II

RANDOM VARIABLES, DISTRIBUTION FUNCTIONS, AND EXPECTATION

1 INTRODUCTION AND SUMMARY

The purpose of this chapter is to introduce the concepts of *random variable*, *distribution and density functions*, and *expectation*. It is primarily a "definitions-and-their-understanding" chapter; although some other results are given as well. The definitions of random variable and cumulative distribution function are given in Sec. 2, and the definitions of density functions are given in Sec. 3. These definitions are easily stated since each is just a particular function. The cumulative distribution function exists and is defined for each random variable; whereas, a density function is defined only for particular random variables. Expectations of functions of random variables are the underlying concept of all of Sec. 4. This concept is introduced by considering two particular, yet extremely important, expectations. These two are the mean and variance, defined in Subsecs. 4.1 and 4.2, respectively. Subsection 4.3 is devoted to the definition and properties of expectation of a function of a random variable. A very important result in the chapter appears in Subsec. 4.4 as the Chebyshev inequality and a generalization thereof. It is nice to be able to attain so famous a result so soon and with so little weaponry. The Jensen inequality is given in

Subsec. 4.5. Moments and moment generating functions, which are expectations of particular functions, are considered in the final subsection. One major unproven result, that of the uniqueness of the moment generating function, is given there. Also included is a brief discussion of some measures of some characteristics, such as location and dispersion, of distribution or density functions.

This chapter provides an introduction to the language of *distribution theory*. Only the univariate case is considered; the bivariate and multivariate cases will be considered in Chap. IV. It serves as a preface to, or even as a companion to, Chap. III, where a number of parametric families of distribution functions is presented. Chapter III gives many examples of the concepts defined in Chap. II.

2 RANDOM VARIABLE AND CUMULATIVE DISTRIBUTION FUNCTION

2.1 Introduction

In Chap. I we defined what we meant by a probability space, which we denoted by the triplet $(\Omega, \mathscr{A}, P[\cdot])$. We started with a conceptual random experiment; we called the totality of possible outcomes of this experiment the sample space and denoted it by Ω. $\mathscr{A}$ was used to denote a collection of subsets, called events, of the sample space. Finally our probability function $P[\cdot]$ was a set function having domain $\mathscr{A}$ and counterdomain the interval $[0, 1]$. Our object was, and still is, to assess probabilities of events. In other words, we want to model our random experiment so as to be able to give values to the probabilities of events. The notion of *random variable*, to be defined presently, will be used to *describe* events, and a *cumulative distribution function* will be used to give the probabilities of certain events defined in terms of random variables; so both concepts will assist us in defining probabilities of events, our goal. One advantage that a cumulative distribution function will have over its counterpart, the probability function (they both give probabilities of events), is that it is a function with domain the real line and counterdomain the interval $[0, 1]$. Thus we will be able to graph it. It will become a convenient tool in modeling random experiments. In fact, we will often model a random experiment by assuming certain things about a random variable and its distribution function and in so doing completely bypass describing the probability space.

2.2 Definitions

We commence by defining a random variable.

Definition 1 Random Variable For a given probability space $(\Omega, \mathscr{A}, P[\cdot])$, a *random variable*, denoted by X or $X(\cdot)$, is a function with domain Ω and counterdomain the real line. The function $X(\cdot)$ must be such that the set A_r, defined by $A_r = \{\omega\colon X(\omega) \le r\}$, belongs to $\mathscr{A}$ for every real number r. ////

If one thinks in terms of a random experiment, Ω is the totality of outcomes of that random experiment, and the function, or random variable, $X(\cdot)$ with domain Ω makes some real number correspond to each outcome of the experiment. That is the important part of our definition. The fact that we also require the collection of ω's for which $X(\omega) \le r$ to be an event (i.e., an element of $\mathscr{A}$) for each real number r is not much of a restriction for our purposes since our intention is to use the notion of random variable only in describing events. We will seldom be interested in a random variable per se; rather we will be interested in events defined in terms of random variables. One might note that the $P[\cdot]$ of our probability space $(\Omega, \mathscr{A}, P[\cdot])$ is not used in our definition.

The use of words "random" and "variable" in the above definition is unfortunate since their use cannot be convincingly justified. The expression "random variable" is a misnomer that has gained such widespread use that it would be foolish for us to try to rename it.

In our definition we denoted a random variable by either $X(\cdot)$ or X. Although $X(\cdot)$ is a more complete notation, one that emphasizes that a random variable is a function, we will usually use the shorter notation of X. For many experiments, there is a need to define more than one random variable; hence further notations are necessary. We will try to use capital Latin letters with or without affixes from near the end of the alphabet to denote random variables. Also, we use the corresponding small letter to denote a value of the random variable.

EXAMPLE 1 Consider the experiment of tossing a single coin. Let the random variable X denote the number of heads. $\Omega = \{\text{head, tail}\}$, and $X(\omega) = 1$ if $\omega = \text{head}$, and $X(\omega) = 0$ if $\omega = \text{tail}$; so, the random variable X associates a real number with each outcome of the experiment. We called X a random variable so mathematically speaking we should show

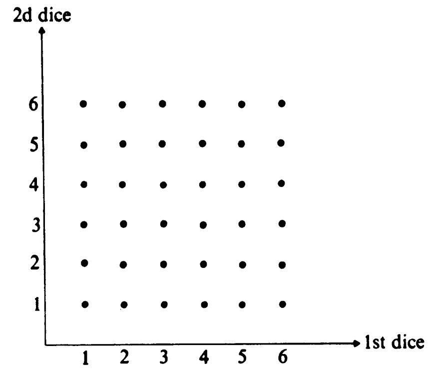


FIGURE 1

that it satisfies the definition; that is, we should show that $\{\omega: X(\omega) \leq r\}$ belongs to $\mathscr{A}$ for every real number r. $\mathscr{A}$ consists of the four subsets: ϕ, {head}, {tail}, and Ω. Now, if $r < 0$, $\{\omega: X(\omega) \leq r\} = \phi$; and if $0 \leq r < 1$, $\{\omega: X(\omega) \leq r\} = \{\text{tail}\}$; and if $r \geq 1$, $\{\omega: X(\omega) \leq r\} = \Omega = \{\text{head}, \text{tail}\}$. Hence, for each r the set $\{\omega: X(\omega) \leq r\}$ belongs to $\mathscr{A}$; so $X(\cdot)$ is a random variable. ////

EXAMPLE 2 Consider the experiment of tossing two dice. Ω can be described by the 36 points displayed in Fig. 1. $\Omega = \{(i, j): i = 1, \ldots, 6 \text{ and } j = 1, \ldots, 6\}$. Several random variables can be defined; for instance, let X denote the sum of the upturned faces; so $X(\omega) = i + j$ if $\omega = (i, j)$. Also, let Y denote the absolute difference between the upturned faces; then $Y(\omega) = |i - j|$ if $\omega = (i, j)$. It can be shown that both X and Y are random variables. We see that X can take on the values 2, 3, ..., 12 and Y can take on the values 0, 1, ..., 5. ////

In both of the above examples we described the random variables in terms of the random experiment rather than in specifying their functional form; such will usually be the case.

Definition 2 Cumulative distribution function The *cumulative distribution function* of a random variable X, denoted by $F_X(\cdot)$, is defined to be that function with domain the real line and counterdomain the interval

[0, 1] which satisfies $F_X(x) = P[X \le x] = P[\{\omega: X(\omega) \le x\}]$ for every real number x. ////

A cumulative distribution function is uniquely defined for each random variable. If it is known, it can be used to find probabilities of events defined in terms of its corresponding random variable. (One might note that it is in this definition that we use the requirement that $\{\omega: X(\omega) \le r\}$ belong to $\mathscr{A}$ for every real r which appears in our definition of random variable X.) Note that different random variables can have the same cumulative distribution function. See Example 4 below.

The use of each of the three words in the expression "cumulative distribution function" is justifiable. A cumulative distribution function is first of all a *function*; it is a *distribution* function inasmuch as it tells us how the values of the random variable are distributed, and it is a *cumulative* distribution function since it gives the distribution of values in cumulative form. Many writers omit the word "cumulative" in this definition. Examples and properties of cumulative distribution functions follow.

EXAMPLE 3 Consider again the experiment of tossing a single coin. Assume that the coin is fair. Let X denote the number of heads. Then,

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } 0 \le x < 1 \\ 1 & \text{if } 1 \le x. \end{cases}$$

Or $F_X(x) = \frac{1}{2}I_{[0,1)}(x) + I_{[1,\infty)}(x)$ in our indicator function notation. ////

EXAMPLE 4 In the experiment of tossing two fair dice, let Y denote the absolute difference. The cumulative distribution of Y, $F_Y(\cdot)$, is sketched in Fig. 2. Also, let X_k denote the value on the upturned face of the kth die for $k = 1, 2$. X_1 and X_2 are different random variables, yet both have the same cumulative distribution function, which is $F_{X_k}(x) = \sum_{i=1}^{5} \frac{i}{6} I_{[i,i+1)}(x) + I_{[6,\infty)}(x)$ and is sketched in Fig. 3. ////

Careful scrutiny of the definition and above examples might indicate the following properties of any cumulative distribution function $F_X(\cdot)$.

