2

Random Variables

2.1 INTRODUCTION

In Chapter 1 we mentioned that there are situations in which not all subsets of the sample space Ω can belong to the event class $\mathscr{F}$, and that difficulties of this type generally arise when Ω is uncountable. Such spaces may arise physically as approximations to discrete spaces with a very large number of points. For example, if a person is picked at random in the United States and his age recorded, a complete description of this experiment would involve a probability space with approximately 200 million points (if the data are recorded accurately enough, no two people have the same age). A more convenient way to describe the experiment is to group the data, for example, into 10-year intervals. We may define a function $q(x)$, $x = 5, 15, 25, \ldots$, so that $q(x)$ is the number of people, say in millions, between $x - 5$ and $x + 5$ years (see Figure 2.1.1).

For example, if $q(15) = 40$, there are 40 million people between the ages of 10 and 20 or, on the average, 4 million per year over that 10-year span. Now if we want the probability that a person picked at random will be between 14 and 16, we can get a reasonable figure by taking the average number of people per year [$4 = q(15)/10$)] and multiplying by the number of years (2) to obtain (roughly) 8 million people, then dividing by the total population to obtain a probability of $8/200 = .04$.

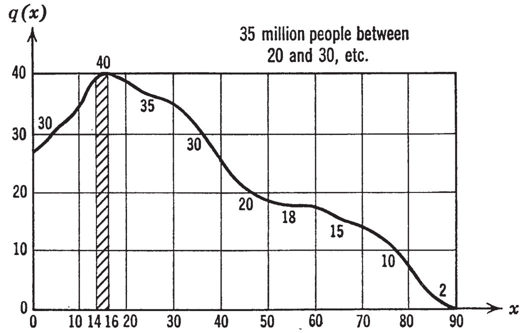


FIGURE 2.1.1 Age Statistics.

If we connect the values of $q(x)$ by a smooth curve, essentially what we are doing is evaluating $(1/200)\int_{14}^{16}[q(x)/10]\,dx$ to find the probability that a person picked at random will be between 14 and 16 years old. In general, we estimate the number of people between ages a and b by $\int_a^b [q(x)/10]\,dx$ so that $q(x)/10$ is the age density, that is, the number of people per unit age. We estimate the probability of obtaining an age between a and b by $\int_a^b [q(x)/2000]\,dx$; thus $q(x)/2000$ is the probability density, or probability per unit age. Thus we are led to the idea of assigning probabilities by means of an integral. We are taking Ω as (a subset of) the reals, and assigning $P(B) = \int_B f(x)\,dx$, where f is a real-valued function defined on the reals. There are several immediate questions, namely, what sigma field we are using, what functions f are allowed, what we mean by $\int_B f(x)\,dx$, and how we know that the resulting P is a probability.

For the moment suppose that we restrict ourselves to continuous or piecewise continuous f. Then we can certainly talk about $\int_B f(x)\,dx$, at least when B is an interval, and the integral is in the Riemann sense. Thus the appropriate sigma field $\mathscr{F}$ should contain the intervals, and hence must be at least as big as the smallest sigma field $\mathscr{B}$ containing the intervals ($\mathscr{B}$ exists; it can be described as the intersection of all sigma fields containing the intervals). The sigma field $\mathscr{B} = \mathscr{B}(E^1)$ is called the class of *Borel sets* of the reals E^1. Intuitively we may think of $\mathscr{B}$ being generated by starting with the intervals and repeatedly forming new sets by taking countable unions (and countable intersections) and complements in all possible ways (it turns out that there are subsets of E^1 that are not Borel sets).

Thus our problem will be to construct probability measures on the class of Borel sets of E^1. The reason for considering only the Borel sets rather than all subsets of E^1 is this. Suppose that we require that $P(B) = \int_B f(x)\,dx$

for all intervals B, where f is a particular nonnegative continuous function defined on E^1, and $\int_{-\infty}^{\infty} f(x)\, dx = 1$. There is no probability measure on the class of all subsets of E^1 satisfying this requirement, but there is such a measure on the Borel sets.

Before elaborating on these ideas, it is convenient to introduce the concept of a random variable; we do this in the next section.

2.2 DEFINITION OF A RANDOM VARIABLE

Intuitively, a random variable is a quantity that is measured in connection with a random experiment. If Ω is a sample space, and the outcome of the experiment is ω, a measuring process is carried out to obtain a number $R(\omega)$. Thus *a random variable is a real-valued function on a sample space*. (The formal definition, which is postponed until later in the section, is somewhat more restrictive.)

▶ **Example 1.** Throw a coin 10 times, and let R be the number of heads. We take $\Omega =$ all sequences of length 10 with components H and T; 2^{10} points altogether. A typical sample point is $\omega = HHTHTTHHTH$. For this point $R(\omega) = 6$. Another random variable, R_1, is the number of times a head is followed immediately by a tail. For the point ω above, $R_1(\omega) = 3$. ◀

▶ **Example 2.** Pick a person at random from a certain population and measure his height and weight. We may take the sample space to be the plane E^2, that is, the set of all pairs (x, y) of real numbers, with the first coordinate x representing the height and the second coordinate y the weight (we can take care of the requirement that height and weight be nonnegative by assigning probability 0 to the complement of the first quadrant). Let R_1 be the height of the person selected, and let R_2 be the weight. Then $R_1(x, y) = x$, $R_2(x, y) = y$. As another example, let R_3 be twice the height plus the cube root of the weight; that is, $R_3 = 2R_1 + \sqrt[3]{R_2}$. Then $R_3(x, y) = 2R_1(x, y) + \sqrt[3]{R_2(x, y)} = 2x + \sqrt[3]{y}$. ◀

▶ **Example 3.** Throw two dice. We may take the sample space to be the set of all pairs of integers (x, y), $x, y = 1, 2, \ldots, 6$ (36 points in all).

Let $R_1 =$ the result of the first toss. Then $R_1(x, y) = x$.

Let $R_2 =$ the sum of the two faces. Then $R_2(x, y) = x + y$.

Let $R_3 = 1$ if at least one face is an even number; $R_3 = 0$ otherwise.

Then $R_3(6, 5) = 1$; $R_3(3, 6) = 1$; $R_3(1, 3) = 0$, and so on. ◀

▶ **Example 4.** Imagine that we can observe the times at which electrons are emitted from the cathode of a vacuum tube, starting at time $t = 0$. As a sample space, we may take all infinite sequences of positive real numbers, with the components representing the emission times. Assume that the emission process never stops. Typical sample points might be $\omega_1 = (.2, 1.5, 6.3, \ldots)$, $\omega_2 = (.01, .5, .9, 1.7, \ldots)$. If R_1 is the number of electrons emitted before $t = 1$, then $R_1(\omega_1) = 1$, $R_1(\omega_2) = 3$. If R_2 is the time at which the first electron is emitted, then $R_2(\omega_1) = .2$, $R_2(\omega_2) = .01$. ◀

If we are interested in a random variable R defined on a given sample space, we generally want to know the probability of events involving R. Physical measurements of a quantity R generally lead to statements of the form $a \le R \le b$, and it is natural to ask for the probability that R will lie between a and b in a given performance of the experiment. Thus we are looking for $P\{\omega: a \le R(\omega) \le b\}$ (or, equally well, $P\{\omega: a < R(\omega) \le b\}$, and so on). For example, if a coin is tossed independently n times, with probability p of coming up heads on a given toss, and if R is the number of heads, we have seen in Chapter 1 that

$$P\{\omega: a \le R(\omega) \le b\} = \sum_{k=a}^{b} \binom{n}{k} p^k (1-p)^{n-k}$$

Notation. $\{\omega: a \le R(\omega) \le b\}$ will often be abbreviated to $\{a \le R \le b\}$.

As another example, if two unbiased dice are tossed independently, and R_2 is the sum of the faces (Example 3 above), then $P\{R_2 = 6\} = P\{(5, 1), (1, 5), (4, 2), (2, 4), (3, 3)\} = 5/36$.

In general an "event involving R" corresponds to a statement that the value of R lies in a set B; that is, the event is of the form $\{\omega: R(\omega) \in B\}$. Intuitively, if $P\{\omega: R(\omega) \in I\}$ is known for all intervals I, then $P\{\omega: R(\omega) \in B\}$ is determined for any "well-behaved" set B, the reason being that any such set can be built up from intervals. For example, $P\{0 \le R < 2$ or $R > 3\}$ $(= P\{R \in [0, 2) \cup (3, \infty)\}) = P\{0 \le R < 2\} + P\{R > 3\}$. Thus it appears that in order to describe the nature of R completely, it is sufficient to know $P\{R \in I\}$ for each interval I. We consider in more detail the problem of characterizing a random variable in the next section; in the remainder of this section we give the formal definition of a random variable.

* For the concept of random variable to fit in with our established model for a probability space, the sets $\{a \le R \le b\}$ must be events; that is, they must belong to the sigma field $\mathscr{F}$. Thus a first restriction on R is that for all real a, b, the sets $\{\omega: a \le R(\omega) \le b\}$ are in $\mathscr{F}$. Thus we can talk intelligently about the event that R lies between a and b.

A question now comes up: Suppose that the sets $\{a \le R \le b\}$ are in $\mathscr{F}$

for all a, b. Can we talk about the event that R belongs to a set B of reals, for B more general than a closed interval?

For example, let $B = [a, b)$ be an interval closed on the left, open on the right. Then

$$a \leq R(\omega) < b \text{ iff } a \leq R(\omega) \leq b - \frac{1}{n} \qquad \text{for at least one } n = 1, 2, \ldots$$

Thus

$$\{\omega: a \leq R(\omega) < b\} = \bigcup_{n=1}^{\infty} \left\{\omega: a \leq R(\omega) \leq b - \frac{1}{n}\right\}$$

and this set is a countable union of sets in $\mathscr{F}$, hence belongs to $\mathscr{F}$. In a similar fashion we can handle all types of intervals. Thus $\{\omega: R(\omega) \in B\} \in \mathscr{F}$ for all intervals B.

In fact $\{\omega: R(\omega) \in B\}$ belongs to $\mathscr{F}$ for all Borel sets B. The sequence of steps by which this is proved is outlined in Problem 1.

We are now ready for the formal definition.

DEFINITION. A *random variable* on the probability space $(\Omega, \mathscr{F}, P)$ is a real valued function R defined on Ω, such that for every Borel subset B of the reals, $\{\omega: R(\omega) \in B\}$ belongs to $\mathscr{F}$.

Notice that the probability P is not involved in the definition at all; if R is a random variable on $(\Omega, \mathscr{F}, P)$ and the probability measure is changed, R is still a random variable. Notice also that, by the above discussion, to check whether a given function R is a random variable it is sufficient to know that $\{\omega: a \leq R(\omega) \leq b\} \in \mathscr{F}$ for all real a, b. In fact (Problem 2) it is sufficient that $\{\omega: R(\omega) < b\} \in \mathscr{F}$ for all real b (or, equally well, $\{\omega: R(\omega) \leq b\} \in \mathscr{F}$ for all real b; or $\{\omega: R(\omega) > a\} \in \mathscr{F}$ for all real a; or $\{\omega: R(\omega) \geq a\} \in \mathscr{F}$ for all real a; the argument is essentially the same in all cases).

Notice that if $\mathscr{F}$ consists of all subsets of Ω, $\{\omega: R(\omega) \in B\}$ automatically belongs to $\mathscr{F}$, so that in this case any real-valued function on the sample space is a random variable. Examples 1 and 3 fall into this category.

Now let us consider Example 2. We take Ω = the plane E^2, $\mathscr{F}$ = the class of Borel subsets of E^2, that is, the smallest sigma field containing all rectangles (we shall use "rectangle" in a very broad sense, allowing open, closed, or semiclosed rectangles, as well as infinite rectangular strips).

To check that R_1 is a random variable, we have

$$\{(x, y): a \leq R_1(x, y) \leq b\} = \{(x, y): a \leq x \leq b\}$$

which is a rectangular strip and hence a set in $\mathscr{F}$. Similarly, R_2 is a random variable. For R_3, see Problem 3.

Example 2 generalizes as follows. Take $\Omega = E^n$ = all n-tuples of real numbers, $\mathscr{F}$ the smallest sigma field containing the n-dimensional "intervals." [If $a = (a_1, \ldots, a_n)$, $b = (b_1, \ldots, b_n)$, the interval (a, b) is defined as $\{x \in E^n\colon a_i < x_i < b_i,\ i = 1, \ldots, n\}$; closed and semiclosed intervals are defined similarly.] The coordinate functions, given by $R_1(x_1, \ldots, x_n) = x_1$, $R_2(x_1, \ldots, x_n) = x_2, \ldots, R_n(x_1, \ldots, x_n) = x_n$, are random variables.

Example 4 involves some serious complications, since the sample points are infinite sequences of real numbers. We postpone the discussion of situations of this type until much later (Chapter 6).

PROBLEMS

*1. Let R be a real-valued function on a sample space Ω, and let $\mathscr{C}$ be the collection of all subsets B of E^1 such that $\{\omega\colon R(\omega) \in B\} \in \mathscr{F}$.
   (a) Show that $\mathscr{C}$ is a sigma field.
   (b) If all intervals belong to $\mathscr{C}$, that is, if $\{\omega\colon R(\omega) \in B\} \in \mathscr{F}$ when B is an interval, show that all Borel sets belong to $\mathscr{C}$. Conclude that R is a random variable.

*2. Let R be a real-valued function on a sample space Ω, and assume $\{\omega\colon R(\omega) < b\} \in \mathscr{F}$ for all real b. Show that R is a random variable.

*3. In Example 2, show that R_3 is a random variable. Do this by showing that if R_1 and R_2 are random variables, so is $R_1 + R_2$; if R is a random variable, so is aR for any real a; if R is a random variable, so is $\sqrt[3]{R}$.

2.3 CLASSIFICATION OF RANDOM VARIABLES

If R is a random variable on the probability space $(\Omega, \mathscr{F}, P)$, we are generally interested in calculating probabilities of events involving R, that is, $P\{\omega\colon R(\omega) \in B\}$ for various (Borel) sets B. The way in which these probabilities are calculated will depend on the particular nature of R; in this section we examine some standard classes of random variables.

The random variable R is said to be *discrete* iff the set of possible values of R is finite or countably infinite. In this case, if $x_1, x_2, \ldots$ are the values of R that belong to B, then

$$\begin{aligned} P\{R \in B\} &= P\{R = x_1 \quad \text{or} \quad R = x_2 \quad \text{or} \quad \cdots\} \\ &= P\{R = x_1\} + P\{R = x_2\} + \cdots = \sum_{x \in B} p_R(x) \end{aligned}$$

where $p_R(x)$, x real, is the *probability function* of R, defined by $p_R(x) = P\{R = x\}$. Thus the probability of an event involving R is found by summing the probability function over the set of points favorable to the event. In particular, the probability function determines the probability of all events involving R.

▶ **Example 1.** Let R be the number of heads in two independent tosses of a coin, with the probability of heads being .6 on a given toss. Take $\Omega = \{HH, HT, TH, TT\}$ with probabilities .36, .24, .24, .16 assigned to the four points of Ω; take $\mathscr{F}$ = all subsets. Then R has three possible values, namely, 0, 1, and 2, and $P\{R = 0\} = .16$, $P\{R = 1\} = .48$, $P\{R = 2\} = .36$, by inspection or by using the binomial formula

$$P\{R = k\} = \binom{n}{k} p^k(1 - p)^{n-k} \quad ◀$$

Another way of characterizing R is by means of the *distribution function*, defined by

$$F_R(x) = P\{R \le x\}, \qquad x \text{ real}$$

(see Figure 2.3.1 for a sketch of F_R and p_R in Example 1).

Observe that, for example, $P\{R \le 1\} = p_R(0) + p_R(1) = .64$, but if $0 < x < 1$, we have $P\{R \le x\} = p_R(0) = .16$. Thus F_R has a discontinuity at $x = 1$, of magnitude $.48 = p_R(1)$. In general, if R is discrete, and $P\{R = x_n\} = p_n$, $n = 1, 2, \ldots$, where the p_n are >0 and $\sum_n p_n = 1$, then F_R has a jump of magnitude p_n at $x = x_n$; F_R is constant between jumps.

