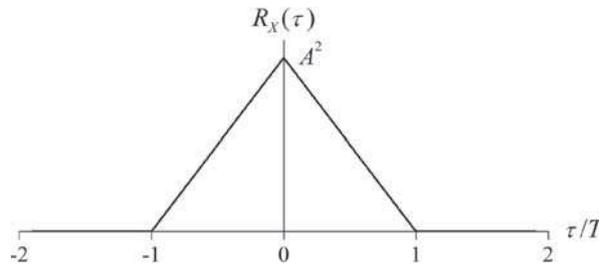


Fig. 1.16 Autocorrelation function of a bipolar binary random wave

1.14.2.3 Power Spectral Density (PSD)

One of the most important properties of the autocorrelation function states that $S_X(f)$, the *power spectral density* (PSD) of the random process $X(t)$, is determined by the Fourier transform of the autocorrelation function $R_X(\tau)$. This transform and the corresponding inverse are given respectively by

$$\begin{aligned} S_X(f) &= \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau \text{ and} \\ R_X(\tau) &= \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau} df. \end{aligned} \quad (1.180)$$

The relations in (1.180) are known as the *Wiener-Khintchine* relations.

From the point of view of the study of communications systems, the power spectral density (PSD) is an important characteristic of a voltage or current random process, since it reveals the frequency content of the process. In other words, it measures the distribution of the power of a random signal in the frequency domain and for this reason it is measured in watts/hertz.

A fast-varying random process has an autocorrelation function more concentrated around $\tau = 0$, yielding a wide PSD. A slowly-varying random process has an $R_X(\tau)$ decaying slowly with τ and, as a consequence, it has a narrowband PSD.

The PSD has some important properties:

$$\begin{aligned} 1. \int_{-\infty}^{\infty} S_X(f) df &= R_X(0) = E[X^2(t)] \\ 2. S_X(f) &= S_X(-f). \end{aligned} \quad (1.181)$$

That is, the total average power of a voltage or current random process developed over a 1Ω resistance, which corresponds to the second moment of the process, can

be calculated as the area under the curve of the PSD. Moreover, from (1.173), the PSD is an even function of the frequency.

There are many other concepts associated to the power spectral density. By now it suffices to understand those presented above. Further analysis on the PSD will be presented in this chapter in the context of the transmission of random processes through linear systems. The power spectral density will also be considered with more details in Chap. 2, Sect. 2.6.

Example 1.9 – Now let us find the power spectral density of the stationary binary random wave considered in Example 1.8. From Chap. 2, Table 2.4, the Fourier transform of a triangular function is

$$\mathfrak{F}\left[\begin{cases} 1 - \frac{|t|}{T}, & |t| < T \\ 0, & |t| \geq T \end{cases}\right] = T \frac{\sin^2(\pi f T)}{(\pi f T)^2}. \quad (1.182)$$

Then, the power spectral density of the binary random wave will be given by

$$S_X(f) = \mathfrak{F}[R_X(\tau)] = A^2 T \frac{\sin^2(\pi f T)}{(\pi f T)^2} = A^2 T \operatorname{sinc}^2(f T). \quad (1.183)$$

This PSD is plotted in Fig. 1.17 for $A = 1$ and $T = 0.1$, using two different scales: (a) conventional linear scale, in W/Hz, and (b) logarithmic scale, in dBm/Hz. The PSD in dBm/Hz is obtained by means of

$$S_X(f)_{\text{dBm/Hz}} = 10 \log \left(\frac{S_X(f)_{\text{W/Hz}}}{0.001} \right). \quad (1.184)$$

The use of the scale in dBm/Hz was motivated by the fact that this is a very common scale encountered in communication systems analysis. Moreover, it is typically used in a real spectrum analyzer. When expressed in this way, lower values of the PSD are emphasized to permit measurements. In a linear scale, low-level lobes would not be even noticed, as can be readily verified in Fig. 1.17(a).

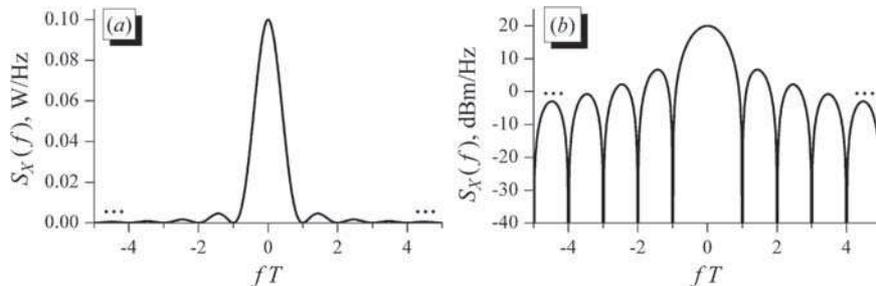


Fig. 1.17 Power spectral density of a bipolar binary random wave in two different scales

From (1.173), $E[X^2(t)] = R_X(0)$. Since $E[X(t)] = 0$, the second moment of the process also corresponds to its average AC power. Observing Fig. 1.16 we see that $R_X(0) = A^2$. Since $A = 1$ in this example, $E[X^2(t)] = 1$ watt.

From the PSD in Fig. 1.17(a), it is apparent that most of the signal power is concentrated in the *main lobe* of the frequency spectrum. Making a coarse approximation of this lobe as a triangle, its area would be $(2/T) \times (A^2T)/2 = A^2 = 1$. This result shows us that, in fact, the average power of a random process can be determined by the area under its power spectral density.