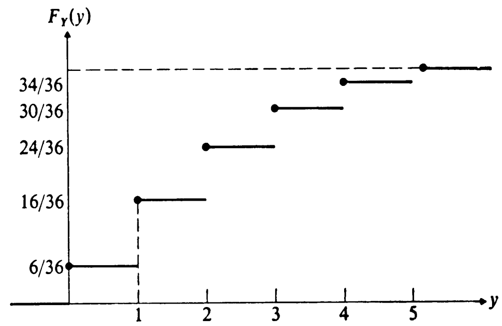


FIGURE 2

Properties of a Cumulative Distribution Function $F_X(\cdot)$

(i) $F_X(-\infty) \equiv \lim_{x \to -\infty} F_X(x) = 0$, and $F_X(+\infty) \equiv \lim_{x \to +\infty} F_X(x) = 1$.

(ii) $F_X(\cdot)$ is a monotone, nondecreasing function; that is, $F_X(a) \leq F_X(b)$ for $a < b$.

(iii) $F_X(\cdot)$ is continuous from the right; that is,

$$\lim_{0 < h \to 0} F_X(x + h) = F_X(x).$$

Except for (ii), we will not prove these properties. Note that the event $\{\omega\colon X(\omega) \leq b\} = \{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ and $\{X \leq a\} \cap \{a < X \leq b\} = \phi$; hence, $F_X(b) = P[X \leq b] = P[X \leq a] + P[a < X \leq b] \geq P[X \leq a] = F_X(a)$ which proves (ii). Property (iii), the continuity of $F_X(\cdot)$ from the right, results from our defining $F_X(x)$ to be $P[X \leq x]$. If we had defined, as some authors do, $F_X(x)$ to be $P[X < x]$, then $F_X(\cdot)$ would have been continuous from the left.

Definition 3 Cumulative distribution function Any function $F(\cdot)$ with domain the real line and counterdomain the interval [0, 1] satisfying the above three properties is defined to be a *cumulative distribution function.* ////

This definition allows us to use the term "cumulative distribution function" without mentioning random variable.

After defining what is meant by continuous and discrete random variables in the first two subsections of the next section, we will give another property that cumulative distribution functions possess, the property of decomposition into three parts.

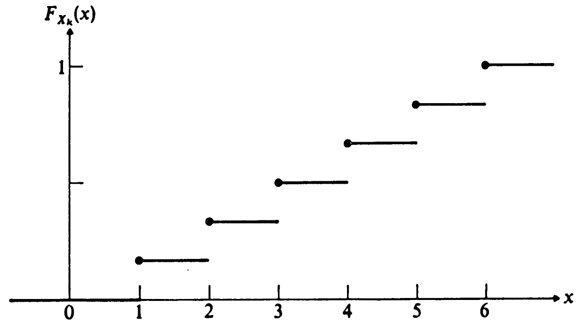


FIGURE 3

The cumulative distribution functions defined here are univariate; the introduction of bivariate and multivariate cumulative distribution functions will be deferred until Chap. IV.

3 DENSITY FUNCTIONS

Random variable and the cumulative distribution function of a random variable have been defined. The cumulative distribution function described the distribution of values of the random variable. For two distinct classes of random variables, the distribution of values can be described more simply by using *density functions.* These two classes, distinguished by the words "discrete" and "continuous," are considered in the next two subsections.

3.1 Discrete Random Variables

Definition 4 Discrete random variable A random variable X will be defined to be *discrete* if the range of X is countable. If a random variable X is discrete, then its corresponding cumulative distribution function $F_X(\cdot)$ will be defined to be *discrete.* ////

By the range of X being countable we mean that there exists a finite or denumerable set of real numbers, say $x_1, x_2, x_3, \ldots$, such that X takes on values only in that set. If X is discrete with distinct values $x_1, x_2, \ldots, x_n, \ldots$, then $\Omega = \bigcup_n \{\omega\colon X(\omega) = x_n\} = \bigcup_n \{X = x_n\}$, and $\{X = x_i\} \cap \{X = x_j\} = \phi$ for $i \neq j$; hence $1 = P[\Omega] = \sum_n P[X = x_n]$ by the third axiom of probability.

Definition 5 Discrete density function of a discrete random variable If X is a discrete random variable with distinct values $x_1, x_2, \ldots, x_n, \ldots,$ then the function, denoted by $f_X(\cdot)$ and defined by

$$f_X(x) = \begin{cases} P[X = x_j] & \text{if } x = x_j, j = 1, 2, \ldots, n, \ldots \\ 0 & \text{if } x \neq x_j \end{cases} \tag{1}$$

is defined to be the *discrete density function* of X. ////

The values of a discrete random variable are often called *mass points*; and, $f_X(x_j)$ denotes the *mass* associated with the *mass point* x_j. *Probability mass function*, *discrete frequency function*, and *probability function* are other terms used in place of *discrete density function.* Also, the notation $p_X(\cdot)$ is sometimes used intead of $f_X(\cdot)$ for discrete density functions. $f_X(\cdot)$ is a function with domain the real line and counterdomain the interval [0, 1]. If we use the indicator function,

$$f_X(x) = \sum_{n=1}^{\infty} P[X = x_n] I_{\{x_n\}}(x), \tag{2}$$

where $I_{\{x_n\}}(x) = 1$ if $x = x_n$ and $I_{\{x_n\}}(x) = 0$ if $x \neq x_n$.

Theorem 1 Let X be a discrete random variable. $F_X(\cdot)$ can be obtained from $f_X(\cdot)$, and vice versa.

PROOF Denote the mass points of X by $x_1, x_2, \ldots$. Suppose $f_X(\cdot)$ is given; then $F_X(x) = \sum_{\{j: x_j \leq x\}} f_X(x_j)$. Conversely, suppose $F_X(\cdot)$ is given; then $f_X(x_j) = F_X(x_j) - \lim_{0<h\to 0} F_X(x_j - h)$; hence $f_X(x_j)$ can be found for each mass point x_j; however, $f_X(x) = 0$ for $x \neq x_j, j = 1, 2, \ldots,$ so $f_X(x)$ is determined for all real numbers. ////

EXAMPLE 5 To illustrate what is meant in Theorem 1, consider the experiment of tossing a single die. Let X denote the number of spots on the upper face:

$$f_X(x) = \left(\frac{1}{6}\right) I_{\{1, 2, \ldots, 6\}}(x),$$

and

$$F_X(x) = \sum_{i=1}^{5} (i/6) I_{[i, i+1)}(x) + I_{[6, \infty)}(x).$$

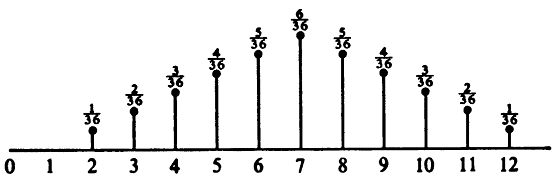


FIGURE 4

According to Theorem 1, for given $f_X(\cdot)$, $F_X(x)$ can be found for any x; for instance, if $x = 2.5$,

$$F_X(2.5) = \sum_{\{j:\, x_j \leq 2.5\}} f_X(x_j) = f_X(1) + f_X(2) = \frac{2}{6}.$$

And, if $F_X(\cdot)$ is given, $f_X(x)$ can be found for any x. For example, for $x = 3$,

$$f_X(3) = F_X(3) - \lim_{0<h\to 0} F_X(3-h) = \left(\frac{3}{6}\right) - \left(\frac{2}{6}\right) = \frac{1}{6}. \qquad ////$$

The cumulative distribution function of a discrete random variable has steps at the mass points; that is, at the mass point x_j, $F_X(\cdot)$ has a step of size $f_X(x_j)$, and $F_X(\cdot)$ is flat between mass points.

EXAMPLE 6 Consider the experiment of tossing two dice. Let X denote the total of the upturned faces. The mass points of X are 2, 3, ..., 12. $f_X(\cdot)$ is sketched in Fig. 4. Let Y denote the absolute difference of the upturned faces; then $f_Y(\cdot)$ is given in tabular form by

y	0	1	2	3	4	5
$f_Y(y)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

////

The discrete density function tells us how likely or probable each of the values of a discrete random variable is. It also enables one to calculate the probability of events described in terms of the discrete random variable X. For example, let X have mass points $x_1, x_2, \ldots, x_n, \ldots$; then $P[a < X \leq b] = \sum_{j:\{a<x_j\leq b\}} f_X(x_j)$ for $a < b$.

Definition 6 Discrete density function Any function $f(\cdot)$ with domain the real line and counterdomain [0, 1] is defined to be a *discrete density function* if for some countable set $x_1, x_2, \ldots, x_n, \ldots,$

(i) $f(x_j) > 0$ for $j = 1, 2, \ldots.$

(ii) $f(x) = 0$ for $x \neq x_j; j = 1, 2, \ldots.$

(iii) $\sum f(x_j) = 1$, where the summation is over the points $x_1, x_2, \ldots, x_n, \ldots.$ ////

This definition allows us to speak of discrete density functions without reference to some random variable. Hence we can talk about properties that a density function might have without referring to a random variable.