In the discrete case, if we are given the probability function, we can construct the distribution function, and, conversely, given F_R, we can construct

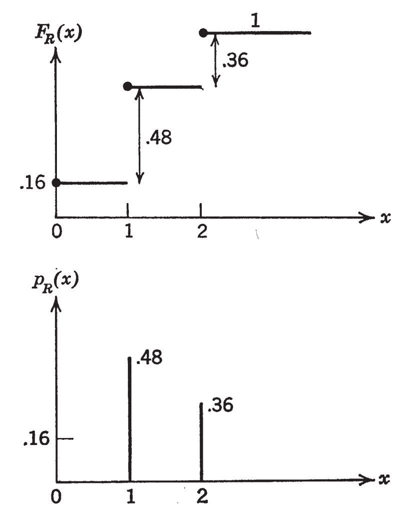


FIGURE 2.3.1 Distribution and Probability Functions of a Discrete Random Variable.

p_R. Knowledge of either function is sufficient to determine the probability of all events involving R.

We now consider the case introduced in Section 2.1, where probabilities are assigned by means of an integral.

Let f be a nonnegative Riemann integrable† function defined on E^1 with $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Take $\Omega = E^1$, $\mathscr{F}$ = Borel sets. We would like to write, for each $B \in \mathscr{F}$,

$$P(B) = \int_B f(x)\,dx$$

but this makes sense only if B is an interval. However, the following result is applicable.

Theorem 1. *Let f be a nonnegative real-valued function on E^1, with $\int_{-\infty}^{\infty} f(x)\,dx = 1$. There is a unique probability measure P defined on the Borel subsets of E^1, such that $P(B) = \int_B f(x)\,dx$ for all intervals $B = (a, b]$.*

The theorem belongs to the domain of measure and integration theory, and will not be proved here.

The theorem allows us to talk about the integral of f over an arbitrary Borel set B. We simply define $\int_B f(x)\,dx$ as $P(B)$, where P is the probability measure given by the theorem.

The uniqueness part of the theorem may then be phrased as follows. If Q is a probability measure on the Borel subsets of E^1 and $Q(B) = \int_B f(x)\,dx$ for all intervals $B = (a, b]$, then $Q(B) = \int_B f(x)\,dx$ for all Borel sets B.

If R is defined on Ω by $R(\omega) = \omega$ (so that the outcome of the experiment is identified with the value of R), then

$$P\{\omega: R(\omega) \in B\} = P(B) = \int_B f(x)\,dx$$

In particular, the distribution function of R is given by

$$F_R(x) = P\{\omega: R(\omega) \le x\} = P(-\infty, x] = \int_{-\infty}^{x} f(t)\,dt$$

so that F_R is represented as an integral.

DEFINITION. The random variable R is said to be *absolutely continuous* iff there is a nonnegative function $f = f_R$ defined on E^1 such that

$$F_R(x) = \int_{-\infty}^{x} f_R(t)\,dt \qquad \text{for all real } x \tag{2.3.1}$$

f_R is called the *density function* of R. We shall see in Section 2.5 that $F_R(x)$ must approach 1 as $x \to \infty$; hence $\int_{-\infty}^{\infty} f_R(x)\,dx = 1$.

† "Integrable" will from now on mean "Riemann integrable."

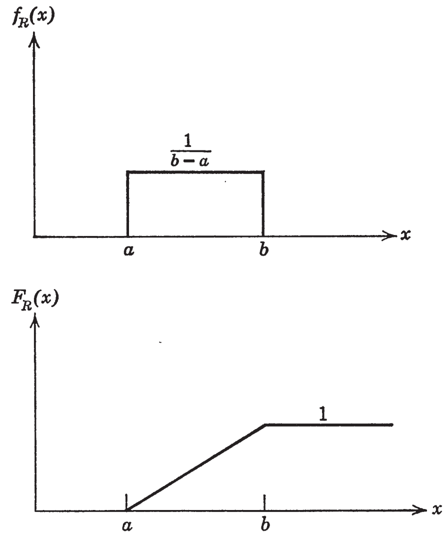


FIGURE 2.3.2 Distribution and Density Functions of a Uniformly Distributed Random Variable.

▶ **Example 2.** A number R is chosen at random between a and b; R is assumed to be *uniformly distributed*; that is, the probability that R will fall into an interval of length c depends only on c, not on the position of the interval within $[a, b]$.

We take $\Omega = E^1$, $\mathscr{F}$ = Borel sets, $R(\omega) = \omega, f(x) = f_R(x) = 1/(b-a)$, $a \le x \le b; f(x) = 0$, $x > b$ or $x < a$. Define $P(B) = \int_B f(x)\,dx$. In particular, if B is a subinterval of $[a, b]$, then $P(B) =$ (length of $B)/(b-a)$. The density and distribution function of R are shown in Figure 2.3.2. ◀

NOTE. The values of F_R are *probabilities*, but the values of f_R are not; probabilities are found by *integrating* f_R.

$$F_R(x) = P\{R \le x\} = \int_{-\infty}^{x} f_R(t)\,dt$$

If R is absolutely continuous, then

$$P\{a < R \le b\} = \int_a^b f_R(x)\,dx, \qquad a < b$$

For $\{R \le b\}$ is the disjoint union of the events $\{R \le a\}$ and $\{a < R \le b\}$; hence $P\{R \le b\} = P\{R \le a\} + P\{a < R \le b\}$. It follows that

$$P\{a < R \leq b\} = F_R(b) - F_R(a) \tag{2.3.2}$$
$$= \int_{-\infty}^{b} f_R(x)\,dx - \int_{-\infty}^{a} f_R(x)\,dx = \int_{a}^{b} f_R(x)\,dx$$

Thus, if $Q(B) = P\{R \in B\}$, we have $Q(B) = \int_B f_R(x)\,dx$ when B is an interval $(a, b]$. By Theorem 1, $Q(B) = \int_B f_R(x)\,dx$ for all Borel sets B. Therefore, *if R is absolutely continuous,*

$$P\{R \in B\} = \int_B f_R(x)\,dx \qquad \text{for all Borel sets } B$$

The basic point is that the *density function f_R determines the probability of all events involving R.*

If R is absolutely continuous, then

$$P\{R = c\} = P\{c \leq R \leq c\} = \int_c^c f_R(x)\,dx = 0$$

The event $\{R = c\}$ is in general *not impossible*; for example, if R is uniformly distributed between a and b, each event $\{R = x\}$, $a \leq x \leq b$, is possible; that is, the set $\{\omega: R(\omega) = x\}$ is not empty. But the event $\{R = c\}$ has probability 0. This does not contradict the axioms of probability. The definition of a probability measure requires that if the event A is impossible (i.e., $A = \varnothing$) then $P(A) = 0$; the *converse* need not be true. Intuitively, if R is uniformly distributed between a and b, it should be expected that all events $\{R = x\}$, $a \leq x \leq b$, will have the same probability. Any probability other than 0 will lead to a contradiction, since there are infinitely many points x between a and b.

As a consequence of the fact that $P\{R = x\} = 0$ in the absolutely continuous case, we have

$$\begin{aligned} P\{a \leq R \leq b\} &= P\{a < R \leq b\} = P\{a \leq R < b\} \\ &= P\{a < R < b\} \\ &= \int_a^b f_R(x)\,dx \\ &= F_R(b) - F_R(a) \end{aligned} \tag{2.3.3}$$

Notice also that although in the discrete case the probability function of R determines the probability of all events involving R, in the absolutely continuous case it gives no information at all, since $p_R(x) = P\{R = x\} = 0$ for all x. However, the distribution function of R is still adequate, since F_R determines f_R. If f_R is continuous, it may be obtained from F_R by differentiation; that is,

$$\frac{d}{dx}\int_{-\infty}^{x} f_R(t)\,dt = f_R(x)$$

(the fundamental theorem of calculus). The general proof that F_R determines f_R is measure-theoretic, and we shall not pursue it here.

If f_R is continuous, we have just seen that F_R is differentiable, and its derivative is f_R. In general, if R is absolutely continuous, $F_R(x)$ will be a continuous function of x, but again we shall not pursue this.

We shall show in Section 2.5 that the distribution function of an *arbitrary* random variable must be nondecreasing [$a < b$ implies $F_R(a) \le F_R(b)$], must approach 1 as $x \to \infty$, and must approach 0 as $x \to -\infty$.

▶ **Example 3.** Let R be time of emission of the first electron from the cathode of a vacuum tube. Under certain physical assumptions, it turns out that R has the following density function:

$$\begin{aligned} f_R(x) &= \lambda e^{-\lambda x}, & x &\ge 0 \\ &= 0 & x &< 0 \end{aligned} \qquad (\lambda \text{ constant})$$

(see Figure 2.3.3).

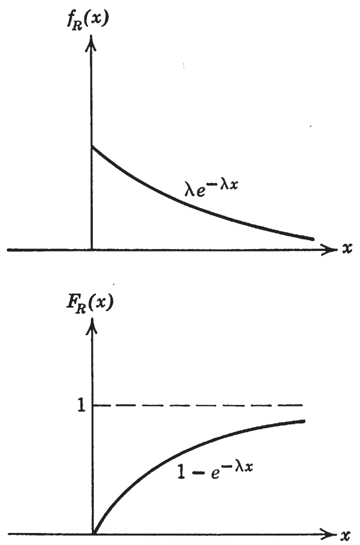


FIGURE 2.3.3 Exponential Density and Distribution Functions.

$$F_R(x) = \int_{-\infty}^{x} f_R(t)\,dt$$

so $F_R(x) = 0$, $x < 0$,

and if $x \ge 0$,

$$F_R(x) = \int_{-\infty}^{0} f_R(t)\,dt + \int_0^x f_R(t)\,dt = \int_0^x \lambda e^{-\lambda t}\,dt = 1 - e^{-\lambda x}$$

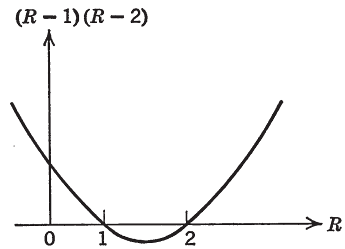


FIGURE 2.3.4 Calculation of Probabilities.

We calculate some probabilities of events involving R:

$$P\{1 \le R \le 2\} = \int_1^2 \lambda e^{-\lambda x}\, dx = e^{-\lambda} - e^{-2\lambda} = F_R(2) - F_R(1)$$

$$\begin{aligned} P\{(R-1)(R-2) \ge 0\} &= P\{R \le 1 \quad \text{or} \quad R \ge 2\} \\ &= P\{R \le 1\} + P\{R \ge 2\} \\ &= \int_0^1 \lambda e^{-\lambda x}\, dx + \int_2^\infty \lambda e^{-\lambda x}\, dx \\ &= 1 - e^{-\lambda} + e^{-2\lambda} \end{aligned}$$

(see Figure 2.3.4). ◀

REMARK. You will often see the statement "Let R be an absolutely continuous random variable with density function f," with no reference made to the underlying probability space. However, we have seen that we can always supply an appropriate space, as follows. Take $\Omega = E^1$, $\mathscr{F}$ = Borel sets, $P(B) = \int_B f(x)\, dx$ for all $B \in \mathscr{F}$. If $R(\omega) = \omega$, $\omega \in \Omega$, then R is absolutely continuous and has density f.

In a sense, it does not make any difference how we arrive at Ω and P; we may equally well use a different Ω and P and a different R, as long as R is absolutely continuous with density f. No matter what construction we use, we get the same essential result, namely,

$$P\{R \in B\} = \int_B f(x)\, dx$$

Thus questions about probabilities of events involving R are answered completely by knowledge of the density f.

PROBLEMS

1. An absolutely continuous random variable R has a density function $f(x) = (1/2)e^{-|x|}$.
   (a) Sketch the distribution function of R.
   (b) Find the probability of each of the following events.

   (1) $\{|R| \leq 2\}$
   (2) $\{|R| \leq 2 \text{ or } R \geq 0\}$
   (3) $\{|R| \leq 2 \text{ and } R \leq -1\}$
   (4) $\{|R| + |R - 3| \leq 3\}$
   (5) $\{R^3 - R^2 - R - 2 \leq 0\}$
   (6) $\{e^{\sin \pi R} \geq 1\}$
   (7) $\{R \text{ is irrational}\}$ $(= \{\omega\colon R(\omega)$ is an irrational number$\})$

2. Consider a sequence of five Bernoulli trials. Let R be the number of times that a head is followed immediately by a tail. For example, if $\omega = HHTHT$ then $R(\omega) = 2$, since a head is followed directly by a tail at trials 2 and 3, and also at trials 4 and 5. Find the probability function of R.

2.4 FUNCTIONS OF A RANDOM VARIABLE

A general problem that arises in many branches of science is the following. Given a system of some sort, to which an input is applied; knowledge of some of the characteristics of the system, together with knowledge of the input, will allow some estimate of the behavior at the output. We formulate a special case of this problem. Given a random variable R_1 on a probability space, and a real-valued function g on the reals, we define a random variable R_2 by $R_2 = g(R_1)$; that is, $R_2(\omega) = g(R_1(\omega))$, $\omega \in \Omega$. R_1 plays the role of the input, and g the role of the system; the output R_2 is a random variable defined on the same space as R_1. Given the function g and the distribution or density function of the random variable R_1, the problem is to find the distribution or density function of R_2.

NOTE. If R_1 is a random variable and we set $R_2 = g(R_1)$, the question arises as to whether R_2 is in fact a random variable. The answer is yes if g is continuous or piecewise continuous; we shall consider this problem in greater detail in Section 2.7.

▶ **Example 1.** Let R_1 be absolutely continuous, with the density f_1 given in Figure 2.4.1. Let $R_2 = R_1^2$; that is, $R_2(\omega) = R_1^2(\omega)$, $\omega \in \Omega$. Find the distribution or density function of R_2.

We shall indicate two approaches to the problem.

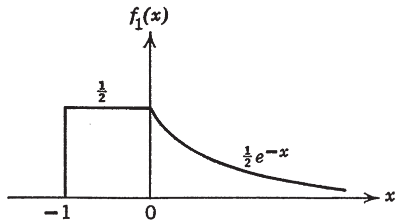


FIGURE 2.4.1

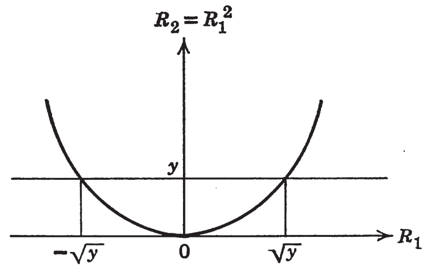


FIGURE 2.4.2

DISTRIBUTION FUNCTION METHOD. In this method the distribution function F_2 of R_2 is found directly, by expressing the event $\{R_2 \le y\}$ in terms of the random variable R_1. First, since $R_2 \ge 0$, we have $F_2(y) = P\{R_2 \le y\} = 0$ for $y < 0$.

If $y \ge 0$, then $R_2 \le y$ iff $-\sqrt{y} \le R_1 \le \sqrt{y}$ (see Figure 2.4.2). Thus, if $y \ge 0$,

$$P\{R_2 \le y\} = P\{-\sqrt{y} \le R_1 \le \sqrt{y}\} = \int_{-\sqrt{y}}^{\sqrt{y}} f_1(x)\,dx$$

In particular, if $0 \le y \le 1$, then

$$F_2(y) = \int_{-\sqrt{y}}^{\sqrt{y}} f_1(x)\,dx = \int_{-\sqrt{y}}^{0} \tfrac{1}{2}\,dx + \int_0^{\sqrt{y}} \tfrac{1}{2}e^{-x}\,dx = \tfrac{1}{2}\sqrt{y} + \tfrac{1}{2}(1 - e^{-\sqrt{y}})$$

(see Figure 2.4.3).

If $y > 1$,

$$F_2(y) = \int_{-\sqrt{y}}^{\sqrt{y}} f_1(x)\,dx = \int_{-\sqrt{y}}^{-1} 0\,dx + \int_{-1}^{0} \tfrac{1}{2}\,dx + \int_0^{\sqrt{y}} \tfrac{1}{2}e^{-x}\,dx$$
$$= \tfrac{1}{2} + \tfrac{1}{2}(1 - e^{-\sqrt{y}})$$

(see Figure 2.4.4). A sketch of F_2 is given in Figure 2.4.5.