1.14.2.4 Cross-Covariance and Cross-Correlation Functions

The cross-covariance and the cross-correlation functions follow the same reasoning adopted for the autocovariance and the autocorrelation functions, the only difference being the fact that, now, we are dealing with two different random processes $X(t)$ and $Y(t)$, instead of one. The expressions must then reflect this.

The cross-covariance and the cross-correlation functions of two stationary random processes follow the relation

$$\text{cov}[X(t), Y(t + \tau)] = R_{XY}(\tau) - \mu_X^2 \mu_Y^2. \quad (1.185)$$

If $\text{cov}[X(t), Y(t + \tau)] = 0$, the random processes are said *uncorrelated*. If the cross-correlation $R_{XY}(\tau) = 0$, the random processes are said *orthogonal*. Note that if two random processes are orthogonal, they will be uncorrelated if and only if one of their means or both are zero.

The cross-correlation function for ergodic processes can be estimated by

$$\hat{R}_{XY}(\tau) = \frac{1}{2T} \int_{-T}^T x(t + \tau)y(t)dt. \quad (1.186)$$

If the processes are band-limited such that they have no or almost zero frequency content for $|f| > W$, the following discrete-time average can be applied:

$$\hat{R}_{XY}(\tau) = \frac{1}{n} \sum_{i=1}^n x(t_i)y(t_i + \tau), \quad (1.187)$$

where $x(t_j)$ and $y(t_j + \tau)$ are obtained by sampling a single sample-function of the random processes $X(t)$ and $Y(t + \tau)$ at a rate $2W$ samples per second or greater.

It is worth mentioning that if the random processes are cyclostationary, all the statistics presented above can be computed by considering a time interval of just one period of the processes.

Simulation 1.8 – Estimating Averages of Random Processes



File – CD drive:\Simulations\Probability\Random_process.vsm

Default simulation settings: Frequency = 100 Hz; End = 3 seconds. Auto Restart and Retain State enabled. Number of sample functions: 10. Export to file blocks disabled. Upper (continuous-time, continuous-valued) process selected.

In this experiment we revisit the diagram considered in Simulation 1.7, now with the aim of obtaining some statistical averages of the simulated random processes.

Open the simulation file indicated in the header and set the number of sample-functions to 9,000. Enable the export blocks for “X1.dat” and “X2.dat” and run the simulation. Wait for a while until the end of the simulation. Disable the export blocks for “X1.dat” and “X2.dat” and enable the export block for “X3.dat”. Change the number of sample-functions to 300 and run the simulation. Now you have three files, each having 9,000 samples: “X1.dat” has samples taken across the process $X(t)$ when sampled at $t = 1$ s, “X2.dat” has samples taken across the process $X(t)$ sampled at $t = 1.4$ s and “X3.dat” has samples taken along one sample function of the process $X(t)$.

As an exercise, load the above samples into a mathematical software tool, like *Mathcad*, *Matlab* and *Mathematica* or into a spread-sheet like *Excel*. Implement the sample averages that follow and try to obtain similar results.

(a) Ensemble mean of $X(t)$:

$$\begin{aligned}\hat{\mu}_X &= \frac{1}{n} \sum_{j=1}^n X(t, \xi_j) \\ &= \frac{1}{n} \sum_{j=1}^n X(1, \xi_j) = \frac{1}{9000} \sum_{j=1}^{9000} X1_j \cong 0.0103 \\ &= \frac{1}{n} \sum_{j=1}^n X(1.4, \xi_j) = \frac{1}{9000} \sum_{j=1}^{9000} X2_j \cong 0.01.\end{aligned}\quad (1.188)$$

Note that the estimated means are essentially the same, which is a consequence of the stationarity of the random process. Additionally, they are close to zero, the actual mean value of the continuous-time, continuous-valued process.

(b) Time average mean of $X(t)$:

$$\hat{\mu}_X = \frac{1}{n} \sum_{j=1}^n x(t_j) \quad (1.189)$$

$$= \frac{1}{9000} \sum_{j=1}^{9000} X3_j \cong -0.004.$$

Note that the mean obtained from the samples of a single sample-function of $X(t)$ is also close to the actual value, indicating that such random process is ergodic.

(c) Autocorrelation of $X(t)$ for $\tau = 0.4$ s:

$$\begin{aligned} \hat{R}_X(\tau) &= \frac{1}{n} \sum_{j=1}^n X(t, \xi_j)X(t + \tau, \xi_j) \\ &= \frac{1}{9000} \sum_{j=1}^{9000} X(1, \xi_j)X(1 + 0.4, \xi_j) \\ &= \frac{1}{9000} \sum_{j=1}^{9000} X1_j X2_j \cong -6.6 \times 10^{-3}. \end{aligned} \quad (1.190)$$

This value indicates that the random variables $X(t_1) = X_1$ and $X(t_2) = X_2$ have a low correlation for the value of $\tau = t_2 - t_1 = 0.4$ seconds. For $\tau = t_2 - t_1 = 0$ we shall find:

(d) Autocorrelation of $X(t)$ for $\tau = 0$:

$$\begin{aligned} \hat{R}_X(0) &= \frac{1}{9000} \sum_{j=1}^{9000} X(1, \xi_j)X(1, \xi_j) \\ &= \frac{1}{9000} \sum_{j=1}^{9000} X(1.4, \xi_j)X(1.4, \xi_j) \\ &= \frac{1}{9000} \sum_{j=1}^{9000} X1_j^2 \cong \frac{1}{9000} \sum_{j=1}^{9000} X2_j^2 \cong 1.09. \end{aligned} \quad (1.191)$$

This is indeed the approximate total average power of the continuous-time, continuous-valued random process.