3.2 Continuous Random Variables

Definition 7 Continuous random variable A random variable X is called *continuous* if there exists a function $f_X(\cdot)$ such that $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ for every real number x. The cumulative distribution function $F_X(\cdot)$ of a continuous random variable X is called *absolutely continuous.* ////

Definition 8 Probability density function of a continuous random variable If X is a continuous random variable, the function $f_X(\cdot)$ in $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$ is called the *probability density function* of X. ////

Other names that are used instead of probability density function include *density function*, *continuous density function*, and *integrating density function.*

Note that strictly speaking *the* probability density function $f_X(\cdot)$ of a random variable X is not uniquely defined. All that the definition requires is that the integral of $f_X(\cdot)$ gives $F_X(x)$ for every x, and more than one function $f_X(\cdot)$ may satisfy such requirement. For example, suppose $F_X(x) = xI_{[0,1)}(x) + I_{[1,\infty)}(x)$; then $f_X(u) = I_{(0,1)}(u)$ satisfies $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$ for every x, and so $f_X(\cdot)$ is a probability density function of X. However $f_X(u) = I_{(0,\frac{1}{2})}(u) + 69I_{\{\frac{1}{2}\}}(u) + I_{(\frac{1}{2},1)}(u)$ also satisfies $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$. (The idea is that if the value of a function is changed at only a "few" points, then its integral is unchanged.) In practice a unique choice of $f_X(\cdot)$ is often dictated by continuity considerations and for this reason we will usually allow ourselves the liberty of speaking of *the* probability density when in fact *a* probability density is more correct.

One should point out that the word "continuous" in "continuous random variable" is not used in its usual sense. Although a random variable is a function and the notion of a continuous function is fairly well established in mathematics, "continuous" here is not used in that usual mathematical sense. In fact it is not clear in what sense it is used. Two possible justifications do come to mind. In contrasting discrete random variables with continuous random variables, one notes that a discrete random variable takes on a finite or denumerable set of values whereas a continuous random variable takes on a nondenumerable set of values. Possibly it is the connection between "nondenumerable" and "continuum" that justifies use of the word "continuous." All the continuous random variables that we shall encounter will take on a continuum of values. The second justification arises when one notes that the absolute continuity of the cumulative distribution function is the regular mathematical definition of an absolutely continuous function (in words, a function is called absolutely continuous if it can be written as the integral of its derivative); the "continuous," then, in a corresponding continuous random variable could be considered just an abbreviation of "absolutely continuous."

Theorem 2 Let X be a continuous random variable. Then $F_X(\cdot)$ can be obtained from an $f_X(\cdot)$, and vice versa.

PROOF If X is a continuous random variable and an $f_X(\cdot)$ is given, then $F_X(x)$ is obtained by integrating $f_X(\cdot)$; that is, $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$. On the other hand, if $F_X(\cdot)$ is given, then an $f_X(x)$ can be obtained by differentiation; that is, $f_X(x) = dF_X(x)/dx$ for those points x for which $F_X(x)$ is differentiable. ////

The notations for discrete density function and probability density function are the same, yet they have quite different interpretations. For discrete random variables $f_X(x) = P[X = x]$, which is not true for continuous random variables. For continuous random variables,

$$f_X(x) = \frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x + \Delta x) - F_X(x - \Delta x)}{2\Delta x};$$

hence $f_X(x)2\Delta x \approx F_X(x + \Delta x) - F_X(x - \Delta x) = P[x - \Delta x < X \leq x + \Delta x]$; that is, the probability that X is in a *small* interval containing the value x is approximately equal to $f_X(x)$ times the width of the interval. For discrete random variables $f_X(\cdot)$ is a function with domain the real line and counterdomain the interval [0, 1]; whereas, for continuous random variables $f_X(\cdot)$ is a function with domain the real line and counterdomain the infinite interval $[0, \infty)$.

Remark We will use the term "density function" without the modifier of "discrete" or "probability" to represent either kind of density. ////

EXAMPLE 7 Let X be the random variable representing the length of a telephone conversation. One could model this experiment by assuming that the distribution of X is given by $F_X(x) = (1 - e^{-\lambda x})I_{[0, \infty)}(x)$, where λ is some positive number. The corresponding probability density function would be given by $f_X(x) = \lambda e^{-\lambda x}I_{[0, \infty)}(x)$. If we assume that telephone conversations are measured in minutes, $P[5 < X \leq 10] = \int_5^{10} \lambda e^{-x\lambda}\, dx = e^{-5\lambda} - e^{-10\lambda} = e^{-1} - e^{-2} \approx .23$ for $\lambda = \frac{1}{5}$, or $P[5 < X \leq 10] = P[X \leq 10] - P[X \leq 5] = (1 - e^{-\lambda 10}) - (1 - e^{-\lambda 5}) = e^{-1} - e^{-2}$ for $\lambda = \frac{1}{5}$. ////

The probability density function is used to calculate the probability of events defined in terms of the corresponding continuous random variable X. For example, $P[a < X \leq b] = \int_a^b f_X(x)\, dx$ for $a < b$.

Definition 9 Probability density function Any function $f(\cdot)$ with domain the real line and counterdomain $[0, \infty)$ is defined to be a *probability density function* if and only if

(i) $f(x) \geq 0$ for all x.

(ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$. ////

With this definition we can speak of probability density functions without reference to random variables. We might note that a probability density function of a continuous random variable as defined in Definition 8 does indeed possess the two properties in the above definition.

3.3 Other Random Variables

Not all random variables are either continuous or discrete, or not all cumulative distribution functions are either absolutely continuous or discrete.

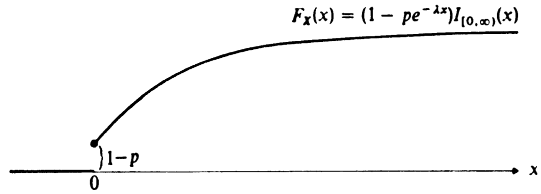


FIGURE 5

EXAMPLE 8 Consider the experiment of recording the delay that a motorist encounters at a one-way traffic stop sign. Let X be the random variable that represents the delay that the motorist experiences after making the required stop. There is a certain probability that there will be no opposing traffic so that the motorist will be able to proceed with no delay. On the other hand, if the motorist has to wait, he may have to wait for any of a continuum of possible times. This experiment could be modeled by assuming that X has a cumulative distribution function given by $F_X(x) = (1 - pe^{-\lambda x})I_{[0, \infty)}(x)$. This $F_X(x)$ has a jump of $1 - p$ at $x = 0$ but is continuous for $x > 0$. See Fig. 5. ////

Many practical examples of cumulative distribution functions that are partly discrete and partly absolutely continuous can be given. Yet there are still other types of cumulative distribution functions. There are continuous cumulative distribution functions, called *singular continuous*, whose derivative is 0 at almost all points. We will not consider such distribution functions other than to note the following result.

Decomposition of a cumulative distribution function Any cumulative distribution function $F(x)$ may be represented in the form

$$F(x) = p_1 F^{\mathrm{d}}(x) + p_2 F^{\mathrm{ac}}(x) + p_3 F^{\mathrm{sc}}(x), \qquad \text{where } p_i \geq 0,\ i = 1, 2, 3. \tag{3}$$

$\sum_{i=1}^{3} p_i = 1$, and $F^{\mathrm{d}}(\cdot)$, $F^{\mathrm{ac}}(\cdot)$, and $F^{\mathrm{sc}}(\cdot)$ are each cumulative distribution functions with $F^{\mathrm{d}}(\cdot)$ discrete, $F^{\mathrm{ac}}(\cdot)$ absolutely continuous, and $F^{\mathrm{sc}}(\cdot)$ singular continuous.

Cumulative distributions studied in this book will have at most a discrete part and an absolutely continuous part; that is, the p_3 in Eq. (3) will always be 0 for the $F(\cdot)$ that we will study.

EXAMPLE 9 To illustrate how the decomposition of a cumulative distribution function can be implemented, consider $F_X(x) = (1 - pe^{-\lambda x})I_{[0, \infty)}(x)$ as in Example 8. $F_X(x) = (1 - p)F^{d}(x) + pF^{ac}(x)$, where $F^{d}(x) = I_{[0, \infty)}(x)$ and $F^{ac}(x) = (1 - e^{-\lambda x})I_{[0, \infty)}(x)$. Note that $F_X(x) = (1 - p)F^{d}(x) + pF^{ac}(x) = (1 - p)I_{[0, \infty)}(x) + p(1 - e^{-\lambda x})I_{[0, \infty)}(x) = (1 - pe^{-\lambda x})I_{[0, \infty)}(x)$. ////

A density function corresponding to a cumulative distribution that is partly discrete and partly absolutely continuous could be defined as follows: If $F(x) = (1 - p)F^{d}(x) + pF^{ac}(x)$, where $0 < p < 1$ and $F^{d}(\cdot)$ and $F^{ac}(\cdot)$ are, respectively, discrete and absolutely continuous cumulative distribution functions, let the density function $f(x)$ corresponding to $F(x)$ be defined by $f(x) = (1 - p)f^{d}(x) + pf^{ac}(x)$, where $f^{d}(\cdot)$ is the discrete density function corresponding to $F^{d}(\cdot)$ and $f^{ac}(\cdot)$ is the probability density function corresponding to $F^{ac}(\cdot)$. Such a density function would require careful interpretation; so when considering cumulative distribution functions that are partly discrete and partly continuous, we will tend to work with the cumulative distribution function itself rather than with a density function.

Remark In future chapters we will frequently have to state that a random variable has a certain distribution. We will make such a statement by giving either the cumulative distribution function or the density function of the random variable of interest. ////

4 EXPECTATIONS AND MOMENTS

An extremely useful concept in problems involving random variables or distributions is that of *expectation*. The subsections of this section give definitions and results regarding expectations.