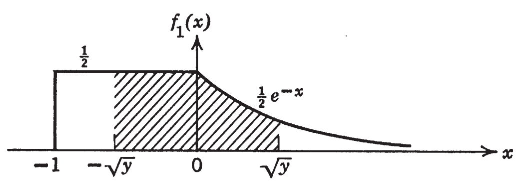


FIGURE 2.4.3

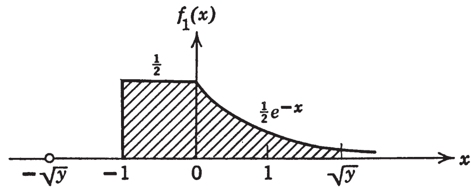


FIGURE 2.4.4

We would like to conclude, by inspection of the distribution function F_2, that the random variable R_2 is absolutely continuous. We should be able to find the density f_2 of R_2 by differentiating F_2.

$$
\begin{aligned}
f_2(y) = \frac{dF_2(y)}{dy} &= 0, \qquad y < 0 \\
&= \frac{1}{4\sqrt{y}}\,(1 + e^{-\sqrt{y}}), \qquad 0 < y < 1 \\
&= \frac{1}{4\sqrt{y}}\,e^{-\sqrt{y}}, \qquad y > 1
\end{aligned}
$$

(see Figure 2.4.6).

It may be verified that $F_2(y)$ is given by $\int_{-\infty}^{y} f_2(t)\,dt$, so that f_2 is in fact the density of R_2. Thus, in this case, if we differentiate F_2 and then integrate the derivative, we get back to F_2.

It is reasonable to expect that a random variable R, whose distribution function is continuous everywhere and defined by an explicit formula or collection of formulas, will be absolutely continuous. The following result will cover almost all situations encountered in practice.

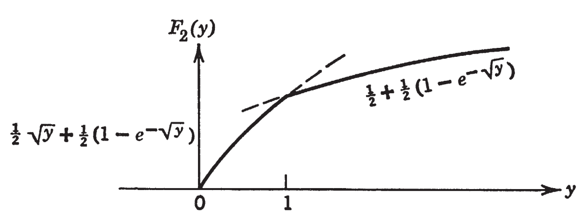


FIGURE 2.4.5

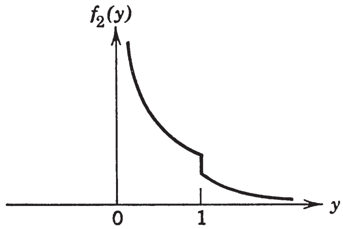


FIGURE 2.4.6

Let R be a random variable with distribution function F. Suppose that

$$\left.\begin{array}{ll} (1) & F \text{ is continuous for all } x \\ (2) & F \text{ is differentiable everywhere except possibly at a finite} \\ & \text{number of points} \\ (3) & \text{The derivative } f(x) = F'(x) \text{ is continuous except possibly at} \\ & \text{a finite number of points} \end{array}\right\} \qquad (2.4.1)$$

Then R is an absolutely continuous random variable with density function f. (The proof involves the application of the fundamental theorem of calculus; see Problem 6.)

NOTE. Density functions need not be continuous or bounded. Also, in this case there is an ambiguity in the values of $f_2(y)$ at $y = 0$ and $y = 1$, since F_2 is not differentiable at these points. However, any values may be assumed, since changing a function at a single point, or a finite or countably infinite number of points, or in fact on a set of total length (Lebesgue measure) zero, does not change the integral.

DENSITY FUNCTION METHOD. In this approach we develop an explicit formula for the density of R_2 in terms of that of R_1.

We first give an informal description. The probability that R_2 will lie in the small interval $[y, y + dy]$ is

$$\int_y^{y+dy} f_2(t)\, dt$$

which is roughly $f_2(y)\, dy$ if f_2 is well-behaved near y. But if we set $h_1(y) = \sqrt{y}$, $h_2(y) = -\sqrt{y}$, $y \geq 0$, then (see Figure 2.4.7)

$$P\{y \leq R_2 \leq y + dy\} = P\{h_1(y) \leq R_1 \leq h_1(y + dy)\} + P\{h_2(y + dy) \leq R_1 \leq h_2(y)\}$$

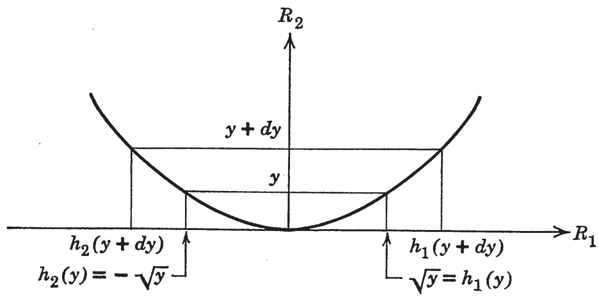


FIGURE 2.4.7

Hence

$$f_2(y)\,dy = f_1(h_1(y))\frac{[h_1(y+dy) - h_1(y)]\,dy}{dy} + f_1(h_2(y))\frac{[h_2(y) - h_2(y+dy)]\,dy}{dy}$$

Let $dy \to 0$ to obtain

$$\begin{aligned} f_2(y) &= f_1(h_1(y))h_1'(y) + f_1(h_2(y))(-h_2'(y)) \\ &= f_1(h_1(y))\,|h_1'(y)| + f_1(h_2(y))\,|h_2'(y)| \end{aligned}$$

In this case

$$f_2(y) = f_1(\sqrt{y})\left|\frac{d}{dy}\sqrt{y}\right| + f_1(-\sqrt{y})\left|\frac{d}{dy} - \sqrt{y}\right| = \frac{1}{2\sqrt{y}}[f_1(\sqrt{y}) + f_1(-\sqrt{y})]$$

Now (see Figure 2.4.1), if $0 < y < 1$,

$$f_2(y) = \frac{1}{2\sqrt{y}}[\tfrac{1}{2}e^{-\sqrt{y}} + \tfrac{1}{2}] = \frac{1}{4\sqrt{y}}(1 + e^{-\sqrt{y}})$$

If $y > 1$,

$$f_2(y) = \frac{1}{2\sqrt{y}}[\tfrac{1}{2}e^{-\sqrt{y}} + 0] = \frac{1}{4\sqrt{y}}e^{-\sqrt{y}}$$

as before.

Similar reasoning shows that in general

$$\underline{f_2(y) = f_1(h_1(y))\,|h_1'(y)| + \cdots + f_1(h_n(y))|h_n'(y)|}$$

where $h_1(y), \ldots, h_n(y)$ are the values of R_1 corresponding to $R_2 = y$.

Here is the formal statement. Suppose that the domain of g can be written as the union of intervals $I_1, I_2, \ldots, I_n$. Assume that over the interval I_j, g is strictly increasing or strictly decreasing and is differentiable (except possibly at the end points), with h_j = inverse of g over I_j. Let F_1 satisfy the three conditions (2.4.1). Then

$$f_2(y) = \sum_{j=1}^{n} f_1(h_j(y))\, |h_j'(y)|$$

where

$$f_1(h_j(y))\, |h_j'(y)|$$

is interpreted as 0 if $y \notin$ the domain of h_j. For the proof, see Problem 7.

REMARK. If we have $R_1 = h(R_2)$, where h is a one-to-one differentiable function, and $R_2 = g(R_1)$, where g is the inverse of h, then

$$h'(y) = \left.\frac{1}{g'(x)}\right]_{x=h(y)}$$

Thus we may write

$$f_2(y) = \sum_{j=1}^{n} \frac{f_1(h_j(y))}{|g'(x)|_{x=h_j(y)}}$$

where $R_2 = g(R_1)$.

In the present example we have

$$g(x) = x^2, \qquad h_1(y) = \sqrt{y}, \qquad h_2(y) = -\sqrt{y}$$

so that

$$f_2(y) = \frac{f_1(\sqrt{y})}{|2x|_{x=\sqrt{y}}} + \frac{f_1(-\sqrt{y})}{|2x|_{x=-\sqrt{y}}} = \frac{1}{2\sqrt{y}}\,[f_1(\sqrt{y}) + f_1(-\sqrt{y})]$$

as before. ◀

▶ **Example 2.** Let R_1 be uniformly distributed between 0 and 2π; that is,

$$\begin{aligned} f_1(x) &= \frac{1}{2\pi}, & 0 \le x \le 2\pi \\ &= 0 & \text{elsewhere} \end{aligned}$$

Let $R_2 = \sin R_1$ (see Figure 2.4.8).

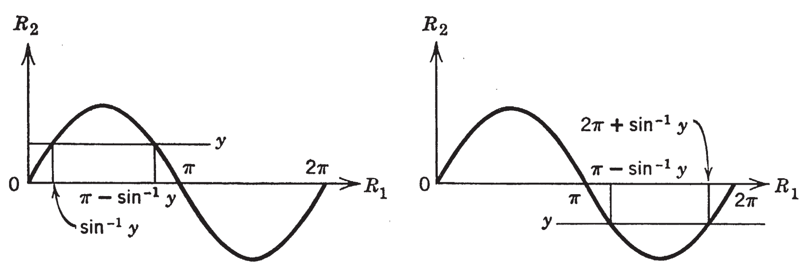


FIGURE 2.4.8

DISTRIBUTION FUNCTION METHOD. If $0 \le y \le 1$,

$$\begin{aligned} F_2(y) &= P\{R_2 \le y\} \\ &= P\{0 \le R_1 \le \sin^{-1} y\} + P\{\pi - \sin^{-1} y \le R_1 \le 2\pi\} \\ &= \int_0^{\sin^{-1} y} \frac{1}{2\pi}\, dx + \int_{\pi - \sin^{-1} y}^{2\pi} \frac{1}{2\pi}\, dx \\ &= \frac{1}{2} + \frac{1}{\pi} \sin^{-1} y \end{aligned}$$

where the branch of the arc sin function is chosen so that $-\pi/2 \le \sin^{-1} y \le \pi/2$.

If $-1 \le y \le 0$,

$$\begin{aligned} F_2(y) &= P\{\pi - \sin^{-1} y \le R_1 \le 2\pi + \sin^{-1} y\} \\ &= \frac{1}{2} + \frac{1}{\pi} \sin^{-1} y \end{aligned}$$

as above, and

$$f_2(y) = F_2'(y) = \frac{1}{\pi\sqrt{1 - y^2}}, \qquad -1 < y < 1$$

DENSITY FUNCTION METHOD. If $0 < y < 1$,

$$\begin{aligned} f_2(y) &= f_1(\sin^{-1} y) \left| \frac{d}{dy} \sin^{-1} y \right| + f_1(\pi - \sin^{-1} y) \left| \frac{d}{dy} (\pi - \sin^{-1} y) \right| \\ &= \frac{1}{\pi\sqrt{1 - y^2}} \end{aligned}$$

Similarly,

$$f_2(y) = \frac{1}{\pi\sqrt{1 - y^2}} \qquad \text{for } -1 < y < 0$$

$[f_2(y) = 0, |y| > 1]$ (see Figure 2.4.9). ◀

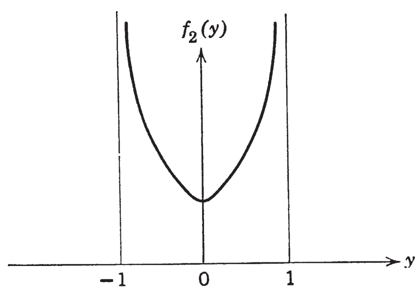


FIGURE 2.4.9

PROBLEMS

1. Let R_1 be absolutely continuous with density

$$f_1(x) = e^{-x}, \qquad x \geq 0; \qquad f_1(x) = 0, \qquad x < 0$$

Define

$$\begin{aligned} R_2 &= R_1 && \text{if } R_1 \leq 1 \\ &= \frac{1}{R_1} && \text{if } R_1 > 1 \end{aligned}$$

Show that R_2 is absolutely continuous and find its density.

2. An absolutely continuous random variable R_1 is uniformly distributed between -1 and $+1$. Find and sketch either the density or the distribution function of the random variable R_2, where $R_2 = e^{-R_1}$.

3. Let R_1 have density $f_1(x) = 1/x^2$, $x \geq 1$; $f_1(x) = 0$, $x < 1$. Define

$$\begin{aligned} R_2 &= 2R_1 && \text{for } R_1 \leq 2 \\ &= R_1^2 && \text{for } R_1 > 2 \end{aligned}$$

Find the density of R_2.

4. Let R_1 be as in Problem 3, and define

$$\begin{aligned} R_2 &= 2R_1 && \text{for } R_1 \leq 2 \\ &= 5 && \text{for } R_1 > 2 \end{aligned}$$

Find and sketch the distribution function of R_2; is R_2 absolutely continuous?

5. (a) Let R_1 have distribution function

$$\begin{aligned} F_1(x) &= 1 - e^{-x}, && x \geq 0 \\ &= 0, && x < 0 \end{aligned}$$

Define

$$\begin{aligned} R_2 &= 1 - e^{-R_1}, && R_1 \geq 0 \\ &= 0, && R_1 < 0 \end{aligned}$$

Show that R_2 is uniformly distributed between 0 and 1.

(b) In general, if a random variable R_1 has a continuous distribution function $g(x) = F_1(x)$ and we define a random variable R_2 by $R_2 = g(R_1)$, show that R_2 is uniformly distributed between 0 and 1.

6. If R is a random variable with distribution function F, where F is continuous everywhere and has a continuous derivative f at all but a finite number of points, show that R is absolutely continuous with density f.

7. Establish the validity of the formula

$$f_2(y) = \sum_{j=1}^{n} f_1(h_j(y))\,|h_j'(y)|$$

under the conditions given in the text.

8. Let R_1 be chosen at random between 0 and 1, with density f_1 [so that $\int_0^1 f_1(y)\,dy = 1$]. Let R_2 be the second digit in the decimal expansion of R_1. (To avoid ambiguity, write, for example, .3 as .3000 $\cdots$, not .2999 $\cdots$.)

(a) Show that $R_2 = k$ iff $i + 10^{-1}k \le 10R_1 < i + 10^{-1}(k+1)$ for some $i = 0, 1, \ldots, 9$. Hence

$$P\{R_2 = k\} = \sum_{i=0}^{9} \int_{10^{-1}i+10^{-2}k}^{10^{-1}i+10^{-2}k+10^{-2}} f_1(y)\,dy, \qquad k = 0, 1, \ldots, 9$$

(b) If R is uniformly distributed between 0 and 1, and $R_1 = \sqrt{R}$, find the probability function of R_2 = the second digit in the decimal expansion of R_1.

9. A projectile is fired with initial velocity v_0 at an angle θ uniformly distributed between 0 and $\pi/2$ (see Figure P.2.4.9). If R is the distance from the launch site

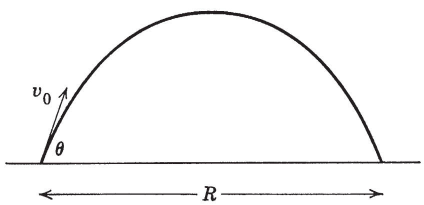


FIGURE P.2.4.9

to the point at which the projectile returns to earth, find the density of R (consider only the effect of gravity).

2.5 PROPERTIES OF DISTRIBUTION FUNCTIONS

We shall establish some general properties of the distribution function of an arbitrary random variable. We need two facts about probability measures.

Theorem 1. *Let $(\Omega, \mathscr{F}, P)$ be a probability space.*

(a) *If $A_1, A_2, \ldots$ is an expanding sequence of sets in $\mathscr{F}$, that is, $A_n \subset A_{n+1}$ for all n, and $A = \bigcup_{n=1}^{\infty} A_n$, then $P(A) = \lim_{n\to\infty} P(A_n)$.*

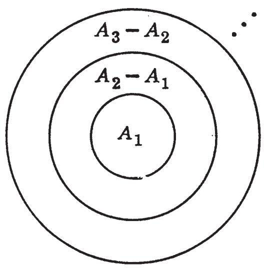


FIGURE 2.5.1 Expanding Sequence.