(e) Time average variance of $X(t)$:

$$\begin{aligned} \hat{\sigma}_X^2 &= \frac{1}{n-1} \sum_{j=1}^n [x(t_j) - \hat{\mu}_X]^2 \\ &= \frac{1}{8999} \sum_{j=1}^{9000} [X3_j + 0.004]^2 \cong 1.04. \end{aligned} \quad (1.192)$$

Note that this value makes sense, since for a zero-mean voltage random waveform, the DC power is zero and the total average power is equal to the AC power.

Finally, let us illustrate how a time average autocorrelation could be performed from “X3.dat”. We need to multiply, sample-by-sample, $x(t_j)$ and $x(t_j + \tau)$, sum-up the result and divide per n , according to (1.172). To do this for $\tau = 0.4$ seconds we must calculate:

(f) Time average autocorrelation of $X(t)$:

$$\begin{aligned}\hat{R}_X(\tau) &= \frac{1}{n} \sum_{j=1}^n x(t_j)x(t_j + \tau) \\ &= \frac{1}{9000} \sum_{j=1}^{9000} X3_j X3_{(j+\tau/0.1)\bmod 9000} \cong -3.3 \times 10^{-3}.\end{aligned}\tag{1.193}$$

where $(j + \tau/0.1)\bmod(9000)$ produces a *circular-shift* of $\tau/0.1 = 0.4/0.1 = 4$ samples in the sequence of samples. Since the sample space is 0.1 seconds, 4 samples correspond to 0.4 seconds, which is the value desired for τ . In fact, this is an adequate method for estimating time averages for cyclostationary random processes. However, since the process under consideration is stationary, cutting-out the desired number of samples and putting them at the beginning of the sequence will produce the desired time-shift effect without producing significant estimation errors for $\tau \ll n$.

For the sake of brevity we shall stop the calculations here. As an exercise, change the value of one of the sampling instants, simulating another value for τ , and repeat the estimations above. Make other estimations from the files you have.

Recall that the theoretical autocorrelation function of a random process corresponding to a random sequence of bipolar pulses is a triangular function of τ . Try to produce small changes in the separation between the sampling instants in the simulation and estimate the autocorrelation. Likewise the theoretical result, you will observe a decay in the correlation with an increase of $|\tau|$, starting from $\tau = 0$.

Explore inside the individual blocks. Try to understand how they were implemented. Create and investigate for yourself new situations and configurations of the simulation parameters and try to reach your own conclusions.

1.14.3 Random Processes Through Linear Systems

While studying communication signals and systems, it is very common to find examples where random signals suffer some sort of filtering effect when going through linear systems. Typical exemplifying situations in which this occurs are the filtering of a modulated signal before transmission and the filtering of the received signal before it is further processed by the receiver.

In this section we analyze the behavior of random processes through linear systems, based on the model depicted in Fig. 1.18. A random process $X(t)$ is applied to the input of a linear system having impulse response $h(t)$ and frequency response $H(f)$. The output of the system is the random process $Y(t)$. In Fig. 1.18, sample-functions of $X(t)$ and $Y(t)$, corresponding to the waveforms $x(t)$ and $y(t)$, are shown for illustration purposes. Note that $y(t)$ was smoothed by the effect of filtering as compared to $x(t)$. In other words, the output process typically will have slower variations than the input process. This means that the autocorrelation function of the output process is wider relative to the autocorrelation function of the input process. As a consequence, the frequency content of $Y(t)$ will occupy a lower bandwidth than the frequency content of $X(t)$. These comments were inserted here only to illustrate a typical filtering effect, since there are also filters designed to produce other results different from a bandwidth reduction. For example, a filter can be used to change the phase of the signal, keeping the bandwidth approximately unchanged.

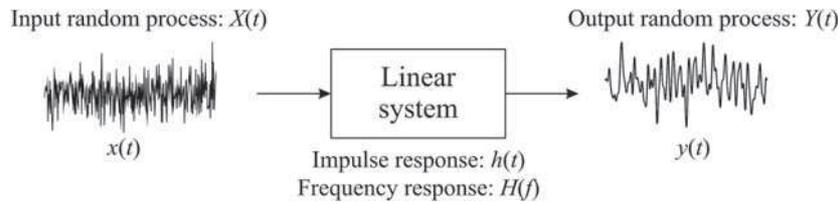


Fig. 1.18 A random process through a linear system

We are now able to start the discussion on the statistics of the random processes that are influenced by a linear system. Again we shall consider only real and stationary random processes, unless otherwise mentioned.