4.1 Mean

Definition 10 Mean Let X be a random variable. The *mean* of X, denoted by μ_X or $\mathscr{E}[X]$, is defined by:

(i)
$$\mathscr{E}[X] = \sum x_j f_X(x_j) \tag{4}$$

if X is discrete with mass points $x_1, x_2, \ldots, x_j, \ldots$.

(ii)
$$\mathscr{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx \tag{5}$$

if X is continuous with probability density function $f_X(x)$.

(iii)
$$\mathscr{E}[X] = \int_0^{\infty} [1 - F_X(x)]\, dx - \int_{-\infty}^0 F_X(x)\, dx \tag{6}$$

for an arbitrary random variable X. ////

In (i), $\mathscr{E}[X]$ is defined to be the indicated series provided that the series is absolutely convergent; otherwise, we say that the mean does not exist. And in (ii), $\mathscr{E}[X]$ is defined to be the indicated integral if the integral exists; otherwise, we say that the mean does not exist. Finally, in (iii), we require that both integrals be finite for the existence of $\mathscr{E}[X]$.

Note what the definition says: In $\sum_j x_j f_X(x_j)$, the summand is the jth value of the random variable X multiplied by the probability that X equals that jth value, and then the summation is over all values. So $\mathscr{E}[X]$ is an "average" of the values that the random variable takes on, where each value is weighted by the probability that the random variable is equal to that value. Values that are more probable receive more weight. The same is true in integral form in (ii). There the value x is multiplied by the approximate probability that X equals the value x, namely $f_X(x)\, dx$, and then integrated over all values.

Several remarks are in order.

Remark In the definition of a mean of a random variable, only density functions [in (i) and (ii)] or distribution functions [in (iii)] were used; hence we have really defined the mean for these functions without reference to random variables. We then call the defined mean the mean of the cumulative distribution function or of the appropriate density function. Hence, we can and will speak of the mean of a distribution or density function as well as the mean of a random variable. ////

Remark $\mathscr{E}[X]$ is the center of gravity (or *centroid*) of the unit mass that is determined by the density function of X. So the mean of X is a measure of where the values of the random variable X are "centered." Other measures of "location" or "center" of a random variable or its corresponding density are given in Subsec. 4.6. ////

Remark (iii) of the definition is for all random variables; whereas, (i) is for discrete random variables, and (ii) is for continuous random variables. Of course, $\mathscr{E}[X]$ could have been defined by just giving (iii). The reason for including (i) and (ii) is that they are more intuitive for their respective cases. It can be proved, although we will not do it, that (i) follows from (iii) in the case of discrete random variables and (ii) follows from (iii) in the case of continuous random variables. Our main use of (iii) will be in finding the mean of a random variable X that is neither discrete nor continuous. See Example 12 below. ////

EXAMPLE 10 Consider the experiment of tossing two dice. Let X denote the total of the two dice and Y their absolute difference. The discrete density functions for X and Y are given in Example 6.

$$\mathscr{E}[Y] = \sum y_j f_Y(y_j) = \sum_{i=0}^{5} i f_Y(i) = 0 \cdot \tfrac{6}{36} + 1 \cdot \tfrac{10}{36}$$

$$+ 2 \cdot \tfrac{8}{36} + 3 \cdot \tfrac{6}{36} + 4 \cdot \tfrac{4}{36} + 5 \cdot \tfrac{2}{36} = \tfrac{70}{36}.$$

$$\mathscr{E}[X] = \sum_{i=2}^{12} i f_X(i) = 7.$$

Note that $\mathscr{E}[Y]$ is not one of the possible values of Y. ////

EXAMPLE 11 Let X be a continuous random variable with probability density function $f_X(x) = \lambda e^{-\lambda x} I_{[0, \infty)}(x)$.

$$\mathscr{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^{\infty} x \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

The corresponding cumulative distribution function is

$$F_X(x) = (1 - e^{-\lambda x}) I_{[0, \infty)}(x); \text{ so } \mathscr{E}[X] = \int_0^{\infty} [1 - F_X(x)]\, dx$$

$$- \int_{-\infty}^{0} F_X(x)\, dx = \int_0^{\infty} (1 - 1 + e^{-\lambda x})\, dx = 1/\lambda. \qquad ////$$

EXAMPLE 12 Let X be a random variable with cumulative distribution function given by $F_X(x) = (1 - pe^{-\lambda x}) I_{[0, \infty)}(x)$; then

$$\mathscr{E}[X] = \int_0^{\infty} [1 - F_X(x)]\, dx - \int_{-\infty}^{0} F_X(x)\, dx = \int_0^{\infty} pe^{-\lambda x}\, dx = \frac{p}{\lambda}.$$

Here, we have used Eq. (6) to find the mean of a random variable that is partly discrete and partly continuous. ////

EXAMPLE 13 Let X be a random variable with probability density function given by $f_X(x) = x^{-2}I_{[1,\infty)}(x)$; then

$$\mathscr{E}[X] = \int_1^\infty x\frac{dx}{x^2} = \lim_{b\to\infty}\log_e b = \infty,$$

so we say that $\mathscr{E}[X]$ does not exist. We might also say that the mean of X is infinite since it is clear here that the integral that defines the mean is infinite. ////

4.2 Variance

The mean of a random variable X, defined in the previous subsection, was a measure of *central location* of the density of X. The *variance* of a random variable X will be a measure of the *spread* or *dispersion* of the density of X.

Definition 11 Variance Let X be a random variable, and let μ_X be $\mathscr{E}[X]$. The *variance of* X, denoted by σ_X^2 or var $[X]$, is defined by

(i)
$$\text{var}[X] = \sum_j (x_j - \mu_X)^2 f_X(x_j) \tag{7}$$

if X is discrete with mass points $x_1, x_2, \ldots, x_j, \ldots$.

(ii)
$$\text{var}[X] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx \tag{8}$$

if X is continuous with probability density function $f_X(x)$.

(iii)
$$\text{var}[X] = \int_0^\infty 2x[1 - F_X(x) + F_X(-x)]\, dx - \mu_X^2 \tag{9}$$

for an arbitrary random variable X. ////

The variances are defined only if the series in (i) is convergent or if the integrals in (ii) and (iii) exist. Again, the variance of a random variable is defined in terms of the density function or cumulative distribution function of the random variable; hence variance could be defined in terms of these functions without reference to a random variable.

Note what the definition says: In (i), the square of the difference between the jth value of the random variable X and the mean of X is multiplied by the probability that X equals the jth value, and then these terms are summed. More weight is assigned to the more probable squared differences. A similar comment applies for (ii). Variance is a *measure of spread* since if the values of a random variable X tend to be far from their mean, the variance of X will be larger than the variance of a comparable random variable Y whose values tend to be near their mean. It is clear from (i) and (ii) and true for (iii) that variance is nonnegative. We saw that a mean was the center of gravity of a density; similarly (for those readers familiar with elementary physics or mechanics), variance represents the moment of inertia of the same density with respect to a perpendicular axis through the center of gravity.

Definition 12 Standard deviation If X is a random variable, the *standard deviation* of X, denoted by σ_X, is defined as $+\sqrt{\text{var}[X]}$. ////

The standard deviation of a random variable, like the variance, is a measure of the spread or dispersion of the values of the random variable. In many applications it is preferable to the variance as such a measure since it will have the same measurement units as the random variable itself.

EXAMPLE 14 Let X be the total of the two dice in the experiment of tossing two dice.

$$\begin{aligned}\text{var}[X] &= \sum (x_j - \mu_X)^2 f_X(x_j) \\ &= (2-7)^2\tfrac{1}{36} + (3-7)^2\tfrac{2}{36} + (4-7)^2\tfrac{3}{36} + (5-7)^2\tfrac{4}{36} \\ &\quad + (6-7)^2\tfrac{5}{36} + (7-7)^2\tfrac{6}{36} + (8-7)^2\tfrac{5}{36} + (9-7)^2\tfrac{4}{36} \\ &\quad + (10-7)^2\tfrac{3}{36} + (11-7)^2\tfrac{2}{36} + (12-7)^2\tfrac{1}{36} = \tfrac{210}{36}.\end{aligned}$$ ////

EXAMPLE 15 Let X be a random variable with probability density given by $f_X(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x)$; then

$$\begin{aligned}\text{Var}[X] &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\,dx \\ &= \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x}\,dx \\ &= \frac{1}{\lambda^2}.\end{aligned}$$ ////

EXAMPLE 16 Let X be a random variable with cumulative distribution given by $F_X(x) = (1 - pe^{-\lambda x}) I_{[0,\infty)}(x)$; then

$$\begin{aligned}\text{Var}[X] &= \int_0^{\infty} 2x[1 - F(x) + F(-x)]\,dx - \mu_X^2 \\ &= \int_0^{\infty} 2xpe^{-\lambda x}\,dx - \left(\frac{p}{\lambda}\right)^2 \\ &= 2\frac{p}{\lambda^2} - \left(\frac{p}{\lambda}\right)^2 = \frac{p(2-p)}{\lambda^2}.\end{aligned}$$ ////

4.3 Expected Value of a Function of a Random Variable

We defined the expectation of an arbitrary random variable X, called the mean of X, in Subsec. 4.1. In this subsection, we will define the expectation of a function of a random variable for discrete or continuous random variables.