(b) *If $A_1, A_2, \ldots$ is a contracting sequence of sets in $\mathscr{F}$, that is, $A_{n+1} \subset A_n$ for all n, and $A = \bigcap_{n=1}^{\infty} A_n$, then $P(A) = \lim_{n\to\infty} P(A_n)$.*

PROOF.
(a) We can write

$$A = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \cdots \cup (A_n - A_{n-1}) \cdots$$

(see Figure 2.5.1; note this is the expansion (1.3.11) in the special case of an expanding sequence). Since this is a disjoint union,

$$\begin{aligned} P(A) &= P(A_1) + P(A_2 - A_1) + P(A_3 - A_2) + \cdots \\ &= P(A_1) + P(A_2) - P(A_1) + P(A_3) - P(A_2) + \cdots \qquad \text{since } A_n \subset A_{n+1} \\ &= \lim_{n\to\infty} P(A_n) \end{aligned}$$

(b) If $A = \bigcap_{n=1}^{\infty} A_n$, then, by the DeMorgan laws, $A^c = \bigcup_{n=1}^{\infty} A_n^{\ c}$. Now $A_{n+1} \subset A_n$; hence $A_n^{\ c} \subset A_{n+1}^c$. Thus the sets $A_n^{\ c}$ form an expanding sequence, so, by (a), $P(A_n^{\ c}) \to P(A^c)$; that is, $1 - P(A_n) \to 1 - P(A)$. The result follows.

Theorem 2. *Let F be the distribution function of an arbitrary random variable R. Then*

1. *$F(x)$ is nondecreasing; that is, $a < b$ implies $F(a) \le F(b)$*

For we have shown [see (2.3.2)] that $F(b) - F(a) = P\{a < R \le b\} \ge 0$.

$$2. \quad \lim_{x\to\infty} F(x) = 1$$

Let x_n, $n = 1, 2, \ldots$ be a sequence of real numbers increasing to $+\infty$. Let $A_n = \{R \le x_n\}$. Then the A_n form an expanding sequence. (Since $x_n \le x_{n+1}$, $R \le x_n$ implies $R \le x_{n+1}$.) Now $\bigcup_{n=1}^{\infty} A_n = \Omega$, since, given any point

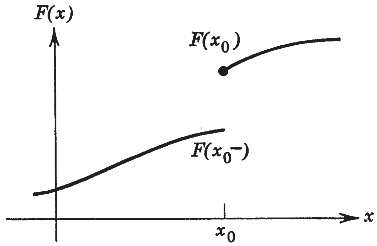


FIGURE 2.5.2 Right Continuity of Distribution Functions.

$\omega \in \Omega$, $R(\omega)$ is a real number; hence, for sufficiently large n, $R(\omega) \le x_n$, so that $\omega \in A_n$. Thus $P(A_n) \to P(\Omega) = 1$, that is, $\lim_{n\to\infty} F(x_n) = 1$.

$$3. \quad \lim_{x\to-\infty} F(x) = 0$$

Let x_n, $n = 1, 2, \ldots$ be a sequence of real numbers decreasing to $-\infty$. Let $A_n = \{R \le x_n\}$. Then the A_n form a contracting sequence. (Since $x_{n+1} \le x_n$, $R \le x_{n+1}$ implies $R \le x_n$.) Now $\bigcap_{n=1}^{\infty} A_n = \varnothing$, since if ω is any point of Ω, $R(\omega)$ cannot always be $\le x_n$ because $x_n \to -\infty$. Thus $P(A_n) \to P(\varnothing) = 0$; that is, $F(x_n) \to 0$.

4. *F is continuous from the right;* that is, $\lim_{x\to x_0^+} F(x) = F(x_0)$

Hence F assumes the upper value at any discontinuity; see Figure 2.5.2.

Let x_n approach x_0 from above; that is, let x_n, $n = 1, 2, \ldots$ be a (strictly) decreasing sequence whose limit is x_0. As before, let $A_n = \{R \le x_n\}$. The A_n form a contracting sequence whose limit (intersection) is $A = \{R \le x_0\}$. In order to show that $\bigcap_{n=1}^{\infty} A_n = \{R \le x_0\}$, we reason as follows. If $R(\omega) \le x_n$ for all n, then, since $x_n \to x_0$, $R(\omega) \le x_0$. Conversely, if $R(\omega) \le x_0$, then, since $x_0 \le x_n$ for all n, $R(\omega) \le x_n$ for all n. Thus $P(A_n) \to P(A)$; that is, $F(x_n) \to F(x_0)$.

$$5. \quad \lim_{x\to x_0^-} F(x) = P\{R < x_0\}$$

[We write $F(x_0^-)$ for $\lim_{x\to x_0^-} F(x)$.]

Let x_n, $n = 1, 2, \ldots$ be a (strictly) increasing sequence whose limit is x_0. Again let $A_n = \{R \le x_n\}$. The A_n form an expanding sequence whose union is $\{R < x_0\}$. To show $\bigcup_{n=1}^{\infty} A_n = \{R < x_0\}$, we reason as follows. If $\omega \in$ some A_n, then $R(\omega) \le x_n$, so that $R(\omega) < x_0$. Conversely, if $R(\omega) < x_0$, then, since $x_n \to x_0$, eventually $R(\omega) \le x_n$, so that $\omega \in \bigcup_{n=1}^{\infty} A_n$. Thus $P(A_n) \to P\{R < x_0\}$, and the result follows.

$$6. \quad P\{R = x_0\} = F(x_0) - F(x_0^-)$$

Thus F is continuous at x_0 iff $P\{R = x_0\} = 0$, and if F is discontinuous at x_0, the magnitude of the jump is the probability that $R = x_0$.

For $P\{R \le x_0\} = P\{R < x_0\} + P\{R = x_0\}$, so that

$$F(x_0) = F(x_0^-) + P\{R = x_0\}$$

REMARK. The random variable R is said to be *continuous* iff its distribution function $F_R(x)$ is a continuous function of x for all x. In any reasonable case a continuous random variable will have a density—that is, it will be absolutely continuous—but it is possible to establish the existence of random variables that are continuous but not absolutely continuous.

7. *Let F be a function from the reals to the reals, satisfying properties 1, 2, 3, and 4 above. Then F is the distribution function of some random variable.*

This is a somewhat vague statement. Let us try to clarify it, even though we omit the proof. What we are doing essentially is making the statement "Let R be a random variable with distribution function F." It is up to us to supply the underlying probability space. As we have done before, we take $\Omega = E^1$, $\mathscr{F}$ = Borel sets, $R(\omega) = \omega$. Now if F is to be the distribution function of R, we must have, for $a < b$,

$$P(a, b] = P\{a < R \le b\} = F(b) - F(a) \qquad \text{by (2.3.2)}$$

It turns out that if F satisfies conditions 1–4, there is a unique probability measure P defined on the Borel subsets of E^1 such that $P(a, b] = F(b) - F(a)$ for all real $a, b, a < b$; thus the probabilities of all events involving R are determined by F. If we let $a \to -\infty$, we obtain $P(-\infty, b] = F(b)$, that is, $P\{R \le b\} = F(b)$, so that in fact F is the distribution function of R. In the special case in which $F(x) = \int_{-\infty}^{x} f(t)\,dt$, where f is a nonnegative integrable function and $\int_{-\infty}^{\infty} f(x)\,dx = 1$, $P(a, b] = F(b) - F(a) = \int_a^b f(x)\,dx$. This is exactly the situation we considered in Theorem 1 of Section 2.3.

PROBLEMS

1. Let R be a random variable with the distribution function shown in Figure P.2.5.1; notice that R is neither discrete nor continuous. Find the probability

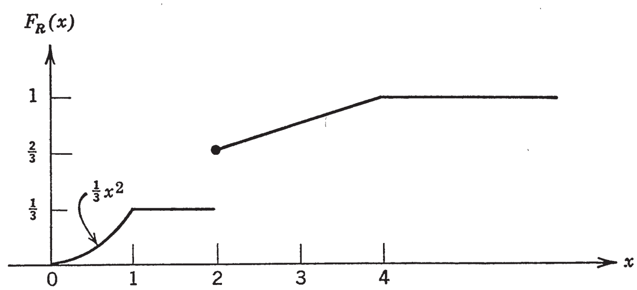


FIGURE P.2.5.1

of the following events.

(a) $\{R = 2\}$

(b) $\{R < 2\}$

(c) $\{R = 2 \text{ or } .5 \le R < 1.5\}$

(d) $\{R = 2 \text{ or } .5 \le R \le 3\}$

2. Let R be an arbitrary random variable with distribution function F. We have seen that $P\{a < R \le b\} = F(b) - F(a)$, $a < b$. Show that

$$P\{a \le R \le b\} = F(b) - F(a^-)$$
$$P\{a \le R < b\} = F(b^-) - F(a^-)$$
$$P\{a < R < b\} = F(b^-) - F(a)$$

(Of course these are all equal if F is continuous at a and b.)

2.6 JOINT DENSITY FUNCTIONS

We are going to investigate situations in which we deal simultaneously with several random variables defined on the same sample space. As an introductory example, suppose that a person is selected at random from a certain population, and his age and weight recorded. We may take as the sample space the set of all pairs (x, y) of real numbers, that is, the Euclidean plane E^2, where we interpret x as the age and y as the weight. Let R_1 be the age of the person selected, and R_2 the weight; that is, $R_1(x, y) = x$, $R_2(x, y) = y$. We wish to assign probabilities to events that involve R_1 and R_2 simultaneously. A cross-section of the available data might appear as shown in Figure 2.6.1. Thus there are 4 million people whose age is between 20 and 25 and (simultaneously) whose weight is between 150 and 160 pounds, and so on. Now suppose that we wish to estimate the number of people between 22 and 23 years, and 154 and 156 pounds. There are 4 million people spread over 5 years and 10 pounds, or 4/50 million per year-pound. We are interested in

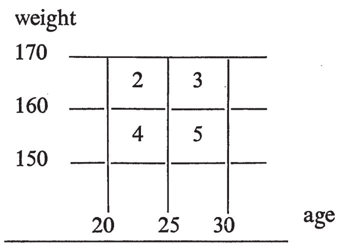


FIGURE 2.6.1 Age-Weight Data (Number of People Is in Millions).

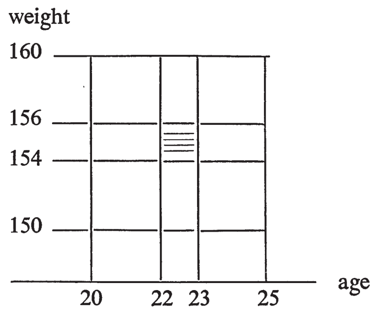


FIGURE 2.6.2 Estimation of Probabilities.

a range of 1 year and 2 pounds, and so our estimate is $4/50 \times 1 \times 2 = 8/50$ million (see Figure 2.6.2). If the total population is 200 million, then

$$P\{22 \leq R_1 \leq 23,\ 154 \leq R_2 \leq 156\}$$

should be approximately

$$\frac{8/50}{200} = .0008$$

NOTATION. $\{22 \leq R_1 \leq 23,\ 154 \leq R_2 \leq 156\}$ means $\{22 \leq R_1 \leq 23$ *and* $154 \leq R_2 \leq 156\}$.

What we are doing is multiplying an age-weight density 4/50 by an area 1×2 to estimate the number of people or, equally well, a probability density $4/[50(200)]$ by an area (1×2) to estimate the probability.

Thus it appears that we should assign probabilities by means of an integral over an area. Let us try to construct an appropriate probability space. We take $\Omega = E^2$, $\mathscr{F}$ = the Borel subsets of E^2. Suppose we have a nonnegative real-valued function f on E^2, with

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$$

Theorem 1 of Section 2.3 holds just as well in the two-dimensional case; there is a unique probability measure P on $\mathscr{F}$ such that $P(B) = \iint_B f(x)\,dx$ for all rectangles B.

If we define $R_1(x, y) = x$, $R_2(x, y) = y$, then

$$P\{(R_1, R_2) \in B\} = P(B) = \iint_B f(x, y)\,dx\,dy$$

For example,

$$P\{a \le R_1 \le b, c \le R_2 \le d\} = \int_{x=a}^{b}\int_{y=c}^{d} f(x, y)\, dx\, dy$$

The *joint distribution function* of two arbitrary random variables R_1 and R_2 is defined by

$$F_{12}(x, y) = P\{R_1 \le x, R_2 \le y\}$$

In the present case we have

$$F_{12}(x, y) = \int_{u=-\infty}^{x}\int_{v=-\infty}^{y} f(u, v)\, du\, dv$$

In general, if R_1 and R_2 are arbitrary random variables defined on a given probability space, the pair (R_1, R_2) is said to be *absolutely continuous* iff there is a nonnegative function $f = f_{12}$ defined on E^2 such that

$$F_{12}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{12}(u, v)\, du\, dv \qquad \text{for all real } x, y \tag{2.6.1}$$

f_{12} is called the *density* of (R_1, R_2) or the *joint density* of R_1 and R_2.

Just as in the one-dimensional case, if (R_1, R_2) is absolutely continuous, it follows that

$$P\{(R_1, R_2) \in B\} = \iint_B f_{12}(x, y)\, dx\, dy$$

for all two-dimensional Borel sets B (see Problem 1). Again, as in the one-dimensional case, if f is a nonnegative function on E^2 with

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$$

we can always find random variables R_1, R_2 such that (R_1, R_2) is absolutely continuous with density f. We take $\Omega = E^2$, $\mathscr{F}$ = Borel sets, $R_1(x, y) = x$, $R_2(x, y) = y$, $P(B) = \iint_B f(x, y)\, dx\, dy$. Even if we use a completely different construction, we get the same result, namely,

$$P\{(R_1, R_2) \in B\} = \iint_B f(x, y)\, dx\, dy$$

We have a similar situation in n dimensions. If the n random variables $R_1, R_2, \ldots, R_n$ are all defined on the same probability space, the *joint distribution function* of $R_1, R_2, \ldots, R_n$ is defined by

$$F_{12\ldots n}(x_1, \ldots, x_n) = P\{R_1 \le x_1, \ldots, R_n \le x_n\}$$

The *random vector* or n-tuple $(R_1, \ldots, R_n)$ is said to be absolutely continuous iff there is a nonnegative function $f_{12\ldots n}$ defined on E^n, called the *density* of $(R_1, \ldots, R_n)$ or the *joint density* of $R_1, \ldots, R_n$, such that

$$F_{12\ldots n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{12\ldots n}(u_1, \ldots, u_n)\, du_1 \cdots du_n \quad (2.6.2)$$

for all real $x_1, \ldots, x_n$.

Notice that $f_{12\ldots n}$ can be recovered from $F_{12\ldots n}$ by differentiation:

$$\frac{\partial^n F_{12\ldots n}(x_1, \ldots, x_n)}{\partial x_1 \cdots \partial x_n} = f_{12\ldots n}(x_1, \ldots, x_n)$$

at least at points where $f_{12\ldots n}$ is continuous.

If $(R_1, \ldots, R_n)$ is absolutely continuous, then

$$P\{(R_1, \ldots, R_n) \in B\} = \int_B \cdots \int f_{12\ldots n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

for all n-dimensional Borel sets B.

If f is a nonnegative function on E^n such that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$$

we can always find random variables $R_1, \ldots, R_n$ such that $(R_1, \ldots, R_n)$ is absolutely continuous with density f. We take $\Omega = E^n$, $\mathscr{F}$ = Borel sets, and define $R_1(x_1, \ldots, x_n) = x_1, \ldots, R_n(x_1, \ldots, x_n) = x_n$. If B is any Borel subset of E^n, we assign

$$P(B) = \int_B \cdots \int f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

Then $(R_1, \ldots, R_n)$ is absolutely continuous with density f.

