CHAPTER TWO

random variables

Often we have reasons to associate one or more numbers (in addition to probabilities) with each possible outcome of an experiment. Such numbers might correspond, for instance, to the cost to us of each experimental outcome, the amount of rainfall during a particular month, or the height and weight of the next football player we meet.

This chapter extends and specializes our earlier work to develop effective methods for the study of experiments whose outcomes may be described numerically.

2-1 Random Variables and Their Event Spaces

For the study of experiments whose outcomes may be specified numerically, we find it useful to introduce the following definition:

A *random variable* is defined by a function which assigns a value of the random variable to each sample point in the sample space of an experiment.

Each performance of the experiment is said to generate an *experimental value* of the random variable. This experimental value of the random variable is equal to the value of the random variable assigned to the sample point which corresponds to the resulting experimental outcome.

Consider the following example, which will be referred to in several sections of this chapter. Our experiment consists of three independent flips of a fair coin. We again use the notation

$$\text{Event } \begin{Bmatrix} H_n \\ T_n \end{Bmatrix} : \begin{Bmatrix} \text{Heads} \\ \text{Tails} \end{Bmatrix} \text{ on the } n\text{th flip}$$

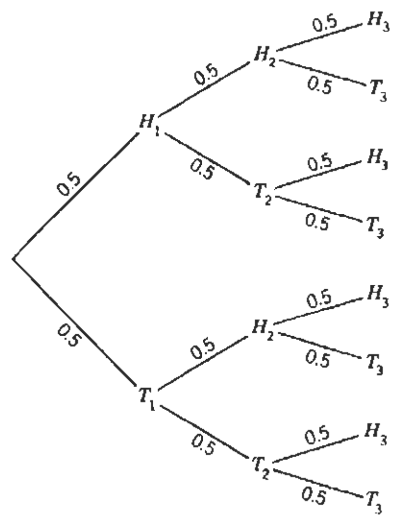
We may define any number of random variables on the sample space of this experiment. We choose the following definitions for two random variables, h and r:

h = total number of heads resulting from the three flips
r = length of longest run resulting from the three flips (a run is a set of successive flips all of which have the same outcome)

We now prepare a fully labeled sequential sample space for this experiment. We include the branch traversal conditional probabilities, the probability of each experimental outcome, and the values of random variables h and r assigned to each sample point. The resulting sample space is shown at the top of the following page.

If this experiment were performed once and the experimental outcome were the event $H_1T_2T_3$, we would say that, for this performance of the experiment, the resulting experimental values of random variables h and r were 1 and 2, respectively.

Although we may require the full sample space to describe the detailed probabilistic structure of an experiment, it may be that our only practical interest in each performance of the experiment will relate to the resulting experimental values of one or more random variables.



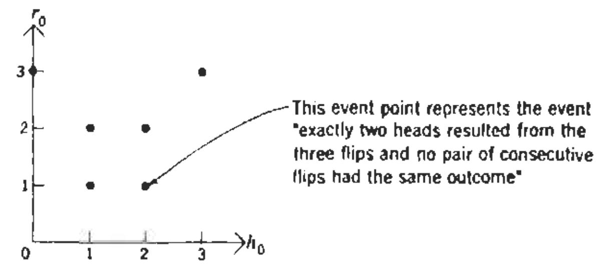
Sample points	$P(\bullet)$	h	r
• $H_1H_2H_3$	0.125	3	3
• $H_1H_2T_3$	0.125	2	2
• $H_1T_2H_3$	0.125	2	1
• $H_1T_2T_3$	0.125	1	2
• $T_1H_2H_3$	0.125	2	2
• $T_1H_2T_3$	0.125	1	1
• $T_1T_2H_3$	0.125	1	2
• $T_1T_2T_3$	0.125	0	3

When this is the case, we may prefer to work in an *event* space which distinguishes among outcomes only in terms of the possible experimental values of the random variables of interest. Let's consider this for the above example.