1.14.3.1 Mean

The relation between the mean of the output random process, μ_Y , and the mean of the input random process, μ_X , is given by

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t) dt = \mu_X \int_{-\infty}^{\infty} h(t) e^{-j2\pi f t} dt \Big|_{f=0} = \mu_X H(0). \quad (1.194)$$

1.14.3.2 Autocorrelation

The autocorrelation of the output random process, $R_Y(\tau)$, can be determined by knowing the system's impulse response $h(t)$ and the autocorrelation of the input random process, $R_X(\tau)$:

$$R_Y(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(u) h(v) R_X(\tau - v + u) du dv. \quad (1.195)$$

1.14.3.3 Power Spectral Density (PSD)

We know from (1.180) that the power spectral density $S_X(f)$ and the autocorrelation function $R_X(\tau)$ of a random process $X(t)$ are related through a Fourier transform pair. Additionally, we have seen that the PSD exhibit some important properties, as given by (1.181). Another very important result relates the PSD at the input and the output of a linear system, as shown by:

$$S_Y(f) = S_X(f)|H(f)|^2, \quad (1.196)$$

where $|H(f)|^2 = H(f)H^*(f)$ is the magnitude squared of the frequency response $H(f)$.

1.14.3.4 Cross-Spectral Density

The cross-spectral density of two random processes is defined as the Fourier transform of their cross-correlation function, that is,

$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j2\pi f\tau} d\tau. \quad (1.197)$$

Differently from the power spectral density, the cross-spectral density seems do not exhibit a clear physical meaning. However, an interpretation can be made with the help of the following example.

Example 1.10 – Suppose that the processes $X(t)$ and $Y(t)$ have zero mean and are jointly stationary. Let us find the PSD of $Z(t) = X(t) + Y(t)$, as given below:

$$\begin{aligned} R_Z(\tau) &= E[Z(t)Z(t + \tau)] \\ &= E\{[X(t) + Y(t)][X(t + \tau) + Y(t + \tau)]\} \\ &= E[X(t)X(t + \tau)] + E[X(t)Y(t + \tau)] \\ &\quad + E[Y(t)X(t + \tau)] + E[Y(t)Y(t + \tau)] \\ &= R_X(\tau) + R_{XY}(\tau) + R_{YX}(\tau) + R_Y(\tau). \end{aligned} \quad (1.198)$$

Taking the Fourier transform of both sides of (1.198) we obtain

$$S_Z(f) = S_X(f) + S_{XY}(f) + S_{YX}(f) + S_Y(f). \quad (1.199)$$

In (1.199) we observe that if the processes $X(t)$ and $Y(t)$ are orthogonal to each other, their cross-correlation will vanish, so that

$$S_Z(f) = S_X(f) + S_Y(f). \quad (1.200)$$

From (1.199) and (1.200) we can conclude that the cross-spectral densities $S_{XY}(f)$ and $S_{YX}(f)$ are components of $S_Z(f)$ created to account for the non-orthogonality between $X(t)$ and $Y(t)$.

We can generalize the above concept and state the following important result: *the power spectral density of the sum of orthogonal random processes is equal to the sum of the individual power spectral densities.*

The cross-spectral density has also applications in linear estimation, as discussed in, for example, [9, p. 614] and [18, p. 426]. Some of its important properties are summarized below:

1. $S_{XY}(f) = S_{YX}^*(f) \Rightarrow R_{XY}(\tau) = R_{YX}(-\tau)$
2. $S_{XY}(f) = H^*(f)S_X(f) \Rightarrow R_{XY}(\tau) = h^*(-\tau) * R_X(\tau)$
3. $S_{YX}(f) = H(f)S_X(f) \Rightarrow R_{YX}(\tau) = h(\tau) * R_X(\tau)$
4. $S_Y(f) = H^*(f)S_{YX}(f) \Rightarrow R_Y(\tau) = h^*(-\tau) * R_{YX}(\tau)$

(1.201)

As we have already mentioned, the concepts related to the power spectral density of random processes are of major importance for the study of communication systems. For this reason we shall return to this topic in a more specific context in Chap. 2, where we shall consider the estimation of the PSD of random signals typically encountered in the analysis of those systems.

1.14.4 The Gaussian Random Process

Likewise the Gaussian random variable, the Gaussian random process often arises in the study of digital transmission systems.

A Gaussian random process has many definitions, some of them general, others more specific and applied to particular situations. In a first and simplified definition, a stationary random process is said to be Gaussian if the probability density function of the random variable derived from samples of the process is Gaussian. It is worth mentioning that if the process is also ergodic, these samples can be collected from a single sample-function, which represents an easy way of verifying if a given process is Gaussian from the point of view of the first definition.

In a more formal definition, let Y be a random variable defined by the following *functional*⁵ relation with a random process $X(t)$:

$$Y = \int_0^T g(t)X(t)dt, \quad (1.202)$$

⁵ Simply stating, a *functional* is a function that takes a function as its argument and returns a scalar

where $g(t)$ is an arbitrary function and T is an arbitrary observation interval. We say that $X(t)$ is a Gaussian random process if Y is a Gaussian random variable.

The functional approach is of special interest for the present context. Recall from theory that the signal at the output of a linear system is determined by the convolution between the input signal and the impulse response of the system. In terms of random processes we can write

$$Y(t) = X(t) * h(t) = \int_{-\infty}^{\infty} X(u)h(t-u)du. \quad (1.203)$$

If, for simplicity purposes, we observe the value of $Y(t)$ at $t = 0$, we shall have a random variable Y . Then we can write

$$Y(0) = Y = \int_{-\infty}^{\infty} X(u)h(u)du. \quad (1.204)$$

But (1.204) has the form of (1.202). Then, if Y is Gaussian-distributed, $X(t)$ is a Gaussian random process. However, since samples of a Gaussian random process are Gaussian-distributed, we conclude that *if the input of a linear system is a Gaussian random process, the output random process is also Gaussian-distributed*. In other words, if a Gaussian random process is filtered, it produces another Gaussian random process as the result. This is an important concept that is frequently applied to the analysis of communication systems.