Definition 13 Expectation Let X be a random variable and $g(\cdot)$ be a function with both domain and counterdomain the real line. The *expectation* or *expected value* of the function $g(\cdot)$ of the random variable X, denoted by $\mathscr{E}[g(X)]$, is defined by:

(i)
$$\mathscr{E}[g(X)] = \sum_j g(x_j) f_X(x_j) \tag{10}$$

if X is discrete with mass points $x_1, x_2, \ldots, x_j, \ldots$ (provided this series is absolutely convergent).

(ii)
$$\mathscr{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \tag{11}$$

if X is continuous with probability density function $f_X(x)$ (provided $\int_{-\infty}^{\infty} |g(x)| f_X(x)\, dx < \infty$).* ////

Expectation or expected value is not really a very good name since it is not necessarily what you "expect." For example, the expected value of a discrete random variable is not necessarily one of the possible values of the discrete random variable, in which case, you would not "expect" to get the expected value. A better name might be "average value" rather than "expected value."

Since $\mathscr{E}[g(X)]$ is defined in terms of the density function of X, it could be defined without reference to a random variable.

Remark If $g(x) = x$, then $\mathscr{E}[g(X)] = \mathscr{E}[X]$ is the mean of X. If $g(x) = (x - \mu_X)^2$, then $\mathscr{E}[g(X)] = \mathscr{E}[(X - \mu_X)^2] = \text{var}[X]$. ////

* $\mathscr{E}[g(X)]$ has been defined here for random variables that are either discrete or continuous; it can be defined for other random variables as well. For the reader who is familiar with the Stieltjes integral, $\mathscr{E}[g(X)]$ is defined as the Stieltjes integral $\int_{-\infty}^{\infty} g(x)\, dF_X(x)$ (provided this integral exists), where $F_X(\cdot)$ is the cumulative distribution function of X. If X is a random variable whose cumulative distribution function is partly discrete and partly continuous, then (according to Subsec. 3.3) $F_X(x) = (1-p)F^d(x) + pF^{ac}(x)$ for some $0 < p < 1$. Now $\mathscr{E}[g(X)]$ can be defined to be $\mathscr{E}[g(X)] = (1-p)\sum g(x_j) f^d(x_j) + p\int_{-\infty}^{\infty} g(x) f^{ac}(x)\, dx$, where $f^d(\cdot)$ is the discrete density function corresponding to $F^d(\cdot)$ and $f^{ac}(\cdot)$ is the probability density function corresponding to $F^{ac}(\cdot)$.

Theorem 3 Below are properties of expected value:

(i) $\mathscr{E}[c] = c$ for a constant c.

(ii) $\mathscr{E}[cg(X)] = c\mathscr{E}[g(X)]$ for a constant c.

(iii) $\mathscr{E}[c_1 g_1(X) + c_2 g_2(X)] = c_1 \mathscr{E}[g_1(X)] + c_2 \mathscr{E}[g_2(X)]$.

(iv) $\mathscr{E}[g_1(X)] \leq \mathscr{E}[g_2(X)]$ if $g_1(x) \leq g_2(x)$ for all x.

PROOF Assume X is continuous. To prove (i), take $g(x) = c$, then

$$\mathscr{E}[g(X)] = \mathscr{E}[c] = \int_{-\infty}^{\infty} c f_X(x)\, dx = c \int_{-\infty}^{\infty} f_X(x)\, dx = c.$$

$$\mathscr{E}[cg(X)] = \int_{-\infty}^{\infty} cg(x) f_X(x)\, dx = c \int_{-\infty}^{\infty} g(x) f_X(x)\, dx = c\mathscr{E}[g(X)],$$

which proves (ii). (iii) is given by

$$\begin{aligned}\mathscr{E}[c_1 g_1(X) + c_2 g_2(X)] &= \int_{-\infty}^{\infty} [c_1 g_1(x) + c_2 g_2(x)] f_X(x)\, dx \\ &= c_1 \int_{-\infty}^{\infty} g_1(x) f_X(x)\, dx + c_2 \int_{-\infty}^{\infty} g_2(x) f_X(x)\, dx \\ &= c_1 \mathscr{E}[g_1(X)] + c_2 \mathscr{E}[g_2(X)].\end{aligned}$$

Finally,

$$0 \leq \mathscr{E}[g_2(X) - g_1(X)] = \mathscr{E}[g_2(X)] - \mathscr{E}[g_1(X)],$$

which gives (iv).

Similar proofs could be presented for the discrete random variable case. ////

Theorem 4 If X is a random variable, $\operatorname{var}[X] = \mathscr{E}[(X - \mathscr{E}[X])^2] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2$ provided $\mathscr{E}[X^2]$ exists.

PROOF (We first note that if $\mathscr{E}[X^2]$ exists, then $\mathscr{E}[X]$ exists.)* By our definitions of variance and $\mathscr{E}[g(X)]$, it follows that $\operatorname{var}[X] = \mathscr{E}[(X - \mathscr{E}[X])^2]$. Now $\mathscr{E}[(X - \mathscr{E}[X])^2] = \mathscr{E}[X^2 - 2X\mathscr{E}[X] + (\mathscr{E}[X])^2] = \mathscr{E}[X^2] - 2(\mathscr{E}[X])^2 + (\mathscr{E}[X])^2 = \mathscr{E}[X^2] - (\mathscr{E}[X])^2$. ////

The above theorem provides us with two methods of calculating a variance, namely $\mathscr{E}[(X - \mu_X)^2]$ or $\mathscr{E}[X^2] - \mu_X^2$. Note that both methods require μ_X.

* Here and in the future we are not going to concern ourselves with checking existence.

$\mathscr{E}[g(X)]$ is used in each of the following three subsections. In Subsec. 4.4 and 4.5 two inequalities involving $\mathscr{E}[g(X)]$ are given. Definitions and examples of $\mathscr{E}[g(X)]$ for particular functions $g(\cdot)$ are given in Subsec. 4.6.

4.4 Chebyshev Inequality

Theorem 5 Let X be a random variable and $g(\cdot)$ a nonnegative function with domain the real line; then

$$P[g(X) \geq k] \leq \frac{\mathscr{E}[g(X)]}{k} \qquad \text{for every } k > 0. \tag{12}$$

PROOF Assume that X is a continuous random variable with probability density function $f_X(\cdot)$; then

$$\begin{aligned}\mathscr{E}[g(X)] &= \int_{-\infty}^{\infty} g(x)f_X(x)\,dx = \int_{\{x:\,g(x)\geq k\}} g(x)f_X(x)\,dx \\ &+ \int_{\{x:\,g(x)<k\}} g(x)f_X(x)\,dx \geq \int_{\{x:\,g(x)\geq k\}} g(x)f_X(x)\,dx \\ &\geq \int_{\{x:\,g(x)\geq k\}} kf_X(x)\,dx = kP[g(X) \geq k].\end{aligned}$$

Divide by k, and the result follows. A similar proof holds for X discrete.

////

Corollary **Chebyshev inequality** If X is a random variable with finite variance,

$$P[|X - \mu_X| \geq r\sigma_X] = P[(X - \mu_X)^2 \geq r^2\sigma_X^2] \leq \frac{1}{r^2} \qquad \text{for every } r > 0. \tag{13}$$

PROOF Take $g(x) = (x - \mu_X)^2$ and $k = r^2\sigma_X^2$ in Eq. (12) of Theorem 5.

////

Remark If X is a random variable with finite variance,

$$P[|X - \mu_X| < r\sigma_X] \geq 1 - \frac{1}{r^2}, \tag{14}$$

which is just a rewriting of Eq. (13). ////

The Chebyshev inequality is used in various ways. We will use it later to prove the law of large numbers. Note what Eq. (14) says:

$$P[\mu_X - r\sigma_X < X < \mu_X + r\sigma_X] \geq 1 - \frac{1}{r^2};$$

that is, the probability that X falls within $r\sigma_X$ units of μ_X is greater than or equal to $1 - 1/r^2$. For $r = 2$, one gets $P[\mu_X - 2\sigma_X < X < \mu_X + 2\sigma_X] \geq \frac{3}{4}$, or for any random variable X having finite variance at least three-fourths of the mass of X falls within two standard deviations of its mean.

Ordinarily, to calculate the probability of an event described in terms of a random variable X, the distribution or density of X is needed; the Chebyshev inequality gives a bound, which does not depend on the distribution of X, for the probability of particular events described in terms of a random variable and its mean and variance.

4.5 Jensen Inequality

Definition 14 Convex function A continuous function $g(\cdot)$ with domain and counterdomain the real line is called *convex* if for every x_0 on the real line, there exists a line which goes through the point $(x_0, g(x_0))$ and lies on or under the graph of the function $g(\cdot)$. ////

Theorem 6 Jensen inequality Let X be a random variable with mean $\mathscr{E}[X]$, and let $g(\cdot)$ be a convex function; then $\mathscr{E}[g(X)] \geq g(\mathscr{E}[X])$.

PROOF Since $g(x)$ is continuous and convex, there exists a line, say $l(x) = a + bx$, satisfying $l(x) = a + bx \leq g(x)$ and $l(\mathscr{E}[X]) = g(\mathscr{E}[X])$. $l(x)$ is a line given by the definition of continuous and convex that goes through the point $(\mathscr{E}[X], g(\mathscr{E}[X]))$. Note that $\mathscr{E}[l(X)] = \mathscr{E}[(a + bX)] = a + b\mathscr{E}[X] = l(\mathscr{E}[X])$; hence $g(\mathscr{E}[X]) = l(\mathscr{E}[X]) = \mathscr{E}[l(X)] \leq \mathscr{E}[g(X)]$ [using property (iv) of expected values (see Theorem 3) for the last inequality]. ////

The Jensen inequality can be used to prove the Rao-Blackwell theorem to appear in Chap. VII. We point out that, in general, $\mathscr{E}[g(X)] \neq g(\mathscr{E}[X])$; for example, note that $g(x) = x^2$ is convex; hence $\mathscr{E}[X^2] \geq (\mathscr{E}[X])^2$, which says that the variance of X, which is $\mathscr{E}[X^2] - (\mathscr{E}[X])^2$, is nonnegative.