Suppose that our only interest in a performance of the experiment has to do with the resulting experimental value of random variable h. We might find it desirable to work with this variable in an event space of the form

The four event points marked along the h_0 axis form a mutually exclusive collectively exhaustive listing of all possible experimental outcomes. The event point at any h_0 corresponds to the event "The experimental value of random variable h generated on a performance of the experiment is equal to h_0," or, in other words, "On a performance of the experiment, random variable h takes on experimental value h_0."

Similarly, if our concern with each performance of the experiment depended only upon the resulting experimental values of random variables h and r, a simple event space would be



An event point in this space with coordinates h_0 and r_0 corresponds to the event "On a performance of the experiment, random variables h and r take on, respectively, experimental values h_0 and r_0." The probability assignment for each of these six event points may, of course, be obtained by collecting these events and their probabilities in the original sequential sample space.

The random variables discussed in our example could take on only experimental values selected from a set of discrete numbers. Such random variables are known as *discrete* random variables. Random variables of another type, known as *continuous* random variables, may take on experimental values anywhere within continuous ranges. Examples of continuous random variables are the exact instantaneous voltage of a noise signal and the precise reading after a spin of an infinitely finely calibrated wheel of fortune (as in the last example of Sec. 1-2).

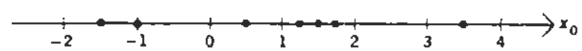
Formally, the distinction between discrete and continuous random variables can be avoided. But the development of our topics is easier to visualize if we first become familiar with matters with regard to discrete random variables and later extend our coverage to the continuous case. Our discussions through Sec. 2-8 deal only with the discrete case, and Sec. 2-9 begins the extension to the continuous case.

2-2 The Probability Mass Function

We have learned that a random variable is defined by a function which assigns a value of that random variable to each sample point. These assigned values of the random variable are said to represent its possible experimental values. Each performance of the experiment generates an experimental value of the random variable. For many purposes, we shall find the resulting experimental value of a random variable to be an adequate characterization of the experimental outcome.

In the previous section, we indicated the form of a simple event space for dealing with a single discrete random variable. To work with a random variable x, we mark, on an x_0 axis, the points corresponding

to all possible experimental values of the random variable. One such event space could be

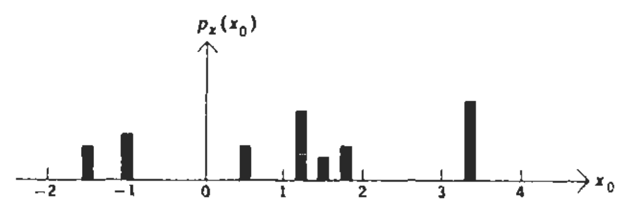


Each point in this event space corresponds to the event "On a performance of the experiment, the resulting experimental value of random variable x is equal to the indicated value of x_0."

We next define a function on this event space which assigns a probability to each event point. The function $p_x(x_0)$ is known as the *probability mass function* (PMF) for discrete random variable x, defined by

$p_x(x_0)$ = probability that the experimental value of random variable x obtained on a performance of the experiment is equal to x_0

We often present the probability mass function as a bar graph drawn over an event space for the random variable. One possible PMF is sketched below:



Since there must be some value of random variable x associated with every sample point, we must have

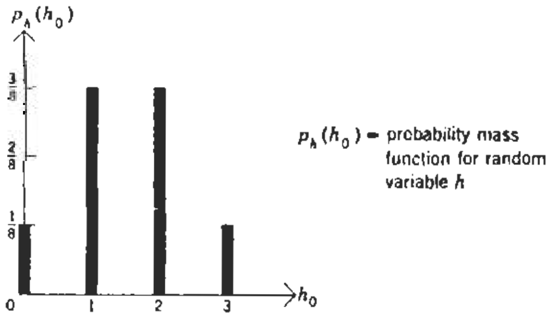
$$\sum_{x_0} p_x(x_0) = 1$$

and, of course, from the axioms of probability theory we also have

$$0 \le p_x(x_0) \le 1 \qquad \text{for all values of } x_0$$

Note that the argument of a PMF is a dummy variable, and the PMF for random variable x could also be written as $p_x(y)$, p_x(☛), or, as some people prefer, $p_x(\cdot)$. We shall generally use the notation $p_x(x_0)$ for a PMF. However, when another notation offers advantages of clarity or brevity for the detailed study of a particular process, we shall adopt it for that purpose.