Simulation 1.9 – Filtered Complex Gaussian Process



File – CD drive:\Simulations\Probability\Gaussian_cplx.vsm

Default simulation settings: Frequency = 100 Hz; End = 200 seconds.
Cutoff frequency of the low-pass filters: 10 Hz (max).

This simulation implements the diagram shown in Fig. 1.19. Comparing this diagram with the Rayleigh random variable generator given in Simulation 1.3 (see file “Continuous_RV.vsm” if necessary), one can notice that they are very similar, except by the insertion of two low-pass filters (LPF) at the output of the Gaussian sources. Moreover, since here we are dealing with random processes, the nomenclature has changed: now we have a complex Gaussian random process $R(t) \exp[j\Theta(t)]$ having its real and imaginary parts composed by independent and identically distributed Gaussian random processes $X'(t)$ and $Y'(t)$, so that

$$\begin{aligned} R(t) &= \sqrt{[X'(t)]^2 + [Y'(t)]^2} \text{ and} \\ \Theta(t) &= \arctan[Y'(t)/X'(t)]. \end{aligned} \quad (1.205)$$

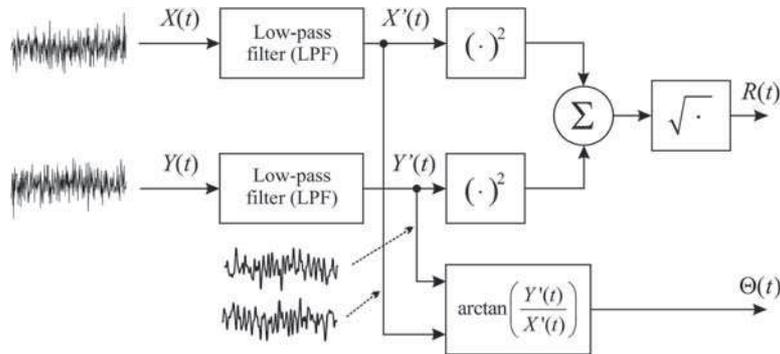


Fig. 1.19 Generation of a filtered, complex Gaussian process

Here, $X'(t)$ and $Y'(t)$ are the low-pass filtered versions of the input Gaussian processes $X(t)$ and $Y(t)$.

Making use of the results obtained in Simulation 1.3, here $R(t)$ is a Rayleigh-distributed random process and $\Theta(t)$ is a uniformly-distributed random process in $(-\pi, \pi)$. However, these processes can manifest themselves as slowly-varying or as fast-varying random processes according to the cutoff frequency used by the low-pass filters. Since a Gaussian input to a linear system will produce a Gaussian output, this is an approach often used to generate random processes having different degrees of correlation between successive samples. Obviously, this will work if such processes can be generated from Gaussian processes, because filtered Gaussian processes are also Gaussian.

Using the default simulation settings, open the upper “Time plot and histogram” block and run the simulation. You can verify that the magnitude of the complex Gaussian random process is indeed Rayleigh-distributed. In addition, you can see the time evolution of a sample-function of the Rayleigh random process and conclude that it refers to a fast-varying process.

Now, reduce the cutoff frequency of the LPF filters to the minimum (0.1 Hz). While looking at the time plot, run the simulation and verify that we have a slowly-varying Rayleigh random process. Note, however, that the histogram does not resemble a Rayleigh PDF anymore. This was expected, since in the observed interval the process does not exhibit enough realizations to be statistically characterized with accuracy. Increase the simulation time to 2,000 seconds and rerun the simulation while observing the histogram of the Rayleigh process. Note that the resemblance with the Rayleigh PDF has been restored.

Using the default simulation settings, now open the lower “Time plot and histogram” block and run the simulation. You can verify that the phase of the complex Gaussian random process is in fact uniformly-distributed. In addition, you can see the time evolution of a sample-function of the uniform random process and conclude that it is also a fast-varying process. Reduce the cutoff frequency of the LPF filters to the minimum value and repeat the procedure described in the last paragraph,

just to confirm that the uniformly-distributed random process also keeps its original distribution, no matter how fast it is.

In Chap. 3 we shall revisit the complex Gaussian random process as representative of the typical behavior of one of the most important channels used by modern communication systems: the wireless channel.

Explore inside the individual blocks. Try to understand how they were implemented. Create and investigate for yourself new situations and configurations of the simulation parameters, especially for the filter's cutoff frequencies, and try to reach your own conclusions.

1.14.5 Thermal Noise Process

In the context of communications systems, the term *noise* can be associated to any impairment, normally random, that affects system performance.

Among the different types of noise, *thermal noise* and *impulse* (or *impulsive*) *noise* deserve especial attention. Thermal noise, often called Gaussian noise or Johnson-Nyquist noise, is a random process generated by the agitation of the electrons inside an electrical conductor due to thermal influence. Impulsive noise is a normally high amplitude pulse noise caused by disturbances having abrupt changes and short durations. For example, switching processes in a circuit may cause impulsive noise. The ignition system of vehicles can also induce impulsive noise in an electronic circuit.