4.6 Moments and Moment Generating Functions

The *moments* (or *raw moments*) of a random variable or of a distribution are the expectations of the powers of the random variable which has the given distribution.

Definition 15 Moments If X is a random variable, the *rth moment* of X, usually denoted by μ_r', is defined as

$$\mu_r' = \mathscr{E}[X^r] \tag{15}$$

if the expectation exists. ////

Note that $\mu_1' = \mathscr{E}[X] = \mu_X$, the mean of X.

Definition 16 Central moments If X is a random variable, the *rth central moment* of X about a is defined as $\mathscr{E}[(X - a)^r]$. If $a = \mu_X$, we have the *rth central moment* of X about μ_X, denoted by μ_r, which is

$$\mu_r = \mathscr{E}[(X - \mu_X)^r]. \tag{16}$$

////

Note that $\mu_1 = \mathscr{E}[(X - \mu_X)] = 0$ and $\mu_2 = \mathscr{E}[(X - \mu_X)^2]$, the variance of X. Also, note that all odd moments of X about μ_X are 0 if the density function of X is symmetrical about μ_X, provided such moments exist.

In the ensuing few paragraphs we will comment on how the first four moments of a random variable or density are used as measures of various characteristics of the corresponding density. For some of these characteristics, other measures can be defined in terms of *quantiles.*

Definition 17 Quantile The *qth quantile* of a random variable X or of its corresponding distribution is denoted by ξ_q and is defined as the smallest number ξ satisfying $F_X(\xi) \geq q$. ////

If X is a continuous random variable, then the *qth quantile* of X is given as the smallest number ξ satisfying $F_X(\xi) = q$. See Fig. 6.

Definition 18 Median The *median* of a random variable X, denoted by med_X, med (X), or $\xi_{.50}$, is the .5th quantile. ////

Remark In some texts the median of X is alternatively defined as any number, say med (X), satisfying $P[X \leq \text{med}(X)] \geq \frac{1}{2}$ and $P[X \geq \text{med}(X)] \geq \frac{1}{2}$. ////

If X is a continuous random variable, then the median of X satisfies

$$\int_{-\infty}^{\text{med}(X)} f_X(x)\, dx = \tfrac{1}{2} = \int_{\text{med}(X)}^{\infty} f_X(x)\, dx;$$

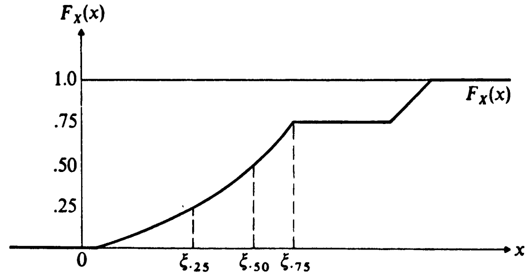


FIGURE 6

so the median of X is any number that has half the mass of X to its right and the other half to its left, which justifies use of the word "median."

We have already mentioned that $\mathscr{E}[X]$, the first moment, locates the "center" of the density of X. The median of X is also used to indicate a central location of the density of X. A third measure of location of the density of X, though not necessarily a measure of central location, is the *mode* of X, which is defined as that point (if such a point exists) at which $f_X(\cdot)$ attains its maximum. Other measures of location [for example, $\frac{1}{2}(\xi_{.25} + \xi_{.75})$] could be devised, but three, mean, median, and mode, are the ones commonly used.

We previously mentioned that the second moment about the mean, the variance of a distribution, measures the spread or dispersion of a distribution. Let us look a little further into the manner in which the variance characterizes the distribution. Suppose that $f_1(x)$ and $f_2(x)$ are two densities with the same mean μ such that

$$\int_{\mu-a}^{\mu+a} [f_1(x) - f_2(x)]\, dx \geq 0 \qquad (17)$$

for every value of a. Two such densities are illustrated in Fig. 7. It can be shown that in this case the variance σ_1^2 in the first density is smaller than the

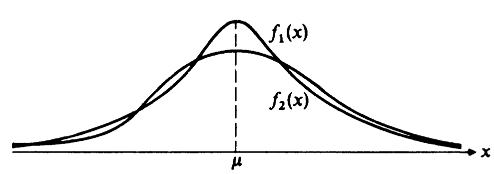
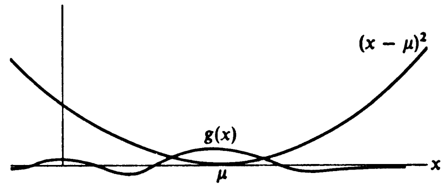


FIGURE 7

FIGURE 8


variance σ_2^2 in the second density. We shall not take the time to prove this in detail, but the argument is roughly this: Let

$$g(x) = f_1(x) - f_2(x),$$

where $f_1(x)$ and $f_2(x)$ satisfy Eq. (17). Since $\int_{-\infty}^{\infty} g(x)\,dx = 0$, the positive area between $g(x)$ and the x axis is equal to the negative area. Furthermore, in view of Eq. (17), every positive element of area $g(x')\,dx'$ may be balanced by a negative element $g(x'')\,dx''$ in such a way that x'' is further from μ than x'. When these elements of area are multiplied by $(x-\mu)^2$, the negative elements will be multiplied by larger factors than their corresponding positive elements (see Fig. 8); hence

$$\int_{-\infty}^{\infty} (x-\mu)^2 g(x)\,dx < 0$$

unless $f_1(x)$ and $f_2(x)$ are equal. Thus it follows that $\sigma_1^2 < \sigma_2^2$. The converse of these statements is not true. That is, if one is told that $\sigma_1^2 < \sigma_2^2$, he cannot conclude that the corresponding densities satisfy Eq. (17) for all values of a; although it can be shown that Eq. (17) must be true for certain values of a. Thus the condition $\sigma_1^2 < \sigma_2^2$ does not give one any precise information about the nature of the corresponding distributions, but it is evident that $f_1(x)$ has more area near the mean than $f_2(x)$, at least for certain intervals about the mean.

We indicated above how variance is used as a measure of spread or dispersion of a distribution. Alternative measures of dispersion can be defined in terms of quantiles. For example $\xi_{.75} - \xi_{.25}$, called the *interquartile range*, is a measure of spread. Also, $\xi_p - \xi_{1-p}$ for some $\frac{1}{2} < p < 1$ is a possible measure of spread.

The third moment μ_3 about the mean is sometimes called a measure of asymmetry, or *skewness*. Symmetrical distributions like those in Fig. 9 can be shown to have $\mu_3 = 0$. A curve shaped like $f_1(x)$ in Fig. 10 is said to be *skewed to the left* and can be shown to have a negative third moment about the mean; one shaped like $f_2(x)$ is called *skewed to the right* and can be shown to have a positive third moment about the mean. Actually, however, knowledge of the

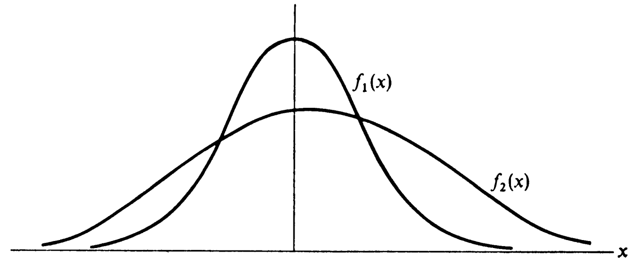


FIGURE 9

third moment gives almost no clue as to the shape of the distribution, and we mention it mainly to point out that fact. Thus, for example, the density $f_3(x)$ in Fig. 10 has $\mu_3 = 0$, but it is far from symmetrical. By changing the curve slightly we could give it either a positive or negative third moment. The ratio μ_3/σ^3, which is unitless, is called the *coefficient of skewness.*

The quantity $\mathscr{s} = (\text{mean} - \text{median})/(\text{standard deviation})$ provides an alternative measure of skewness. It can be proved that $-1 \leq \mathscr{s} \leq 1$.

The fourth moment about the mean is sometimes used as a measure of *excess* or *kurtosis*, which is the degree of flatness of a density near its center. Positive values of $\mu_4/\sigma^4 - 3$, called the *coefficient of excess* or *kurtosis*, are sometimes used to indicate that a density is more peaked around its center than the density of a normal curve (see Subsec. 3.2 of Chap. III), and negative values are sometimes used to indicate that a density is more flat around its center than the density of a normal curve. This measure, however, suffers from the same failing as does the measure of skewness; namely, it does not always measure what it is supposed to.