▶ **Example 1.** Let

$$\begin{aligned} f_{12}(x, y) &= 1 \quad \text{if } 0 \le x \le 1 \quad \text{and} \quad 0 \le y \le 1 \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

(This is the *uniform density* on the unit square.) We may as well take $\Omega = E^2$, $\mathscr{F}$ = Borel sets, $R_1(x, y) = x$, $R_2(x, y) = y$,

$$P(B) = \iint_B f_{12}(x, y)\, dx\, dy$$

Let us calculate the probability that $1/2 \le R_1 + R_2 \le 3/2$. Now

$$\begin{aligned} \{\tfrac{1}{2} \le R_1 + R_2 \le \tfrac{3}{2}\} &= \{(x, y) \colon \tfrac{1}{2} \le R_1(x, y) + R_2(x, y) \le \tfrac{3}{2}\} \\ &= \{(x, y) \colon \tfrac{1}{2} \le x + y \le \tfrac{3}{2}\} \end{aligned}$$

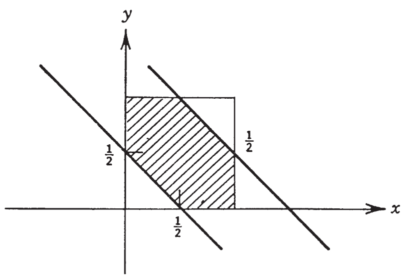


FIGURE 2.6.3
Calculation of $P\{\frac{1}{2} \le R_1 + R_2 \le \frac{3}{2}\}$.

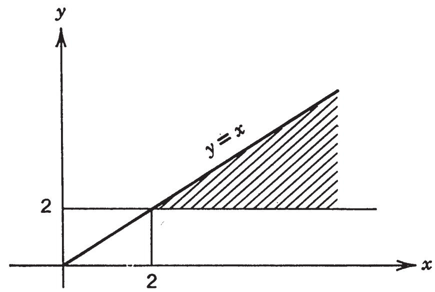


FIGURE 2.6.4
Calculation of $P\{R_1 \ge R_2 \ge 2\}$.

Thus (see Figure 2.6.3)

$$\begin{aligned} P\{\tfrac{1}{2} \le R_1 + R_2 \le \tfrac{3}{2}\} &= \iint\limits_{1/2 \le x+y \le 3/2} f_{12}(x, y)\,dx\,dy \\ &= \iint\limits_{\text{shaded area}} 1\,dx\,dy = \text{shaded area} \\ &= 1 - 2(\tfrac{1}{8}) = \tfrac{3}{4} \end{aligned}$$

If we want the probability that $1/2 \le R_1 \le 3/4$ and $0 \le R_2 \le 1/2$, we obtain

$$P\{\tfrac{1}{2} \le R_1 \le \tfrac{3}{4}, 0 \le R_2 \le \tfrac{1}{2}\} = \int_{x=1/2}^{3/4} \int_{y=0}^{1/2} 1\,dx\,dy = \tfrac{1}{2}(\tfrac{1}{4}) = \tfrac{1}{8} \blacktriangleleft$$

▶ **Example 2.** Let

$$\begin{aligned} f_{12}(x, y) &= e^{-(x+y)}, \qquad x, y \ge 0 \\ &= 0 \qquad \text{elsewhere} \end{aligned}$$

Let us calculate the probability that $R_1 \ge R_2 \ge 2$. We have (see Figure 2.6.4)

$$\begin{aligned} P\{R_1 \ge R_2 \ge 2\} &= \iint\limits_{x \ge y \ge 2} f_{12}(x, y)\,dx\,dy \\ &= \int_2^\infty e^{-x}\,dx \int_2^x e^{-y}\,dy = \int_2^\infty e^{-x}(e^{-2} - e^{-x})\,dx \\ &= e^{-4} - \tfrac{1}{2}e^{-4} = \tfrac{1}{2}e^{-4} \blacktriangleleft \end{aligned}$$

To summarize:

$$P\{(R_1, R_2) \in B\} = \iint\limits_{(x,y) \in B} f_{12}(x, y)\,dx\,dy$$

The probability of any event is found by integrating the density function *over the set defined by the event*. This is perhaps about as close as one can come to a one-sentence summary of the role of density functions in probability theory.

PROBLEMS

1. Let F_{12} be the joint distribution function of R_1 and R_2, where (R_1, R_2) is absolutely continuous with density f_{12}. Show that

$$P\{a_1 < R_1 \le b_1, a_2 < R_2 \le b_2\} = \int_{a_1}^{b_1}\int_{a_2}^{b_2} f_{12}(x, y)\, dx\, dy$$

The uniqueness part of Theorem 2.3.1 (generalized to two dimensions) shows that

$$P\{(R_1, R_2) \in B\} = \iint_B f_{12}(x, y)\, dx\, dy$$

for all two-dimensional Borel sets B.
HINT: If F is the joint distribution function of the random variables R_1 and R_2, show that

$$P\{a_1 < R_1 \le b_1, a_2 < R_2 \le b_2\} = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

2. If F is the joint distribution function of the random variables R_1, R_2, and R_3, express

$$P\{a_1 < R_1 \le b_1, a_2 < R_2 \le b_2, a_3 < R_3 \le b_3\}$$

in terms of F. Can you see a general pattern that will extend this result to n dimensions?

3. If

$$\begin{aligned} F(x, y) &= 1 \qquad \text{for } x + y \ge 0 \\ &= 0 \qquad \text{for } x + y < 0 \end{aligned}$$

(see Figure P.2.6.3.), show that F cannot possibly be the joint distribution

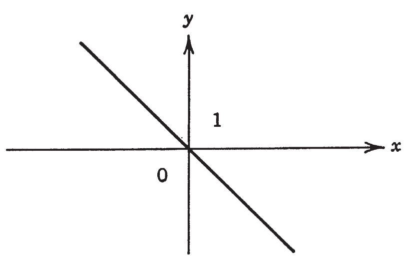


FIGURE P.2.6.3

function of a pair of random variables (see Problem 1.)

4. Let R_1 and R_2 have the following joint density:

$$\begin{aligned} f_{12}(x, y) &= \tfrac{1}{4} \quad \text{if } -1 \le x \le 1 \quad \text{and} \quad -1 \le y \le 1 \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

(This corresponds to R_1 and R_2 being chosen independently, each uniformly distributed between -1 and $+1$; we elaborate on this in the next section.) Find the probability of each of the following events.

(a) $\{R_1 + R_2 \le \frac{1}{2}\}$
(b) $\{R_1 - R_2 \le \frac{1}{2}\}$
(c) $\{R_1R_2 \le \frac{1}{4}\}$
(d) $\left\{\dfrac{R_2}{R_1} \le \dfrac{1}{2}\right\}$
(e) $\left\{\left|\dfrac{R_2}{R_1}\right| \le \dfrac{1}{2}\right\}$
(f) $\{|R_1| + |R_2| \le 1\}$
(g) $\{|R_2| \le e^{R_1}\}$

2.7 RELATIONSHIP BETWEEN JOINT AND INDIVIDUAL DENSITIES; INDEPENDENCE OF RANDOM VARIABLES

If R_1 and R_2 are two random variables defined on the same probability space, we wish to investigate the relation between the characterization of the random variables individually and their characterization simultaneously. We shall consider two problems.

1. If (R_1, R_2) is absolutely continuous, are R_1 and R_2 absolutely continuous, and, if so, how can the individual densities of R_1 and R_2 be found in terms of the joint densities?

2. Given R_1, R_2 (individually) absolutely continuous, is (R_1, R_2) absolutely continuous, and, if so, can the joint density be derived from the individual density?

Problem 1

To go from simultaneous information to individual information is essentially a matter of adding across a row or column. For example, suppose that a group of 14 people has the age-weight distribution shown in Figure 2.7.1. The number of people between 20 and 25 years is found by adding the numbers in the first column; thus $4 + 2 = 6$.

Let us develop this idea a bit further. If R_1 and R_2 are discrete, the *joint probability function* of R_1 and R_2 [or the probability function of the pair

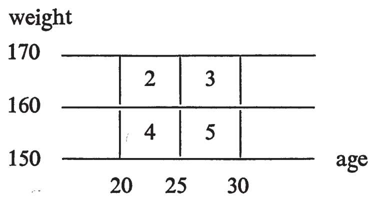


FIGURE 2.7.1 Calculation of Individual Probabilities from Joint Probabilities.

(R_1, R_2)] is defined by

$$p_{12}(x, y) = P\{R_1 = x, R_2 = y\} \qquad x, y \text{ real} \tag{2.7.1}$$

If the possible values of R_2 are $y_1, y_2, \ldots$, then

$$\{R_1 = x\} = \{R_1 = x, R_2 = y_1\} \cup \{R_1 = x, R_2 = y_2\} \cup \cdots$$

since the events $\{R_2 = y_n\}$, $n = 1, 2, \ldots$ are mutually exclusive and exhaustive. Thus the probability function of R_1 is given by

$$p_1(x) = P\{R_1 = x\} = \sum_y p_{12}(x, y) \tag{2.7.2}$$

Similarly,

$$p_2(y) = P\{R_2 = y\} = \sum_x p_{12}(x, y) \tag{2.7.3}$$

There are analogous formulas in higher dimensions, for example,

$$p_{12}(x, y) = \sum_z p_{123}(x, y, z) \qquad p_2(y) = \sum_{x,z} p_{123}(x, y, z)$$

where $p_{123}(x, y, z) = P\{R_1 = x, R_2 = y, R_3 = z\}$.

Now let us return to the absolutely continuous case. If (R_1, R_2) is absolutely continuous with joint density f_{12}, we shall show that R_1 is absolutely continuous (and so is R_2) and find f_1 and f_2 in terms of f_{12}.

For any x_0 we have, intuitively,

$$P\{x_0 \le R_1 \le x_0 + dx_0\} \approx f_1(x_0)\, dx_0 \tag{2.7.4}$$

But

$$\begin{aligned} P\{x_0 \le R_1 \le x_0 + dx_0\} &= P\{x_0 \le R_1 \le x_0 + dx_0, -\infty < R_2 < \infty\} \\ &= \int_{x_0}^{x_0+dx_0} dx \int_{-\infty}^{\infty} f_{12}(x, y)\, dy \end{aligned}$$

(see Figure 2.7.2).

If f_{12} is well-behaved, this is approximately

$$dx_0 \int_{-\infty}^{\infty} f_{12}(x_0, y)\, dy \tag{2.7.5}$$

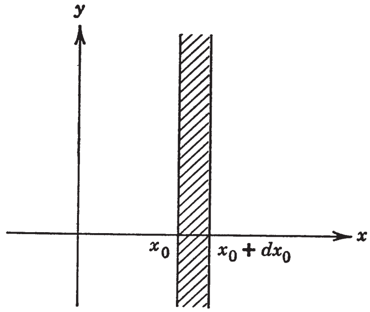


FIGURE 2.7.2 Calculation of Individual Densities from Joint Densities.

From (2.7.4) and (2.7.5) (replacing x_0 by x) we have

$$f_1(x) = \int_{-\infty}^{\infty} f_{12}(x, y)\, dy$$

To verify this formally, we work with the distribution function of R_1.

$$F_1(x_0) = P\{R_1 \le x_0\} = P\{R_1 \le x_0, -\infty < R_2 < \infty\}$$
$$= \int_{x=-\infty}^{x_0} \left[\int_{y=-\infty}^{\infty} f_{12}(x, y)\, dy \right] dx$$

Thus F_1 is represented as an integral, and so R_1 is absolutely continuous with density

$$f_1(x) = \int_{-\infty}^{\infty} f_{12}(x, y)\, dy \tag{2.7.6}$$

Similarly,

$$f_2(y) = \int_{-\infty}^{\infty} f_{12}(x, y)\, dx \tag{2.7.7}$$

In exactly the same way we may establish similar formulas in higher dimensions; for example,

$$f_{12}(x, y) = \int_{-\infty}^{\infty} f_{123}(x, y, z)\, dz \tag{2.7.8}$$

$$f_2(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{123}(x, y, z)\, dx\, dz \tag{2.7.9}$$

The process of obtaining the individual densities from the joint density is sometimes called the calculation of *marginal densities*, because of the similarity to the process of adding across a row or column.

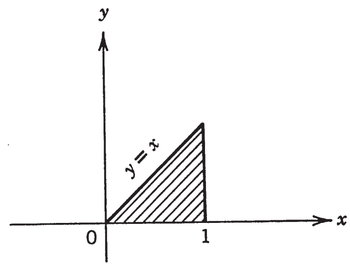


FIGURE 2.7.3

▶ **Example 1.** Let

$$f_{12}(x, y) = 8xy, \qquad 0 \le y \le x \le 1$$
$$= 0 \qquad \text{elsewhere}$$

(see Figure 2.7.3).

$$f_1(x) = \int_{-\infty}^{\infty} f_{12}(x, y)\,dy$$
$$= 0 \qquad \text{if } x < 0 \quad \text{or} \quad x > 1$$

If $0 \le x \le 1$,

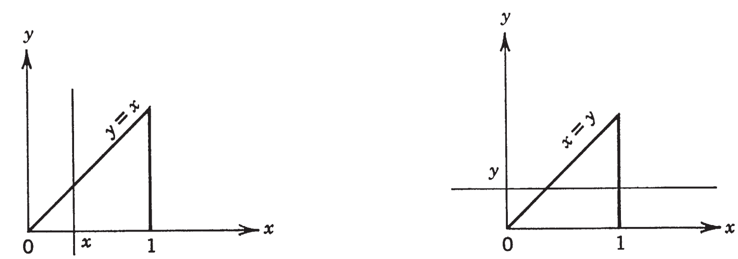
$$f_1(x) = \int_0^x 8xy\,dy = 4x^3 \qquad \text{(Figure 2.7.4a)}$$

$$f_2(y) = \int_{-\infty}^{\infty} f_{12}(x, y)\,dx$$
$$= 0 \qquad \text{if } y < 0 \quad \text{or} \quad y > 1$$

If $0 \le y \le 1$,

$$f_2(y) = \int_y^1 8xy\,dx = 4y(1 - y^2) \qquad \text{(Figure 2.7.4b)}$$

Sketches of f_1 and f_2 are given in Figure 2.7.5. ◀



(a) (b)

FIGURE 2.7.4

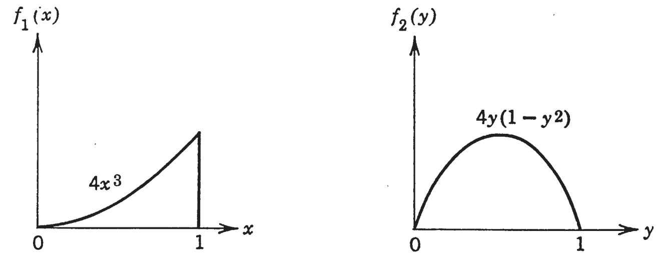


FIGURE 2.7.5

Problem 2

The second problem posed at the beginning of this section has a negative answer; that is, if R_1 and R_2 are each absolutely continuous then (R_1, R_2) is *not* necessarily absolutely continuous. Furthermore, even if (R_1, R_2) is absolutely continuous, $f_1(x)$ and $f_2(y)$ do *not* determine $f_{12}(x, y)$. We give examples later in the section.

However, there is an affirmative answer when the random variables are independent. We have considered the notion of independence of events, and this can be used to define independence of random variables. Intuitively, the random variables $R_1, \ldots, R_n$ are independent if knowledge about some of the R_i does not change the odds about the other R_i's. In other words, if A_i is an event involving R_i alone, that is, if $A_i = \{R_i \in B_i\}$, then the events $A_1, \ldots, A_n$ should be independent. Formally, we define independence as follows.

DEFINITION. Let $R_1, \ldots, R_n$ be random variables on $(\Omega, \mathscr{F}, P)$. $R_1, \ldots, R_n$ are said to be *independent* iff for all Borel subsets $B_1, \ldots, B_n$ of E^1 we have

$$P\{R_1 \in B_1, \ldots, R_n \in B_n\} = P\{R_1 \in B_1\} \cdots P\{R_n \in B_n\}$$

REMARK. If $R_1, \ldots, R_n$ are independent, so are $R_1, \ldots, R_k$ for $k < n$. For

$$\begin{aligned} P\{R_1 \in B_1, \ldots, R_k \in B_k\} &= P\{R_1 \in B_1, \ldots, R_k \in B_k, \\ &\qquad -\infty < R_{k+1} < \infty, \ldots, -\infty < R_n < \infty\} \\ &= P\{R_1 \in B_1\} \ldots P\{R_k \in B_k\} \end{aligned}$$

since $P\{-\infty < R_i < \infty\} = 1$. If $(R_i,\ i \in$ the index set $I)$, is an *arbitrary* family of random variables on the space $(\Omega, \mathscr{F}, P)$, the R_i are said to be independent iff for each finite set of distinct indices $i_1, \ldots, i_k \in I$, $R_{i_1}, \ldots, R_{i_k}$ are independent.