Thermal noise is present in all communication systems and, for this reason, we shall give some details about it in what follows. Additionally, since thermal noise is a Gaussian random process, from the central limit theorem and under certain circumstances the sum of noises coming from many different sources, no matter their individual distributions, can be approximated to a Gaussian noise.

It can be shown that the mean square value of the noise voltage V_{TN} across the terminals of the equivalent Thévenin circuit in Fig. 1.20 is given by [7, pp. 220–224; 5, p. 60]

$$E[V_{TN}^2] = 4kTBR \text{ volts}^2, \quad (1.206)$$

where $k = 1.38047 \times 10^{-27}$ joule/Kelvin (J/K) is the Boltzmann constant, T is the absolute temperature in Kelvin, B is the measuring bandwidth in hertz (Hz) and R is the resistance in ohms (Ω).

In the condition of maximum power transfer, the load connected to the terminals of the circuit in Fig. 1.20 must have R ohms. In this case, the average noise power delivered to this load will be

$$N = \frac{\left(\sqrt{E[V_{TN}^2]}/2\right)^2}{R} = \frac{\left(\sqrt{4kTBR}/2\right)^2}{R} = kTB \text{ watts}. \quad (1.207)$$

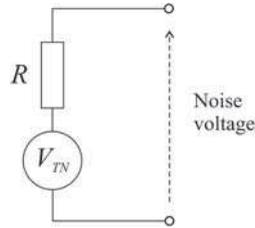


Fig. 1.20 Thévenin equivalent circuit for thermal noise generation

Since the number of electrons in a conductor is very large and their motions are random and independent, the central limit theorem leads us to the conclusion that the *thermal noise voltage is a zero-mean Gaussian random process*.

1.14.5.1 Additive White Gaussian Noise (AWGN)

In communication systems, thermal noise acts by addition to the signal of interest. In this sense we say that the noise is *additive*. Moreover, due to the variety of the electrons motion speeds, ranging from very slow to very high values, the frequency content of the thermal noise will be very high. This fact has led to the flat theoretical model for the power spectral density of this noise, that is,

$$S_W(f) = \frac{N_0}{2} \text{ W/Hz}, \quad -\infty \leq f \leq \infty, \quad (1.208)$$

where $N_0 = kT_e$ is the noise power spectral density produced at the input of a communication receiver circuit for which the *equivalent noise temperature* is T_e . The equivalent noise temperature is a circuit-dependent parameter defined as the temperature that a resistor must be submitted in order to produce the power density N_0 when connected to a noiseless version of the circuit.

Due to the flat model for the power spectral density of the thermal noise we say that it is a *white noise*. This is an allusion to the white color, since its spectrum contains the spectrum of all colors. Then, putting together all terms we usually call the thermal noise an *additive white Gaussian noise* (AWGN).

Since the autocorrelation function and the power spectral density form a Fourier transform pair, we have

$$R_W(\tau) = \mathfrak{F}^{-1}\{S_W(f)\} = \frac{N_0}{2}\delta(\tau), \quad (1.209)$$

where $\mathfrak{F}^{-1}\{\cdot\}$ is the inverse Fourier transform and $\delta(\tau)$ is the Dirac delta function.

The power spectral density (1.208) and the autocorrelation function (1.209) of the AWG noise are plotted in Fig. 1.21.

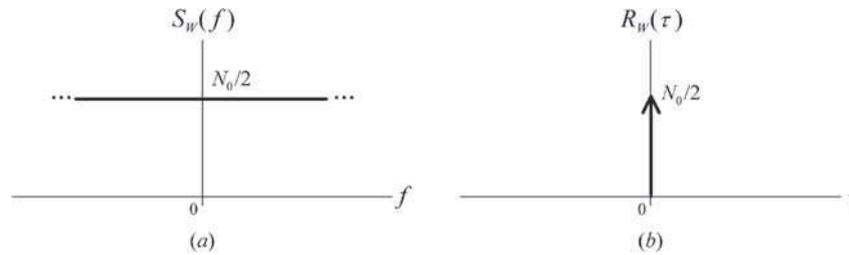


Fig. 1.21 Power spectral density (a) and autocorrelation function (b) of the AWGN

Note that the AWGN is the ultimate process in terms of randomness, since its samples collected Δ seconds apart will produce completely uncorrelated random variables for any value of $\Delta > 0$. Furthermore, note that the total average power of the AWGN is infinite, since it corresponds to the area under $S_W(f)$. This means that this model is unrealizable in practice. However, for most applications it suffices to consider that the thermal noise is white in the overall bandwidth of the system under consideration and zero outside. As a result, the total average noise power becomes finite.

1.14.5.2 Equivalent Noise Bandwidth

The flat noise bandwidth does not apply to all band-limited systems. In other words, even if we consider the noise flat within the system bandwidth, the system's response might not be flat in this range. To solve this problem it is usual to work with an *equivalent noise bandwidth*, defined as the bandwidth of an ideal filter that produces the same noise power produced by the filter that represents the actual system's response. The responses of these filters are illustrated in Fig. 1.22.