While a particular moment or a few of the moments may give little information about a distribution (see Fig. 11 for a sketch of two densities having the same first four moments. See Ref. 40. Also see Prob. 30 in Chap. III.), the entire set of moments $(\mu_1', \mu_2', \mu_3', \ldots)$ will ordinarily determine the distri-

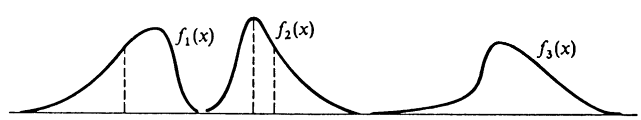


FIGURE 10

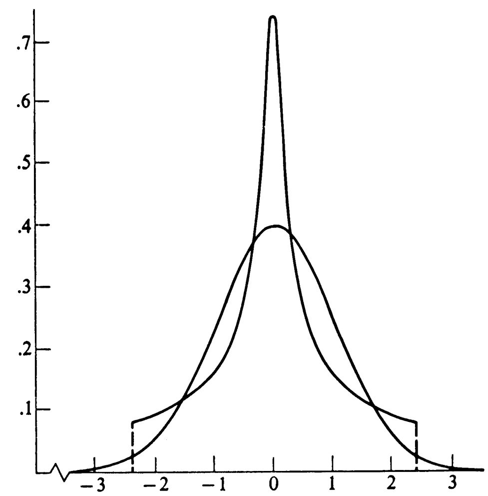


FIGURE 11

bution exactly, and for this reason we shall have occasion to use the moments in theoretical work.

In applied statistics, the first two moments are of great importance, as we shall see, but the third and higher moments are rarely useful. Ordinarily one does not know what distribution function one is working with in a practical problem, and often it makes little difference what the actual shape of the distribution is. But it is usually necessary to know at least the location of the distribution and to have some idea of its dispersion. These characteristics can be estimated by examining a sample drawn from a set of objects known to have the distribution in question. This estimation problem is probably the most important problem in applied statistics, and a large part of this book will be devoted to a study of it.

We now define another kind of moment, *factorial moment.*

Definition 19 Factorial moment If X is a random variable, the *rth factorial moment of X* is defined as (r is a positive integer):

$$\mathscr{E}[X(X-1)\cdots(X-r+1)]. \tag{18}$$

////

For some random variables (usually discrete), factorial moments are easier to calculate than raw moments. However the raw moments can be obtained from the factorial moments and vice versa.

The moments of a density function play an important role in theoretical and applied statistics. In fact, in some cases, if all the moments are known, the density can be determined. This will be discussed briefly at the end of this subsection. Since the moments of a density are important, it would be useful if a function could be found that would give us a representation of all the moments. Such a function is called a *moment generating function.*

Definition 20 Moment generating function Let X be a random variable with density $f_X(\cdot)$. The expected value of e^{tX} is defined to be the *moment generating function* of X if the expected value exists for every value of t in some interval $-h < t < h$; $h > 0$. The moment generating function, denoted by $m_X(t)$ or $m(t)$, is

$$m(t) = \mathscr{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx \tag{19}$$

if the random variable X is continuous and is

$$m(t) = \mathscr{E}[e^{tX}] = \sum_x e^{tx} f_X(x)$$

if the random variable is discrete. ////

One might note that a moment generating function is defined in terms of a density function, and since density functions were defined without reference to random variables (see Definitions 6 and 9), a moment generating function can be discussed without reference to random variables.

If a moment generating function exists, then $m(t)$ is continuously differentiable in some neighborhood of the origin. If we differentiate the moment generating function r times with respect to t, we have

$$\frac{d^r}{dt^r} m(t) = \int_{-\infty}^{\infty} x^r e^{xt} f_X(x)\, dx, \tag{20}$$

and letting $t \to 0$, we find

$$\frac{d^r}{dt^r} m(0) = \mathscr{E}[X^r] = \mu_r', \tag{21}$$

where the symbol on the left is to be interpreted to mean the rth derivative of $m(t)$ evaluated as $t \to 0$. Thus the moments of a distribution may be obtained from the moment generating function by differentiation, hence its name.

If in Eq. (19) we replace e^{xt} by its series expansion, we obtain the series expansion of $m(t)$ in terms of the moments of $f_X(\cdot)$; thus

$$m(t) = \mathscr{E}\left[1 + Xt + \frac{1}{2!}(Xt)^2 + \frac{1}{3!}(Xt)^3 + \cdots\right]$$
$$= 1 + \mu_1' t + \frac{1}{2!}\mu_2' t^2 + \cdots$$
$$= \sum_{i=0}^{\infty} \frac{1}{i!}\mu_i' t^i, \tag{22}$$

from which it is again evident that μ_r' may be obtained from $m(t)$; μ_r' is the coefficient of $t^r/r!$.

EXAMPLE 17 Let X be a random variable with probability density function given by $f_X(x) = \lambda e^{-\lambda x} I_{[0, \infty)}(x)$.

$$m_X(t) = \mathscr{E}[e^{tX}] = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}\,dx = \frac{\lambda}{\lambda - t} \qquad \text{for } t < \lambda.$$

$$m'(t) = \frac{dm(t)}{dt} = \frac{\lambda}{(\lambda - t)^2} \qquad \text{hence} \quad m'(0) = \mathscr{E}[X] = \frac{1}{\lambda}.$$

And
$$m''(t) = \frac{2\lambda}{(\lambda - t)^3}, \qquad \text{so } m''(0) = \mathscr{E}[X^2] = \frac{2}{\lambda^2}.$$
////

EXAMPLE 18 Consider the random variable X having probability density function $f_X(x) = x^{-2} I_{[1, \infty)}(x)$. (See Example 13.) If the moment generating function of X exists, then it is given by $\int_1^{\infty} x^{-2} e^{tx}\,dx$. It can be shown, however, that the integral does not exist for any $t > 0$, and hence the moment generating function does not exist for this random variable X. ////

As with moments, there is also a generating function for factorial moments.

Definition 21 Factorial moment generating function Let X be a random variable. The *factorial moment generating function* is defined as $\mathscr{E}[t^X]$ if this expectation exists. ////

The factorial moment generating function is used to generate factorial moments in the same way as the raw moments are obtained from $\mathscr{E}[e^{tX}]$ except that t approaches 1 instead of 0. It sometimes simplifies finding moments of discrete distributions.

EXAMPLE 19 Suppose X has a discrete density function given by

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for } x = 0, 1, 2, \ldots.$$

Then

$$\mathscr{E}[t^X] = \sum_{x=0}^{\infty} \frac{t^x e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}e^{\lambda t} = e^{\lambda(t-1)}.$$

$$\frac{d}{dt}\mathscr{E}[t^X] = \frac{d}{dt}e^{\lambda(t-1)} = \lambda e^{\lambda(t-1)}; \qquad \text{hence} \quad \frac{d}{dt}\mathscr{E}[t^X]\bigg|_{t=1} = \lambda. \qquad ////$$

In addition to raw moments, central moments, and factorial moments, there are other kinds of moments, called *cumulants*, or *semi-invariants*. Cumulants will be defined in terms of the *cumulant generating function*. We will not make use of cumulants in this book.

Definition 22 Cumulant and cumulant generating function The logarithm of the moment generating function of X is defined to be the *cumulant generating function* of X. The *rth cumulant* of X, denoted by $\kappa_r(X)$ or κ_r, is the coefficient of $t^r/r!$ in the Taylor series expansion of the cumulant generating function. ////

A moment generating function is used, as its name suggests, to generate moments. That, however, will not be its only use for us. An important use will be in determining distributions.

Theorem 7 Let X and Y be two random variables with densities $f_X(\cdot)$ and $f_Y(\cdot)$, respectively. Suppose that $m_X(t)$ and $m_Y(t)$ both exist and are equal for all t in the interval $-h < t < h$ for some $h > 0$. Then the two cumulative distribution functions $F_X(\cdot)$ and $F_Y(\cdot)$ are equal. ////

A proof of the above theorem can be obtained using certain transform theory that is beyond the scope of this book. We should note, however, what the theorem asserts. It says that if we can find the moment generating function of a random variable, then, theoretically, we can find the distribution of the random variable since there is a unique distribution function for a given moment generating function. This theorem will prove to be extremely useful in finding the distribution of certain functions of random variables. In particular, see Sec. 4 of Chap. V.

EXAMPLE 20 Suppose that a random variable X has a moment generating function $m_X(t) = 1/(1-t)$ for $-1 < t < 1$; then we know that the density of X is given by $f_X(x) = e^{-x} I_{[0,\infty)}(x)$ since we showed in Example 17 above that $\lambda e^{-\lambda x} I_{[0,\infty)}(x)$ has $\lambda/(\lambda - t)$ for its moment generating function. ////

Problem of moments We have seen that a density function determines a set of moments μ_1', μ_2', ... when they exist. One of the important problems in theoretical statistics is this: Given a set of moments, what is the density function from which these moments came, and is there only one density function that has these particular moments? We shall give only partial answers. First, there exists a sequence of moments for which there is an infinite (nondenumerable) collection of different distribution functions having these same moments. In general, a sequence of moments μ_1', μ_2', ... does not determine a unique distribution function. However, we did see that if the moment generating function of a random variable did exist, then this moment generating function did uniquely determine the corresponding distribution function. (See Theorem 7 above.) Hence, there are conditions (existence of the moment generating function is a sufficient condition) under which a sequence of moments does uniquely determine a distribution function. The general problem of whether or not a distribution function is determined by its sequence of moments is referred to as the *problem of moments* and will not be discussed further.