We may now give the solution to Problem 2 under the hypothesis of independence.

Theorem 1. *Let $R_1, R_2, \ldots, R_n$ be independent random variables on a given probability space. If each R_i is absolutely continuous with density f_i, then $(R_1, R_2, \ldots, R_n)$ is absolutely continuous; also, for all $x_1, \ldots, x_n$,*

$$f_{12\cdots n}(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n)$$

Thus in this sense the joint density is the product of the individual densities.

PROOF. The joint distribution function of $R_1, \ldots, R_n$ is given by

$$\begin{aligned} F_{12\cdots n}(x_1, \ldots, x_n) &= P\{R_1 \le x_1, \ldots, R_n \le x_n\} \\ &= P\{R_1 \le x_1\} \cdots P\{R_n \le x_n\} \qquad \text{by independence} \\ &= \int_{-\infty}^{x_1} f_1(u_1)\,du_1 \cdots \int_{-\infty}^{x_n} f_n(u_n)\,du_n \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(u_1)\cdots f_n(u_n)\,du_1 \cdots du_n \end{aligned}$$

It follows from the definition of absolute continuity [see (2.6.2)] that $(R_1, \ldots, R_n)$ is absolutely continuous and that the joint density is $f_{12\cdots n}(x_1, \ldots, x_n) = f_1(x_1)\cdots f_n(x_n)$.

Note that we have the following intuitive interpretation (when $n = 2$). From the independence of R_1 and R_2 we obtain

$$\begin{aligned} &P\{x \le R_1 \le x + dx, y \le R_2 \le y + dy\} \\ &\qquad = P\{x \le R_1 \le x + dx\}P\{y \le R_2 \le y + dy\} \end{aligned}$$

If there is a joint density, we have (roughly) $f_{12}(x, y)\,dx\,dy = f_1(x)\,dx\, f_2(y)\,dy$, so that $f_{12}(x, y) = f_1(x)f_2(y)$.

As a consequence of this result, the statement "Let $R_1, \ldots, R_n$ be independent random variables, with R_i having density f_i," is unambiguous in the sense that it completely determines all probabilities of events involving the random vector $(R_1, \ldots, R_n)$; if B is an n-dimensional Borel set,

$$P\{(R_1, \ldots, R_n) \in B\} = \int \cdots \int_B f_1(x_1)\cdots f_n(x_n)\,dx_1 \cdots dx_n$$

We now show that Problem 2 has a negative answer when the hypothesis of independence is dropped. We have seen that if $(R_1, \ldots, R_n)$ is absolutely continuous then each R_i is absolutely continuous, but the converse is false in general if the R_i are not independent; that is, each of the random variables $R_1, \dots, R_n$ can have a density without there being a density for the n-tuple $(R_1, \dots, R_n)$.

▶ **Example 2.** Let R_1 be an absolutely continuous random variable with density f, and take $R_2 \equiv R_1$; that is, $R_2(\omega) = R_1(\omega)$, $\omega \in \Omega$. Then R_2 is absolutely continuous, but (R_1, R_2) is not. For suppose that (R_1, R_2) has a density g. Necessarily $(R_1, R_2) \in L$, where L is the line $y = x$, but

$$P\{(R_1, R_2) \in L\} = \iint_L g(x, y)\, dx\, dy$$

Since L has area 0, the integral on the right is 0. But the probability on the left is 1, a contradiction. ◀

We can also give an example to show that if R_1 and R_2 are each absolutely continuous (but not necessarily independent), then even if (R_1, R_2) is absolutely continuous, the joint density is not determined by the individual densities.

▶ **Example 3.** Let

$$\begin{aligned} f_{12}(x, y) &= \tfrac{1}{4}(1 + xy), \qquad && \begin{matrix} -1 \le x \le 1 \\ -1 \le y \le 1 \end{matrix} \\ &= 0 && \text{elsewhere} \end{aligned}$$

Since

$$\int_{-1}^{1} x\, dx = \int_{-1}^{1} y\, dy = 0,$$

$$\begin{aligned} f_1(x) = \int_{-\infty}^{\infty} f_{12}(x, y)\, dy &= \tfrac{1}{2}, \qquad && -1 \le x \le 1 \\ &= 0 && \text{elsewhere} \end{aligned}$$

$$\begin{aligned} f_2(y) &= \tfrac{1}{2}, \qquad && -1 \le y \le 1 \\ &= 0 && \text{elsewhere} \end{aligned}$$

But if

$$\begin{aligned} f_{12}(x, y) &= \tfrac{1}{4}, \qquad && \begin{matrix} -1 \le x \le 1 \\ -1 \le y \le 1 \end{matrix} \\ &= 0 && \text{elsewhere} \end{aligned}$$

we get the same individual densities. ◀

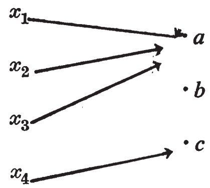


FIGURE 2.7.6

Now intuitively, if R_1 and R_2 are independent, then, say, e^{R_1} and $\sin R_2$ should be independent, since information about e^{R_1} should not change the odds concerning R_2 and hence should not affect $\sin R_2$ either. We shall prove a theorem of this type, but first we need some additional terminology.

If g is a function that maps points in the set D into points in the set E,† and $T \subset E$, we define the *preimage* of T under g as

$$g^{-1}(T) = \{x \in D: g(x) \in T\}$$

For example, let $D = \{x_1, x_2, x_3, x_4\}$, $E = \{a, b, c\}$, $g(x_1) = g(x_2) = g(x_3) = a$, $g(x_4) = c$ (see Figure 2.7.6). We then have

$$\begin{aligned} g^{-1}\{a\} &= \{x_1, x_2, x_3\} \\ g^{-1}\{a, b\} &= \{x_1, x_2, x_3\} \\ g^{-1}\{a, c\} &= \{x_1, x_2, x_3, x_4\} \\ g^{-1}\{b\} &= \varnothing \end{aligned}$$

Note that, by definition of preimage, $x \in g^{-1}(T)$ iff $g(x) \in T$.

Now let $R_1, \ldots, R_n$ be random variables on a given probability space, and let $g_1, \ldots, g_n$ be functions of *one* variable, that is, functions from the reals to the reals. Let $R_1' = g_1(R_1), \ldots, R_n' = g_n(R_n)$; that is, $R_i'(\omega) = g_i(R_i(\omega))$, $\omega \in \Omega$. We assume that the R_i' are also random variables; this will be the case if the g_i are continuous or piecewise continuous. Specifically, we have the following result, which we shall use without proof.

If g is a real-valued function defined on the reals, and g is piecewise continuous, then for each Borel set $B \subset E^1$, $g^{-1}(B)$ is also a Borel subset of E^1. (A function with this property is said to be *Borel measurable.*)

Now we show that if g_i is piecewise continuous or, more generally, Borel measurable, R_i' is a random variable. Let B_i' be a Borel subset of E^1. Then

$$\begin{aligned} R_i'^{-1}(B_i') &= \{\omega: R_i'(\omega) \in B_i'\} \\ &= \{\omega: g_i(R_i(\omega)) \in B_i'\} \\ &= \{\omega: R_i(\omega) \in g_i^{-1}(B_i')\} \in \mathscr{F} \end{aligned}$$

since $g_i^{-1}(B_i')$ is a Borel set.

† A common notation for such a function is $g: D \to E$. It means simply that $g(x)$ is defined and belongs to E for each x in D.

Similarly, if g is a continuous real-valued function defined on E^n, then, for each Borel set $B \subset E^1$, $g^{-1}(B)$ is a Borel subset of E^n. It follows that if $R_1, \ldots, R_n$ are random variables, so is $g(R_1, \ldots, R_n)$.

Theorem 2. *If $R_1, \ldots, R_n$ are independent, then $R_1', \ldots, R_n'$ are also independent.* (For short, "functions of independent random variables are independent.")

PROOF. If $B_1', \ldots, B_n'$ are Borel subsets of E^1, then

$$
\begin{aligned}
P\{R_1' \in B_1', \ldots, R_n' \in B_n'\} &= P\{g_1(R_1) \in B_1' \ \ldots, g_n(R_n) \in B_n'\} \\
&= P\{R_1 \in g_1^{-1}(B_1'), \ldots, R_n \in g_n^{-1}(B_n')\} \\
&= \prod_{i=1}^{n} P\{R_i \in g_i^{-1}(B_i')\} \quad \text{by independence of the } R_i \\
&= \prod_{i=1}^{n} P\{g_i(R_i) \in B_i'\} = \prod_{i=1}^{n} P\{R_i' \in B_i'\}
\end{aligned}
$$

PROBLEMS

1. Let (R_1, R_2) have the following density function.

$$
\begin{aligned}
f_{12}(x, y) &= 4xy && \text{if } 0 \le x \le 1, 0 \le y \le 1, x \ge y \\
&= 6x^2 && \text{if } 0 \le x \le 1, 0 \le y \le 1, x < y \\
&= 0 && \text{elsewhere}
\end{aligned}
$$

(a) Find the individual density functions f_1 and f_2.
(b) If $A = \{R_1 \le \frac{1}{2}\}$, $B = \{R_2 \le \frac{1}{2}\}$, find $P(A \cup B)$.

2. If (R_1, R_2) is absolutely continuous with

$$
\begin{aligned}
f_{12}(x, y) &= 2e^{-(x+y)}, && 0 \le y \le x \\
&= 0 && \text{elsewhere}
\end{aligned}
$$

find $f_1(x)$ and $f_2(y)$.

3. Let (R_1, R_2) be uniformly distributed over the parallelogram with vertices $(-1, 0)$, $(1, 0)$, $(2, 1)$, and $(0, 1)$.
(a) Find and sketch the density functions of R_1 and R_2.
(b) A new random variable R_3 is defined by $R_3 = R_1 + R_2$. Show that R_3 is absolutely continuous, and find and sketch its density.

4. If $R_1, R_2, \ldots, R_n$ are independent, show that the joint distribution function is the product of the individual distribution functions; that is,

$$
F_{12\cdots n}(x_1, x_2, \ldots, x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n) \qquad \text{for all real } x_1, \ldots, x_n
$$

[Conversely, it can be shown that if $F_{12\cdots n}(x_1, \ldots, x_n) = F_1(x_1) \cdots F_n(x_n)$ for all real $x_1, \ldots, x_n$, then $R_1, \ldots, R_n$ are independent.]

5. Show that a random variable R is independent of itself—in other words, R and R are independent—if and only if R is *degenerate*, that is, essentially constant ($P\{R = c\} = 1$ for some c).
6. Under what conditions will R and $\sin R$ be independent? (Use Problem 5 and the result that functions of independent random variables are independent.)
7. If $(R_1, \ldots, R_n)$ is absolutely continuous and $f_{12\cdots n}(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n)$ for all $x_1, \ldots, x_n$, show that $R_1, \ldots, R_n$ are independent.
8. Let (R_1, R_2) be absolutely continuous with density $f_{12}(x, y) = (x + y)/8$, $0 \le x \le 2$, $0 \le y \le 2$; $f_{12}(x, y) = 0$ elsewhere.
   (a) Find the probability that $R_1^2 + R_2 \le 1$.
   (b) Find the conditional probability that exactly one of the random variables R_1, R_2 is ≤ 1, given that at least one of the random variables is ≤ 1.
   (c) Determine whether or not R_1 and R_2 are independent.

2.8 FUNCTIONS OF MORE THAN ONE RANDOM VARIABLE

We are now equipped to consider a wide variety of problems of the following sort. If $R_1, \ldots, R_n$ are random variables with a given joint density, and we define $R = g(R_1, \ldots, R_n)$, we ask for the distribution or density function of R. We shall use a distribution function approach to these problems; that is, we shall find the distribution function of R directly. There is also a density function method, but it is usually not as convenient; the density function approach is outlined in Problem 12. The distribution function method can be described as follows.

$$F_R(z) = P\{R \le z\} = P\{g(R_1, \ldots, R_n) \le z\}$$
$$= \int_{g(x_1, \ldots, x_n) \le z} \cdots \int f_{12\cdots n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

▶ **Example 1.** Let R_1 and R_2 be uniformly distributed between 0 and 1, and independent.

(a) Let $R_3 = R_1 + R_2$. Then, since $f_{12}(x, y) = f_1(x)f_2(y)$ by independence,

$$F_3(z) = P\{R_1 + R_2 \le z\} = \iint_{x+y \le z} f_1(x)f_2(y)\, dx\, dy$$

If $0 \le z \le 1$ (see Figure 2.8.1a),

$$F_3(z) = \iint_{\text{shaded area}} 1\, dx\, dy = \text{shaded area} = \frac{z^2}{2}$$

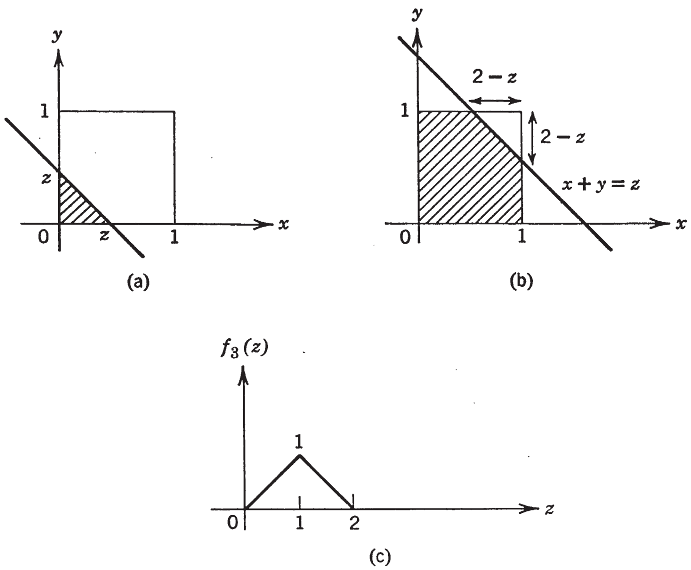


FIGURE 2.8.1 (*a*) Calculation of $F_3(z)$, $0 \le z \le 1$. (*b*) Calculation of $F_3(z)$, $1 \le z \le 2$. (*c*) $f_3(z)$.

If $1 \le z \le 2$ (see Figure 2.8.1b),

$$F_3(z) = \text{shaded area} = 1 - \frac{(2-z)^2}{2}$$

Thus $f_3(z) = z$, $0 \le z \le 1$; $f_3(z) = 2 - z$, $1 \le z \le 2$; $f_3(z) = 0$ elsewhere (see Figure 2.8.1c).