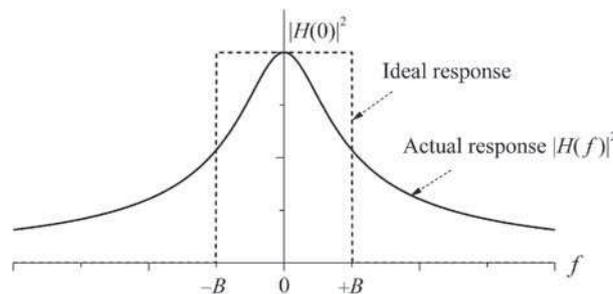


Fig. 1.22 Illustrating the concept of the equivalent noise bandwidth

Considering a white noise applied to the actual filter, the average noise power at the output of this filter will be

$$N_1 = \int_{-\infty}^{\infty} \frac{N_0}{2} |H(f)|^2 df = N_0 \int_0^{\infty} |H(f)|^2 df. \quad (1.210)$$

With the same white noise applied to the ideal filter of bandwidth B , the average noise power in its output will be

$$N_2 = \frac{N_0}{2} H^2(0) 2B. \quad (1.211)$$

Equating N_1 and N_2 given in (1.210) and (1.211), the system's equivalent noise bandwidth is finally given by

$$B = \frac{\int_0^{\infty} |H(f)|^2 df}{H^2(0)}. \quad (1.212)$$

1.15 Summary and Further Reading

In this chapter we studied some concepts about probability and random processes, with focus on those topics which are more directly applicable to the study of digital communication systems.

We dedicated about 75% of the chapter to random variables and about 25% to stochastic processes, since most of the concepts related to random variables are applicable with minor adequacies to random processes. In other words, if one has a good understanding about random variables, the concepts can be easily applied to random processes. For example, the statistics of a random process is in fact obtained from the statistics of the random variables derived from the process.

Many references were already cited throughout the chapter, according to each specific topic covered. In what follows we give a short overview about some of the references listed at the end of the chapter, aiming at helping the reader to find some complementary subject when needed.

The book by A. Papoulis [10] is one of the most complete references in what concerns the formal treatment of probability and random processes, without focusing in a specific application or area.

In the book by L. B. Koralov and Y. G. Sinai [6], basic material found in essentially all good books on probability and stochastic processes are addressed, but we think that the main contribution of the authors is to provide a complementary material not usually covered in detail in most of the similar books.

For a condensed overview of the main topics on probability and random processes applied to the design and analysis of communication systems we recommend Chaps. 10 and 11 of the book by B. P. Lathi [8], Chaps. 1 to 4 of the book by M. B. Pursley [14], Chap. of the book by J. G. Proakis [13] and Chap. 1 and Appendix 1 from the book by S. Haykin [5].

A fundamental analysis of random processes with applications in communications theory is given by B. P. Lathi in [7]. The book by P. Z. Peebles [11], though

applicable to engineering in general, focuses on some typical applications in communication systems analysis.

For a deep treatment of probability and stochastic processes applied to Electrical and Computer Engineering we recommend the books by A. Leon-Garcia [9], H. Stark and J. W. Woods [15] and R. D. Yates and D. J. Goodman [18].

1.16 Additional Problems

Several simulation-oriented problems were already proposed in the body of the simulation work plans throughout the chapter. Most of the following additional propositions are also based on simulation or are research-oriented, requiring some sort of additional reading. A number of exercises involving conventional calculations can be found in the references cited throughout the chapter.

1. Suppose that you want to create your own lottery, in which K integer numbers are drawn at random from a total of N integer numbers ranging from 1 to N . A player can make a bet with J numbers. In a first version of the lottery you have determined that $N = 6$, $K = 2$ and $J = 2$ or 3. Solve the following propositions: (a) Compute C , the number of combinations of 2 numbers in your lottery. List the possible combinations. (b) Determine P_2 , the probability of a player win by making a bet with 2 numbers. (c) Determine P_3 , the probability of a player win by making a bet with 3 numbers. (d) From the results obtained in “a” and “b”, determine an expression for P_J , the probability of a player win by making a bet with J numbers. (e) Now assume that an official lottery has 60 numbers and that draws of 6 numbers are made. Compute P_7 , the probability of a player win by making a bet with 7 numbers.
2. Implement a VisSim/Comm diagram to simulate a discrete memoryless source emitting symbols from the alphabet $\{s_0, s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$, with *a priori* probabilities $\{p_0, p_1, p_2, p_3, p_4, p_5, p_6, p_7\} = \{0.2, 0.28, 0.1, 0.09, 0.01, 0.02, 0.16, 0.14\}$, respectively. *Hint*: use the inverse CDF method for random number generation.
3. Let $\{X_i\}$, $i = 1, 2, \dots, n$, be a set of independent and identically distributed random variables with mean μ_X and variance σ_X^2 . Determine the variance of the random variable Z defined by:

$$Z = \frac{1}{\sqrt{n\sigma_X^2}} \left(\sum_{i=1}^n X_i - n\mu_X \right)$$

4. Suppose that, for reliability improvement purposes, a given machine has n on-board computers operating in parallel and on an independent fashion. The machine will continue to operate if at least one of its computers is operating properly. Considering that a computer fails with probability p , determine the probability that the machine will fail.