PROBLEMS

1 (*a*) Show that the following are probability density functions (p.d.f.'s):

$$f_1(x) = e^{-x} I_{(0,\infty)}(x)$$
$$f_2(x) = 2e^{-2x} I_{(0,\infty)}(x)$$
$$f(x) = (\theta + 1) f_1(x) - \theta f_2(x) \qquad 0 < \theta < 1.$$

(*b*) Prove or disprove: If $f_1(x)$ and $f_2(x)$ are p.d.f.'s and if $\theta_1 + \theta_2 = 1$, then $\theta_1 f_1(x) + \theta_2 f_2(x)$ is a p.d.f.

2 Show that the following is a density function and find its median:

$$f(x) = \frac{\alpha^2(\alpha + 2x)}{x^2(\alpha + x)^2} I_{(\alpha,\infty)}(x) + \frac{x(2\alpha + x)}{\alpha(\alpha + x)^2} I_{(0,\alpha]}(x), \text{ for } \alpha > 0$$

3 Find the constant K so that the following is a p.d.f.

$$f(x) = Kx^2 I_{(-K,K)}(x).$$

4 Suppose that the cumulative distribution function (c.d.f.) $F_X(x)$ can be written as a function of $(x - \alpha)/\beta$, where α and $\beta > 0$ are constants; that is, x, α, and β appear in $F_X(\cdot)$ only in the indicated form.

(*a*) Prove that if α is increased by $\Delta\alpha$, then so is the mean of X.

(*b*) Prove that if β is multiplied by $k (k > 0)$, then so is the standard deviation of X.

5 The experiment is to toss two balls into four boxes in such a way that each ball is equally likely to fall in any box. Let X denote the number of balls in the first box.

(*a*) What is the c.d.f. of X?

(*b*) What is the density function of X?

(*c*) Find the mean and variance of X.

6 A fair coin is tossed until a head appears. Let X denote the number of tosses required.

(*a*) Find the density function of X.

(*b*) Find the mean and variance of X.

(*c*) Find the moment generating function (m.g.f.) of X.

**7* A has two pennies; B has one. They match pennies until one of them has all three. Let X denote the number of trials required to end the game.

(*a*) What is the density function of X?

(*b*) Find the mean and variance of X.

(*c*) What is the probability that B wins the game?

8 Let $f_X(x) = (1/\beta)[1 - |(x - \alpha)/\beta|]I_{(\alpha-\beta,\, \alpha+\beta)}(x)$, where α and β are fixed constants satisfying $-\infty < \alpha < \infty$ and $\beta > 0$.

(*a*) Demonstrate that $f_X(\cdot)$ is a p.d.f., and sketch it.

(*b*) Find the c.d.f. corresponding to $f_X(\cdot)$.

(*c*) Find the mean and variance of X.

(*d*) Find the qth quantile of X.

9 Let $f_X(x) = k(1/\beta)\{1 - [(x - \alpha)/\beta]^2\}I_{(\alpha-\beta,\, \alpha+\beta)}(x)$, where $-\infty < \alpha < \infty$ and $\beta > 0$.

(*a*) Find k so that $f_X(\cdot)$ is a p.d.f., and sketch the p.d.f.

(*b*) Find the mean, median, and variance of X.

(*c*) Find $\mathscr{E}[|X - \alpha|]$.

(*d*) Find the qth quantile of X.

10 Let $f_X(x) = \frac{1}{2}\{\theta I_{(0,\, 1)}(x) + I_{[1,\, 2]}(x) + (1 - \theta)I_{(2,3)}(x)\}$, where θ is a fixed constant satisfying $0 \leq \theta \leq 1$.

(*a*) Find the c.d.f. of X.

(*b*) Find the mean, median, and variance of X.

11 Let $f(x;\, \theta) = \theta f(x;\, 1) + (1 - \theta)f(x;\, 0)$, where θ is a fixed constant satisfying $0 \leq \theta \leq 1$. Assume that $f(\cdot\,; 0)$ and $f(\cdot\,; 1)$ are both p.d.f.'s.

(*a*) Show that $f(\cdot\,; \theta)$ is also a p.d.f.

(*b*) Find the mean and variance of $f(\cdot\,; \theta)$ in terms of the mean and variance of $f(\cdot\,; 0)$ and $f(\cdot\,; 1)$, respectively.

(*c*) Find the m.g.f. of $f(\cdot\,; \theta)$ in terms of the m.g.f.'s of $f(\cdot\,; 0)$ and $f(\cdot\,; 1)$.

12 A bombing plane flies directly above a railroad track. Assume that if a large (small) bomb falls within 40 (15) feet of the track, the track will be sufficiently damaged so that traffic will be disrupted. Let X denote the perpendicular distance from the track that a bomb falls. Assume that

$$f_X(x) = \frac{100 - x}{5000} I_{[0,100)}(x).$$

(*a*) Find the probability that a large bomb will disrupt traffic.
(*b*) If the plane can carry three large (eight small) bombs and uses all three (eight), what is the probability that traffic will be disrupted?

13 (*a*) Let X be a random variable with mean μ and variance σ^2. Show that $\mathscr{E}[(X-b)^2]$, as a function of b, is minimized when $b = \mu$.
*(*b*) Let X be a continuous random variable with median m. Minimize $\mathscr{E}[|X-b|]$ as a function of b. HINT: Show that $\mathscr{E}[|X-b|] = \mathscr{E}[|X-m|] + 2\int_b^m (x-b)f_X(x)\,dx$.

14 (*a*) If X is a random variable such that $\mathscr{E}[X] = 3$ and $\mathscr{E}[X^2] = 13$, use the Chebyshev inequality to determine a lower bound for $P[-2 < X < 8]$.
(*b*) Let X be a discrete random variable with density

$$f_X(x) = \tfrac{1}{8}I_{\{-1\}}(x) + \tfrac{6}{8}I_{\{0\}}(x) + \tfrac{1}{8}I_{\{1\}}(x).$$

For $k = 2$ evaluate $P[|X - \mu_X| \ge k\sigma_X]$. (This shows that in general the Chebyshev inequality cannot be improved.)
(*c*) If X is a random variable with $\mathscr{E}[X] = \mu$ satisfying $P[X \le 0] = 0$, show that $P[X > 2\mu] \le \frac{1}{2}$.

15 Let X be a random variable with p.d.f. given by

$$f_X(x) = |1 - x| I_{[0,2]}(x).$$

Find the mean and variance of X.

16 Let X be a random variable having c.d.f.

$$F_X(x) = pH(x) + (1-p)G(x),$$

where p is a fixed real number satisfying $0 < p < 1$,

$$H(x) = xI_{[0,1)}(x) + I_{[1,\infty)}(x),$$

and

$$G(x) = \tfrac{1}{2}xI_{[0,2)}(x) + I_{[2,\infty)}(x).$$

(*a*) Sketch $F_X(x)$ for $p = \frac{1}{2}$.
(*b*) Give a formula for the p.d.f. of X or the discrete density function of X, whichever is appropriate.
(*c*) Evaluate $P[X \le \frac{1}{2} | X \le 1]$.

17 Does there exist a random variable X for which $P[\mu_X - 2\sigma_X \le X \le \mu_X + 2\sigma_X] = .6$?

18 An urn contains balls numbered 1, 2, 3. First a ball is drawn from the urn, and then a fair coin is tossed the number of times as the number shown on the drawn ball. Find the expected number of heads.

19 If X has distribution given by $P[X=0]=P[X=2]=p$ and $P[X=1]=1-2p$ for $0 \le p \le \frac{1}{2}$, for what p is the variance of X a maximum?

20 If X is a random variable for which $P[X \le 0]=0$ and $\mathscr{E}[X]=\mu<\infty$, prove that $P[X \le \mu t] \ge 1-1/t$ for every $t \ge 1$.

21 Given the c.d.f.

$$\begin{aligned} F_X(x) &= 0 && \text{for } x<0 \\ &= x^2+.2 && \text{for } 0 \le x < .5 \\ &= x && \text{for } .5 \le x < 1 \\ &= 1 && \text{for } 1 \le x. \end{aligned}$$

(*a*) Express $F_X(x)$ in terms of indicator functions.

(*b*) Express $F_X(x)$ in the form

$$aF^{ac}(x)+bF^{d}(x),$$

where $F^{ac}(\cdot)$ is an absolutely continuous c.d.f. and $F^{d}(\cdot)$ is a discrete c.d.f.

(*c*) Find $P[.25<X<.75]$.

(*d*) Find $P[.25<X<.5]$.

22 Let $f(x)=Ke^{-\alpha x}(1-e^{-\alpha x})I_{(0,\infty)}(x)$.

(*a*) Find K such that $f(\cdot)$ is a density function.

(*b*) Find the corresponding c.d.f.

(*c*) Find $P[X>1]$.

23 A coin is tossed four times. Let X denote the number of times a head is followed immediately by a tail. Find the distribution, mean, and variance of X.

24 Let $f_X(x;\theta)=(\theta x+\frac{1}{2})I_{(-1,1)}(x)$, where θ is a constant.

(*a*) For what range of values of θ is $f_X(\cdot\,;\theta)$ a density function?

(*b*) Find the mean and median of X.

(*c*) For what values of θ is var $[X]$ maximized?

25 Let X be a discrete random variable with the nonnegative integers as values. Note that $\mathscr{E}[t^X]=\sum_{j=0}^{\infty} t^j P[X=j]$. Hence, $\mathscr{E}[t^X]$ is a **probability generating function** of X, inasmuch as the coefficient of t^j gives $P[X=j]$. Find $\mathscr{E}[t^X]$ for the random variable of Probs. 6 and 7.