(b) Let $R_3 = R_1R_2$. (Notice that $0 \le R_3 \le 1$.) If $0 \le z \le 1$ (see Figure 2.8.2),

$$F_3(z) = P\{R_1R_2 \le z\} = \iint\limits_{xy \le z} f_{12}(x, y)\,dx\,dy$$

$$= \text{shaded area} = z + \int_z^1 \frac{z}{x}\,dx = z - z \ln z$$

$$f_3(z) = -\ln z \qquad 0 < z \le 1$$

$$= 0 \qquad \text{elsewhere}$$

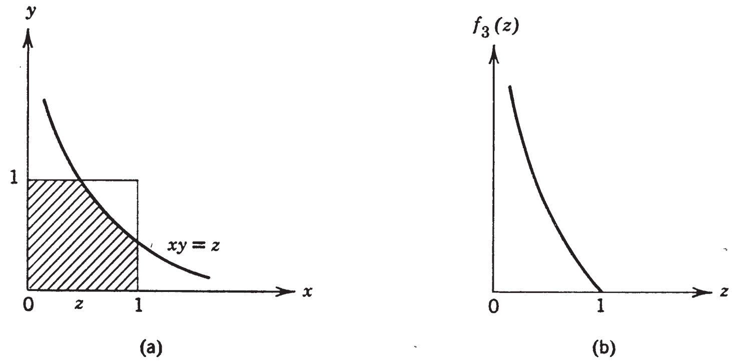


FIGURE 2.8.2

(c) Let $R_3 = \max(R_1, R_2)$. If $0 \le z \le 1$ (see Figure 2.8.3),

$$P\{R_3 \le z\} = P\{R_1 \le z, R_2 \le z\} = \text{shaded area} = z^2$$

[Alternatively, $F_3(z) = P\{R_1 \le z\}P\{R_2 \le z\} = z^2$ by independence.]

$$f_3(z) = 2z, \qquad 0 \le z \le 1 \blacktriangleleft$$

Before the next example, we introduce the *Gaussian* or *normal* density function.

$$f(x) = \frac{1}{\sqrt{2\pi}b} e^{-(x-a)^2/2b^2}, \qquad x \text{ real } (b > 0, a \text{ any real number}) \quad (2.8.1)$$

This is the familiar bell-shaped curve centered at a (Figure 2.8.4); the smaller the value of b, the higher the peak and the more f is concentrated close to $x = a$. To check that this is a legitimate density, we must show that the area

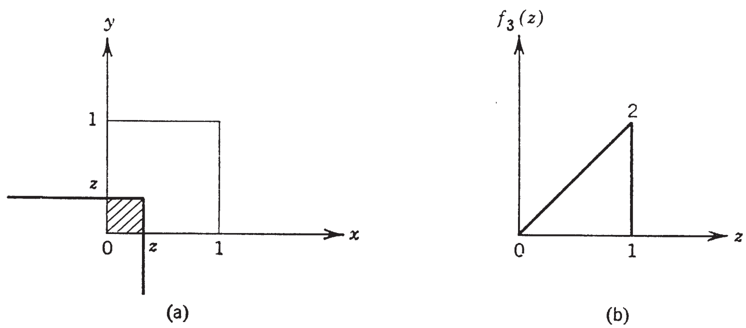


FIGURE 2.8.3

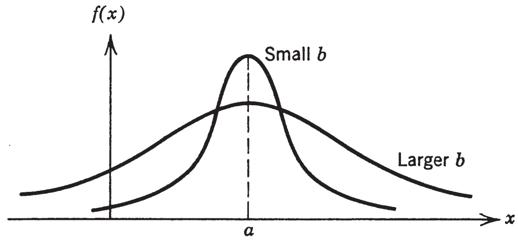


FIGURE 2.8.4 Normal Density.

under f is 1. Let

$$I = \int_{-\infty}^{\infty} e^{-x^2}\, dx$$

Then

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2}\, dx \int_{-\infty}^{\infty} e^{-y^2}\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\, dx\, dy = \text{(in polar coordinates)}$$

$$\int_0^{2\pi} d\theta \int_0^{\infty} re^{-r^2}\, dr = \pi$$

so that

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi} \tag{2.8.2}$$

Thus

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}b} e^{-(x-a)^2/2b^2}\, dx = \left(\text{with } y = \frac{x-a}{\sqrt{2}b}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-y^2}\, dy = 1$$

▶ **Example 2.** Let R_1, R_2, and R_3 be independent, each normally distributed (i.e., having the normal density), with $a = 0$, $b = 1$. Let $R_4 = (R_1^2 + R_2^2 + R_3^2)^{1/2}$; take the positive square root so that $R_4 \geq 0$. (For example, if R_1, R_2, and R_3 are the velocity components of a particle, then R_4 is the speed.) Find the distribution function of R_4.

$$F_4(w) = P\{R_4 \leq w\} = P\{R_1^2 + R_2^2 + R_3^2 \leq w^2\}$$

$$= \iiint\limits_{x^2+y^2+z^2 \leq w^2} (2\pi)^{-3/2} e^{-(x^2+y^2+z^2)/2} dx\, dy\, dz$$

We switch to spherical coordinates:

$$x = r \sin\phi \cos\theta$$
$$y = r \sin\phi \sin\theta$$
$$z = r \cos\phi$$

(ϕ is the "cone angle," and θ the "polar coordinate angle.") Then

$$F_4(w) = \int_0^{2\pi} d\theta \int_0^{\pi} d\phi \int_0^{w} (2\pi)^{-3/2} e^{-r^2/2} r^2 \sin \phi \, dr$$
$$= (2\pi)^{-3/2}(2\pi)(2) \int_0^{w} r^2 e^{-r^2/2} \, dr$$

Thus R_4 has a density given by

$$f_4(w) = \frac{2}{\sqrt{2\pi}} w^2 e^{-w^2/2} \qquad w \geq 0$$
$$= 0, \qquad w < 0 \blacktriangleleft$$

▶ **Example 3.** There are certain situations in which it is possible to avoid all integration in an n-dimensional problem. Suppose that $R_1, \dots, R_n$ are independent and F_i is the distribution function of R_i, $i = 1, 2, \dots, n$.

Let T_k be the kth smallest of the R_i. [For example, if $n = 4$ and $R_1(\omega) = 3$, $R_2(\omega) = 1.5$, $R_3(\omega) = -10$, $R_4(\omega) = 7$, then

$$T_1(\omega) = \min_i R_i(\omega) = R_3(\omega) = -10, \qquad T_2(\omega) = R_2(\omega) = 1.5$$
$$T_3(\omega) = R_1(\omega) = 3, \qquad T_4(\omega) = \max_i R_i(\omega) = R_4(\omega) = 7\,]$$

(Ties may be broken, for example, by favoring the random variable with the smaller subscript.)

We wish to find the distribution function of T_k. When $k = 1$ or $k = n$, the calculation is brief.

$$P\{T_n \leq x\} = P\{\max (R_1, \dots, R_n) \leq x\} = P\{R_1 \leq x, \dots, R_n \leq x\}$$
$$= \prod_{i=1}^{n} P\{R_i \leq x\} \qquad \text{by independence}$$

Thus

$$F_{T_n}(x) = \prod_{i=1}^{n} F_i(x)$$

$$P\{T_1 \leq x\} = 1 - P\{T_1 > x\} = 1 - P\{\min (R_1, \dots, R_n) > x\}$$
$$= 1 - P\{R_1 > x, \dots, R_n > x\} = 1 - \prod_{i=1}^{n} P\{R_i > x\}$$

Thus

$$F_{T_1}(x) = 1 - \prod_{i=1}^{n} (1 - F_i(x))$$

REMARK. We may also calculate $F_{T_1}(x)$ as follows.

$$\begin{aligned}P\{T_1 \le x\} &= P\{\text{at least one } R_i \text{ is } \le x\}\\ &= P(A_1 \cup A_2 \cup \cdots \cup A_n) \qquad \text{where } A_i = \{R_i \le x\}\\ &= P(A_1) + P(A_1{}^c \cap A_2) + \cdots\\ &\quad + P(A_1{}^c \cap \cdots \cap A_{n-1}^c \cap A_n) \qquad \text{by (1.3.11)}\end{aligned}$$

But

$$\begin{aligned}P(A_1{}^c \cap \cdots \cap A_{i-1}^c \cap A_i) &= P\{R_1 > x, \dots, R_{i-1} > x, R_i \le x\}\\ &= (1 - F_1(x)) \cdots (1 - F_{i-1}(x))F_i(x)\end{aligned}$$

Thus

$$\begin{aligned}F_{T_1}(x) &= F_1(x) + (1 - F_1(x))F_2(x) + (1 - F_1(x))(1 - F_2(x))F_3(x)\\ &\quad + \cdots + (1 - F_1(x)) \cdots (1 - F_{n-1}(x))F_n(x)\end{aligned}$$

Hence

$$\begin{aligned}1 - F_{T_1} &= (1 - F_1)[1 - F_2 - (1 - F_2)F_3 - \cdots\\ &\quad - (1 - F_2) \cdots (1 - F_{n-1})F_n]\\ &= (1 - F_1)(1 - F_2)[1 - F_3 - (1 - F_3)F_4\\ &\quad - \cdots - (1 - F_3) \cdots (1 - F_{n-1})F_n]\\ &= \prod_{i=1}^{n} (1 - F_i)\end{aligned}$$

as above.

We now make the simplifying assumption that the R_i are absolutely continuous (as well as independent), each with the same density f. [Note that

$$P\{R_i = R_j\} = \iint_{x_i = x_j} f(x_i)f(x_j)\, dx_i\, dx_j = 0 \qquad (\text{if } i \ne j)$$

Hence

$$P\{R_i = R_j \text{ for at least one } i \ne j\} \le \sum_{i \ne j} P\{R_i = R_j\} = 0$$

Thus ties occur with probability zero and can be ignored.]

We shall show that the T_k are absolutely continuous, and find the density explicitly. We do this intuitively first. We have

$$\begin{aligned}P\{x < T_k < x + dx\} &= P\{x < T_k < x + dx,\ T_k = R_1\}\\ &\quad + P\{x < T_k < x + dx,\ T_k = R_2\} + \cdots + P\{x < T_k < x + dx,\ T_k = R_n\}\end{aligned}$$

by the theorem of total probability. (The events $\{T_k = R_i\}$, $i = 1, \dots, n$, are mutually exclusive and exhaustive.) Thus

$$\begin{aligned}P\{x < T_k < x + dx\} &= nP\{x < T_k < x + dx,\ T_k = R_1\} \qquad \text{by symmetry}\\ &= nP\{T_k = R_1,\ x < R_1 < x + dx\}\end{aligned}$$

Now for R_1 to be the kth smallest and fall between x and $x + dx$, exactly $k - 1$ of the random variables $R_2, \dots, R_n$ must be $<R_1$, and the remaining $n - k$ must be $>R_1$ [and R_1 must lie in $(x, x + dx)$].

Since there are $\binom{n-1}{k-1}$ ways of selecting $k - 1$ distinct objects out of $n - 1$, we have

$$P\{x < T_k < x + dx\} = n\binom{n-1}{k-1}P\{x < R_1 < x + dx,$$
$$R_2 < R_1, \dots, R_k < R_1, R_{k+1} > R_1, \dots, R_n > R_1\}$$

But if R_1 falls in $(x, x + dx)$, $R_i < R_1$ is essentially the same thing as $R_i < x$, so that

$$\begin{aligned} P\{x < T_k < x + dx\} &= n\binom{n-1}{k-1} f(x)\,dx(P\{R_i < x\})^{k-1}(P\{R_j > x\})^{n-k} \\ &= n\binom{n-1}{k-1} f(x)(F(x))^{k-1}(1 - F(x))^{n-k}\,dx \end{aligned}$$

Since $P\{x < T_k < x + dx\} = f_k(x)\,dx$, where f_k is the density of T_k (assumed to exist), we have

$$f_k(x) = n\binom{n-1}{k-1} f(x)(F(x))^{k-1}(1 - F(x))^{n-k}$$

[When $k = n$ we get $nf(x)(F(x))^{n-1} = (d/dx)F(x)^n$, and when $k = 1$ we get $nf(x)(1 - F(x))^{n-1} = (d/dx)(1 - (1 - F(x))^n)$, in agreement with the previous results if all R_i have distribution function F and the density f can be obtained by differentiating F.]

To obtain the result formally, we reason as follows.

$$\begin{aligned} P\{T_k \le x\} &= \sum_{i=1}^{n} P\{T_k \le x, T_k = R_i\} = nP\{T_k \le x, T_k = R_1\} \\ &= nP\{R_1 \le x, \text{ exactly } k - 1 \text{ of the variables } R_2, \dots, R_n \text{ are } \\ &\qquad <R_1, \text{ and the remaining } n - k \text{ variables are } >R_1\} \\ &= n\binom{n-1}{k-1} P\{R_1 \le x, R_2 < R_1, \dots, R_k < R_1, R_{k+1} > R_1, \dots, \\ &\qquad R_n > R_1\} \quad \text{by symmetry} \\ &= n\binom{n-1}{k-1} \int_{x_1=-\infty}^{x} \int_{x_2=-\infty}^{x_1} \cdots \int_{x_k=-\infty}^{x_1} \int_{x_{k+1}=x_1}^{\infty} \\ &\qquad \cdots \int_{x_n=x_1}^{\infty} f(x_1)\cdots f(x_n)\,dx_1 \cdots dx_n \\ &= \int_{-\infty}^{x} n\binom{n-1}{k-1} f(x_1)(F(x_1))^{k-1}(1 - F(x_1))^{n-k}\,dx_1 \end{aligned}$$

The integrand is the density of T_k, in agreement with the intuitive approach. $T_1, \ldots, T_n$ are called the *order statistics* of $R_1, \ldots, R_n$.

REMARK. All events

$$\{R_{i_1} \le x, R_{i_2} < R_{i_1}, \ldots, R_{i_k} < R_{i_1}, R_{i_{k+1}} > R_{i_1}, \ldots, R_{i_n} > R_{i_1}\}$$

have the same probability, namely,

$$\int_{-\infty}^{x} f(x_{i_1})\, dx_{i_1} \int_{-\infty}^{x_{i_1}} f(x_{i_2})\, dx_{i_2} \cdots \int_{-\infty}^{x_{i_1}} f(x_{i_k})\, dx_{i_k} \int_{x_{i_1}}^{\infty} f(x_{i_{k+1}})\, dx_{i_{k+1}} \cdots \int_{x_{i_1}}^{\infty} f(x_{i_n})\, dx_{i_n}$$

This justifies the appeal to symmetry in the above argument. ◀

PROBLEMS

1. Let R_1 and R_2 be independent and uniformly distributed between 0 and 1. Find and sketch the distribution or density function of the random variable $R_3 = R_2/R_1^2$.
2. If R_1 and R_2 are independent random variables, each with the density function $f(x) = e^{-x}$, $x \ge 0$; $f(x) = 0$, $x < 0$, find and sketch the distribution or density function of the random variable R_3, where
(a) $R_3 = R_1 + R_2$
(b) $R_3 = R_2/R_1$
3. Let R_1 and R_2 be independent, absolutely continuous random variables, each normally distributed with parameters $a = 0$ and $b = 1$; that is,

$$f_1(x) = f_2(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Find and sketch the density or distribution function of the random variable $R_3 = R_2/R_1$.

4. Let R_1 and R_2 be independent, absolutely continuous random variables, each uniformly distributed between 0 and 1. Find and sketch the distribution or density function of the random variable R_3, where

$$R_3 = \frac{\max(R_1, R_2)}{\min(R_1, R_2)}$$

REMARK. The example in which $R_3 = \max(R_1, R_2) + \min(R_1, R_2)$ may occur to the reader. However, this yields nothing new, since

$\max(R_1, R_2) + \min(R_1, R_2) = R_1 + R_2$ (the sum of two numbers is the larger plus the smaller).

5. A point-size worm is inside an apple in the form of the sphere $x^2 + y^2 + z^2 = 4a^2$. (Its position is uniformly distributed.) If the apple is eaten down to a core determined by the intersection of the sphere and the cylinder $x^2 + y^2 = a^2$, find the probability that the worm will be eaten.

6. A point (R_1, R_2, R_3) is uniformly distributed over the region in E^3 described by $x^2 + y^2 \le 4$, $0 \le z \le 3x$. Find the probability that $R_3 \le 2R_1$.

7. Solve Problem 6 under the assumption that (R_1, R_2, R_3) has density $f(x, y, z) = kz^2$ over the given region and $f(x, y, z) = 0$ outside the region.

8. Let $T_1, \dots, T_n$ be the order statistics of $R_1, \dots, R_n$, where $R_1, \dots, R_n$ are independent, each with density f. Show that the joint density of $T_1, \dots, T_n$ is given by

$$\begin{aligned} g(x_1, \dots, x_n) &= n!\,f(x_1)\cdots f(x_n), \qquad x_1 < x_2 < \cdots < x_n \\ &= 0 \qquad \text{elsewhere} \end{aligned}$$

HINT: Find $P\{T_1 \le b_1, \dots, T_n \le b_n, R_1 < R_2 < \cdots < R_n\}$.

9. Let R_1, R_2, and R_3 be independent, each with density

$$\begin{aligned} f(x) &= e^{-x}, \qquad x \ge 0 \\ &= 0, \qquad x < 0 \end{aligned}$$

Find the probability that $R_1 \ge 2R_2 \ge 3R_3$.

10. A man and a woman agree to meet at a certain place some time between 11 and 12 o'clock. They agree that the one arriving first will wait z hours, $0 \le z \le 1$, for the other to arrive. Assuming that the arrival times are independent and uniformly distributed, find the probability that they will meet.