5. An electronic equipment has 326 components, each of them with a probability of failure of 0.001. Determine the probability of five or more components in failure, assuming that failures occur independently.
6. A die is thrown two times. Determine the probability of 3 dots in the second throw, given that 2 dots were presented in the first throw.
7. Using VisSim/Comm, implement an experiment capable of simulating the throwing of two dices and the computation of some unconditional and conditional probabilities via the relative frequency definition of probability.
8. Using the moment theorem, compute the mean, the mean square value and variance of the random variable whose PDF is given by:

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

9. The random variables X and Y have the joint PDF $f_{XY}(x, y)$ given below. *a)* determine the marginal PDFs of X and of Y . *b)* Based on the result in “*a*”, can we state that the random variables X and Y are independent? Justify your answer.

$$f_{XY}(x, y) = \begin{cases} \lambda\mu e^{-(\lambda x + \mu y)}, & x \geq 0, y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

10. Do some research and make a dissertation about two goodness-of-fit tests of your choice.
11. Let a white noise process $W(t)$ with power spectral density $N_0/2$ W/Hz be applied to an ideal low-pass filter with bandwidth B Hz and unitary magnitude response. (a) Determine and sketch $S_N(f)$, the PSD of the noise $N(t)$ at the output of the filter. (b) Determine and sketch $R_N(\tau)$, the autocorrelation function of $N(t)$. (c) Compute N , the average power of the noise at the output of the filter.
12. A stationary Gaussian random process $X(t)$ with mean 10 volts is applied to the input of a linear time-invariant system whose impulse response $h(t)$ is given below. The output process $Y(t)$ has variance ψ . (a) Determine $E[Y(t)]$. (b) Write the PDF of $Y(t)$.

$$h(t) = \begin{cases} e^{t/0.2}, & 0 \leq t \leq 0.1 \text{ s} \\ 0, & \text{otherwise} \end{cases}$$

13. The magnitude of the frequency response of an RC low-pass filter is:

$$|H(f)| = \frac{1}{\sqrt{1 + (2\pi fRC)^2}}, \text{ with } R = 500 \Omega \text{ and } C = 10 \mu\text{F}.$$

To the input of this filter is applied a random process $Z(t)$ generated by filtering a white Gaussian noise $W(t)$ by an ideal low-pass filter of bandwidth 50 Hz and unitary gain. The power spectral density of $W(t)$ is 0.001 W/Hz. (a) Determine the equivalent noise bandwidth of the RC filter. (b) Compute the average power of the process at the output of the filter. (c) Sketch $|U(f)|$, the magnitude of the frequency response of an ideal low-pass filter capable of providing the same average power computed in “b”, when excited by $W(t)$.

14. Using VisSim/Comm, implement a simulation capable of demonstrating that when a Gaussian random process is filtered, another Gaussian random process is produced at the output of the filter.
15. Using VisSim/Comm, implement a simulation capable of demonstrating the concept of equivalent noise bandwidth.

References

1. Chung, K. L. and Aitsahlia, F. Elementary Probability Theory with Stochastic Processes and an Introduction to Mathematical Finance. 4th Ed. New York, USA: Springer-Verlag, 2003.
2. Coates, R. F. W., Janacek, G. J. and Lever, K. V. Monte Carlo Simulation and Random Number Generation. IEEE J. Selected Areas in Comm., vol. 6, pp. 58–66, January 1988.
3. Gentle, J. E. Random Number Generation and Monte Carlo Methods. 2nd Ed. New York, USA: Springer Science and Business Media, Inc., 2003.
4. Graf, R. F. Modern Dictionary of Electronics, 7th Ed, USA: Newnes & Butterworth-Heineman, 1999.
5. Haykin, S. Communication Systems. 3rd Ed. New York, USA: John Wiley and Sons, Inc., 1994.
6. Korolov, L. B. and Sinai, Y. G. Theory of Probability and Random Processes. 2nd Ed. Berlin, Heidelberg: Springer-Verlag, 2007.
7. Lathi, B. P. An Introduction to Random Signals and Communication Theory. USA: International Textbook Company, 1968.
8. Lathi, B. P. Modern Digital and Analog Communication Systems. 3rd Ed. New York, USA: Oxford University Press, 1998.
9. Leon-Garcia, A. Probability, Statistics, and Random Processes for Electrical Engineering. 3rd Ed. Upper Saddle River, NJ, USA: Prentice Hall, 2008.
10. Papoulis, A. Probability, Random Variables, and Stochastic Processes. 3rd Ed. New York, USA: McGraw Hill, 1991.
11. Peebles, P. Z. Probability, Random Variables and Random Signal Principles. 2nd Ed. New York, USA: McGraw Hill, 1987.
12. Pham, H. (Editor). Springer Handbook of Engineering Statistics. London, UK: Springer-Verlag, 2006.
13. Proakis, J. G. Digital Communications. 3rd Ed. USA: McGraw Hill, Inc., 1995.
14. Pursley, M. B. Introduction to Digital Communications. Upper Saddle River, New Jersey, USA: Prentice-Hall, 2005.
15. Stark, H. and Woods, J. W. Probability and Random Processes with Applications to Signal Processing. 3rd Ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.
16. Wikipedia – The Free Encyclopedia on line. Available at: <http://en.wikipedia.org> (last access on June 2009).
17. Weisstein, E. Wolfram MathWorld. Available at: <http://mathworld.wolfram.com> (last access on June 2009).
18. Yates, R. D. and Goodman, D. J. Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers. 2nd Ed. USA: John Wiley & Sons, Inc, 2005.