11. If n points $R_1, \dots, R_n$ are picked independently and with uniform density on a straight line of length L, find the probability that no two points will be less than distance d apart; that is, find

$$P\{\min_{i \ne j} |R_i - R_j| \ge d\}$$

HINT: First find $P\{\min_{i \ne j} |R_i - R_j| \ge d, R_1 < R_2 < \cdots < R_n\}$; show that the region of integration defined by this event is

$$\begin{aligned} x_{n-1} + d \le x_n &\le L \\ x_{n-2} + d \le x_{n-1} &\le L - d \\ x_{n-3} + d \le x_{n-2} &\le L - 2d \\ &\vdots \\ x_1 + d \le x_2 &\le L - (n-2)d \\ 0 \le x_1 &\le L - (n-1)d \end{aligned}$$

12. (The density function method for functions of more than one random variable.) Let $(R_1, \dots, R_n)$ be absolutely continuous with density $f_{12\cdots n}(x_1, \dots, x_n)$. Define random variables $W_1, \dots, W_n$ by $W_i = g_i(R_1, \dots, R_n)$, $i = 1, 2, \dots, n$; thus $(W_1, \dots, W_n) = g(R_1, \dots, R_n)$. Assume that g is one-to-one, continuously differentiable with a nonzero Jacobian J_g (hence g has a continuously differentiable inverse h). Show that $(W_1, \dots, W_n)$ is absolutely continuous with density

$$\begin{aligned} f^*_{12\cdots n}(\mathbf{y}) &= f_{12\cdots n}(h(\mathbf{y}))\,|J_h(\mathbf{y})|, \qquad \mathbf{y} = (y_1, \dots, y_n) \\ &= \frac{f_{12\cdots n}(h(\mathbf{y}))}{|J_g(\mathbf{x})|_{\mathbf{x}=h(\mathbf{y})}} \end{aligned}$$

[The result is the same if g is defined only on some open subset D of E^n and $P\{(R_1, \dots, R_n) \in D\} = 1$.]

13. Let R_1 and R_2 be independent random variables, each normally distributed with $a = 0$ and the same b. Define random variables R_0 and θ_0 by

$$\begin{aligned} R_1 &= R_0 \cos \theta_0 \qquad (\text{taking } R_0 \ge 0) \\ R_2 &= R_0 \sin \theta_0 \end{aligned}$$

Show that R_0 and θ_0 are independent, and find their density functions.

14. Let R_1 and R_2 be independent, absolutely continuous, positive random variables and let $R_3 = R_1 R_2$. Show that the density function of R_3 is given by

$$\begin{aligned} f_3(z) &= \int_0^\infty \frac{1}{w} f_1\left(\frac{z}{w}\right) f_2(w)\, dw, \qquad z > 0 \\ &= 0, \qquad z < 0 \end{aligned}$$

Note: This problem may be done by the distribution function method or by applying Problem 12 as follows.

$$\begin{aligned} R_3 &= R_1 R_2 \\ R_4 &= R_2 \end{aligned}$$

Use the results of Problem 12 to obtain $f_{34}(z, w)$ and from this find $f_3(z)$.

15. Because of inefficiency of production, the resistances R_1 and R_2 in Figure P.2.8.15 may be regarded as independent random variables, each uniformly

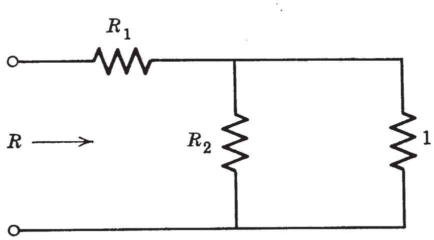


FIGURE P.2.8.15

distributed between 0 and 1 ohm. Find the probability that the total resistance R of the network is $\leq \frac{1}{2}$ ohm.

16. A chamber consists of the inside of the cylinder $x^2 + y^2 = 1$. A particle at the origin is given initial velocity components $v_x = R_1$ and $v_y = R_2$, where R_1 and R_2 are independent random variables, each with normal density $f(x) = (2\pi)^{-1/2} e^{-x^2/2}$. (There is no motion in the z-direction, and no force acting on the particle after the initial "push" at time $t = 0$.) If T is the time at which the particle strikes the wall of the chamber, find the distribution and density functions of T.

2.9 SOME DISCRETE EXAMPLES

In this section we examine some typical problems involving one or more discrete random variables. We first introduce the Poisson distribution, which may be regarded as an approximation to the binomial when the number n of trials is large and the probability p of success on a given trial is small.

Let R_n be the number of successes in n Bernoulli trials, with probability p_n of success on a given trial. We have seen (Section 1.5) that R_n has the *binomial distribution*;† that is, the probability function of R_n is

$$p_{R_n}(k) = \binom{n}{k} p_n^{\,k}(1 - p_n)^{n-k}, \qquad k = 0, 1, \ldots, n$$

We now let $n \to \infty$, $p_n \to 0$ in such a way that $np_n \to \lambda =$ constant. We shall show that

$$p_{R_n}(k) \to \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, \ldots$$

To see this, write

$$\begin{aligned} p_{R_n}(k) &= \frac{n(n-1)\cdots(n-k+1)}{k!} p_n^{\,k}(1 - p_n)^{n-k} \\ &= \frac{(1 - 1/n)(1 - 2/n)\cdots(1 - (k-1)/n)}{k!} (np_n)^k \left(1 - \frac{np_n}{n}\right)^{n-k} \end{aligned}$$

Now $(1 - np_n/n)^{-k} \to 1$ and $(1 - np_n/n)^n \to e^{-\lambda}$ (Problem 1), and the result follows.

We call

$$p(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots \tag{2.9.1}$$

† When a probabilist says he knows the *distribution* of a random variable R, he generally means that he has some way of calculating $P\{R \in B\}$ for all Borel sets B. For example, he might know the distribution function of R, or the probability function if R is discrete, or the density function if R is absolutely continuous. Thus to say that R has the normal distribution means that R has a density given by the formula (2.8.1).

the *Poisson* probability function; a random variable which has this probability function is said to have the *Poisson distribution.* (To check that it is a legitimate probability function:

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2!} + \cdots\right] = e^{-\lambda}e^{\lambda} = 1)$$

We shall show that if R_1 and R_2 are independent, each having the Poisson distribution, then $R_1 + R_2$ also has the Poisson distribution. We first need a characterization of independence in the discrete case.

Theorem 1. *Let $R_1, \dots, R_n$ be discrete random variables on a given probability space, with probability functions $p_1, \dots, p_n$. Let $p_{12\cdots n}$ be the joint probability function of $R_1, \dots, R_n$, defined by*

$$p_{12\cdots n}(x_1, \dots, x_n) = P\{R_1 = x_1, \dots, R_n = x_n\} \tag{2.9.2}$$

Then $R_1, \dots, R_n$ are independent if and only if

$$p_{12\cdots n}(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n) \qquad \textit{for all } x_1, \dots, x_n$$

PROOF. If $R_1, \dots, R_n$ are independent, then

$$\begin{aligned} p_{12\cdots n}(x_1, \dots, x_n) &= P\{R_1 = x_1, \dots, R_n = x_n\} \\ &= P\{R_1 = x_1\} \cdots P\{R_n = x_n\} \qquad \text{by independence} \\ &= p_1(x_1) \cdots p_n(x_n) \end{aligned}$$

Conversely, if $p_{12\cdots n}(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n)$, then for all one-dimensional Borel sets $B_1, \dots, B_n$,

$$\begin{aligned} P\{R_1 \in B_1, \dots, R_n \in B_n\} &= \sum_{x_1 \in B_1, \dots, x_n \in B_n} P\{R_1 = x_1, \dots, R_n = x_n\} \\ &= \sum_{x_1 \in B_1, \dots, x_n \in B_n} p_1(x_1) \cdots p_n(x_n) \\ &= \sum_{x_1 \in B_1} p_1(x_1) \cdots \sum_{x_n \in B_n} p_n(x_n) \\ &= P\{R_1 \in B_1\} \cdots P\{R_n \in B_n\} \end{aligned}$$

Hence $R_1, \dots, R_n$ are independent.

REMARKS. If R_1 and R_2 are not independent, the joint probability function of R_1 and R_2 is not determined by the individual probability functions. For example, if $P\{R_1 = 1,\ R_2 = 1\} = P\{R_1 = 2,\ R_2 = 2\} = a$,

$P\{R_1 = 1, R_2 = 2\} = P\{R_1 = 2, R_2 = 1\} = \frac{1}{2} - a$, $0 \le a \le \frac{1}{2}$, then $P\{R_1 = 1\} = P\{R_1 = 2\} = \frac{1}{2}$ and $P\{R_2 = 1\} = P\{R_2 = 2\} = \frac{1}{2}$. Thus we have uncountably many joint probability functions giving rise to the same individual probability functions.

If we wish to define discrete random variables $R_1, \ldots, R_n$ having a specified joint probability function $p_{12\cdots n}$, there is no difficulty in constructing an appropriate probability space. Take $\Omega = E^n$, $\mathscr{F}$ = all subsets of Ω (since the random variables are discrete, there is no need to restrict to Borel sets), $P(B) = \sum_{(x_1,\ldots,x_n)\in B} p_{12\cdots n}(x_1, \ldots, x_n)$, $B \in \mathscr{F}$.

Now let R_1 and R_2 be independent, with R_i having the Poisson distribution with parameter λ_i, $i = 1, 2$. By Theorem 1, the joint probability function of R_1 and R_2 is

$$p_{12}(j, k) = P\{R_1 = j, R_2 = k\} = e^{-(\lambda_1+\lambda_2)} \frac{\lambda_1^{\,j}\lambda_2^{\,k}}{j!\,k!}$$

We find the probability function of $R_1 + R_2$.

$$\begin{aligned} P\{R_1 + R_2 = m\} &= \sum_{j+k=m} p_{12}(j, k) \\ &= e^{-(\lambda_1+\lambda_2)} \sum_{j=0}^{m} \frac{\lambda_1^{j}\,\lambda_2^{m-j}}{j!\,(m-j)!} \end{aligned}$$

But $(\lambda_1 + \lambda_2)^m = \sum_{j=0}^{m} \binom{m}{j} \lambda_1^j \lambda_2^{m-j}$ by the binomial theorem, so that

$$P\{R_1 + R_2 = m\} = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^m}{m!}, \qquad m = 0, 1, \ldots$$

Thus $R_1 + R_2$ has the Poisson distribution with parameter $\lambda_1 + \lambda_2$.

By induction, it follows that the sum of n independent random variables $R_1, \ldots, R_n$, where R_i is Poisson with parameter λ_i, has the Poisson distribution with parameter $\lambda_1 + \cdots + \lambda_n$.

The use of the Poisson distribution as an approximation to the binomial is illustrated in the problems.

▶ **Example 1.** Six unbiased dice are tossed independently. Let R_1 be the number of ones, R_2 the number of twos; R_1 and R_2 have the binomial distribution with $n = 6$, $p = 1/6$; that is,

$$p_1(k) = p_2(k) = \binom{6}{k}\left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{6-k}, \qquad k = 0, 1, 2, 3, 4, 5, 6$$

Let us find the joint probability function $p_{12}(j, k)$ of R_1 and R_2. This is a multinomial problem.

$$\begin{array}{llll} b_1 = \text{“1 occurs” on a given toss} & p_1 = \frac{1}{6} & n_1 = j & \\ b_2 = \text{“2 occurs”} & p_2 = \frac{1}{6} & n_2 = k & n = 6 \\ b_3 = \text{“3, 4, 5, or 6 occurs”} & p_3 = \frac{2}{3} & n_3 = 6 - j - k & \end{array}$$

Thus

$$p_{12}(j, k) = \frac{6!}{j!\,k!\,(6-j-k)!}\left(\frac{1}{6}\right)^{j+k}\left(\frac{2}{3}\right)^{6-j-k},$$

$$j, k = 0, 1, 2, 3, 4, 5, 6; j + k \le 6$$

Thus the multinomial formula appears as the joint probability function of a number of random variables, each of which is individually binomial.

Now let us find the conditional probability function of R_1 given R_2; that is,

$$\begin{aligned} p_1(j \mid k) &= P\{R_1 = j \mid R_2 = k\} \\ &= \frac{p_{12}(j, k)}{p_2(k)} = \frac{[6!/j!\,k!\,(6-j-k)!](1/6)^{j+k}(4/6)^{6-j-k}}{[6!/k!\,(6-k)!](1/6)^{k}(5/6)^{6-k}} \\ &= \frac{(6-k)!}{j!\,(6-j-k)!}\,\frac{4^{6-j-k}}{5^{6-k}}\left(\frac{5^{-j}}{5^{-j}}\right) = \binom{6-k}{j}\left(\frac{1}{5}\right)^{j}\left(\frac{4}{5}\right)^{6-k-j} \end{aligned}$$

Intuitively, given $R_2 = k$, there are $6 - k$ remaining tosses. The possible outcomes are 1, 3, 4, 5, or 6 (2 is not permitted), all equally likely. Thus, given $R_2 = k$, R_1 should be binomial with $n = 6 - k$, $p = 1/5$. This is verified by the formal calculation above.

REMARK. Since the discrete random variables R_1 and R_2 are independent iff $p_{12}(j, k) = p_1(j)p_2(k)$ for all j, k, it follows that independence is equivalent to $p_1(j \mid k) = p_1(j)$ for all j, k [such that $p_2(k) > 0$].

In the present case $p_1(j \mid k)$ is the binomial probability function with $n = 6 - k$, $p = 1/5$, and $p_1(j)$ is the binomial probability function with $n = 6$, $p = 1/6$. Thus R_1 and R_2 are not independent. This is clear intuitively; for example, if we know that $R_1 = 6$, the odds about R_2 are certainly affected; in fact, R_2 must be 0. ◀

PROBLEMS

1. (a) If $|x| \le 1/2$, show that

$$\ln(1 + x) = x + \theta x^2$$

where $|\theta| \le 1$, θ depending on x.

(b) Show that if $x_n \to \lambda$, then

$$\left(1 - \frac{x_n}{n}\right)^n \to e^{-\lambda}$$

2. If R has the binomial distribution with n large and p small, the Poisson approximation with $\lambda = np$ may be used (a rule of thumb that has been given is that the approximation will be good to several decimal places if $n \geq 100$ and $p \leq .01$). Feller (An Introduction to Probability Theory and Its Applications, vol. 1, John Wiley and Sons, 1950) gives several examples of such random variables:

(i) The number of color-blind people in a large group (or the number of people possessing some other rare characteristic).
(ii) The number of misprints on a page.
(iii) The number of radioactive particles (or particles with some other distinguishing characteristic) passing through a counting device in a given time interval.
(iv) The number of flying bomb hits on a particular area of London during World War II (n is the number of bombs in a given period of time, p the probability that a single bomb will hit the area).
(v) The number of raisins in a cookie.

[Here the assumptions are not entirely clear. Perhaps what is envisioned is that the dough is bombarded by a raisin gun at some stage in the cookie-making process. It would seem that this is simply a peaceful version of example (iv).]

In the following exercises, use the Poisson approximation to calculate the probabilities.

(a) If $p = .001$, how large must n be if $P\{R \geq 1\} \geq .99$?
(b) If $np = 2$, find $P\{R \geq 3\}$.

3. The joint probability function of two discrete random variables R_1 and R_2 is as follows:

$$\begin{aligned} p_{12}(1,1) &= .4 \\ p_{12}(1,2) &= .3 \\ p_{12}(2,1) &= .2 \\ p_{12}(2,2) &= .1 \\ p_{12}(j,k) &= 0 \qquad \text{elsewhere} \end{aligned}$$

(a) Determine whether or not R_1 and R_2 are independent.
(b) Find the probability that $R_1R_2 \leq 2$.

4. Let R_1 and R_2 be independent; assume that R_1 has the binomial distribution with parameters n and p, and R_2 has the binomial distribution with parameters m and p. Find $P\{R_1 = j \mid R_1 + R_2 = k\}$, and interpret the result intuitively. [Note: one approach involves establishing the formula $\binom{n+m}{k} = \sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i}$.]