For an example of the PMF for a random variable, let's return to the experiment of three flips of a fair coin, introduced in the previous section. We may go back to the original sample space of that experiment to collect $p_h(h_0)$, the PMF for the total number of heads resulting from the three flips. We obtain



2-3 Compound Probability Mass Functions

We wish to consider situations in which values of more than one random variable are assigned to each point in the sample space of an experiment. Our discussion will be for two discrete random variables, but the extension to more than two is apparent.

For a performance of an experiment, the probability that random variable x will take on experimental value x_0 *and* random variable y will take on experimental value y_0 may be determined in sample space by summing the probabilities of each sample point which has this compound attribute. To designate the probability assignment in an x_0,y_0 event space, we extend our previous work to define the *compound* (or *joint*) PMF for two random variables x and y,

$p_{x,y}(x_0,y_0)$ = probability that the experimental values of random variables x and y obtained on a performance of the experiment are equal to x_0 and y_0, respectively

A picture of this function would have the possible event points marked on an x_0,y_0 coordinate system with each value of $p_{x,y}(x_0,y_0)$ indicated as a bar perpendicular to the x_0,y_0 plane above each event point. [We use the word *event* point here since each possible (x_0,y_0) point might represent the union of several sample points in the finest-grain description of the experiment.]

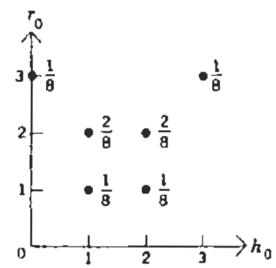
By considering an x_0,y_0 event space and recalling that an event space is a mutually exclusive, collectively exhaustive listing of all possible experimental outcomes, we see that the following relations hold:

$$\sum_{x_0}\sum_{y_0} p_{x,y}(x_0,y_0) = 1$$

$$\sum_{x_0} p_{x,y}(x_0,y_0) = p_y(y_0) \qquad \sum_{y_0} p_{x,y}(x_0,y_0) = p_x(x_0)$$

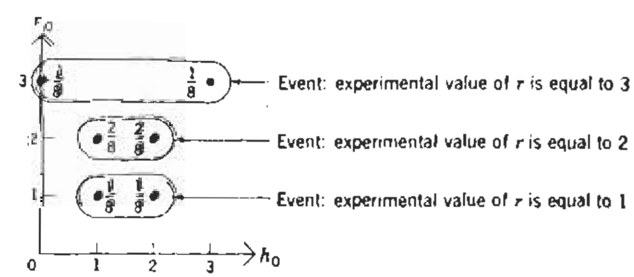
In situations where we are concerned with more than one random variable, the PMF's for single random variables, such as $p_x(x_0)$, are referred to as *marginal* PMF's. No matter how many random variables may be defined on the sample space of the experiment, this function $p_x(x_0)$ always has the same physical interpretation. For instance, $p_x(2)$ is the probability that the experimental value of discrete random variable x resulting from a performance of the experiment will be equal to 2.

Let's return to the example of three flips of a fair coin in Sec. 2-1 to obtain the compound PMF for random variables h (number of heads) and r (length of longest run). By collecting the events of interest and their probabilities from the sequential sample space in Sec. 2-1, we obtain $p_{h,r}(h_0,r_0)$. We indicate the value of $p_{h,r}(h_0,r_0)$ associated with each event by writing it beside the appropriate event point.

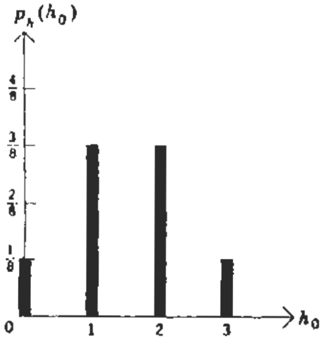


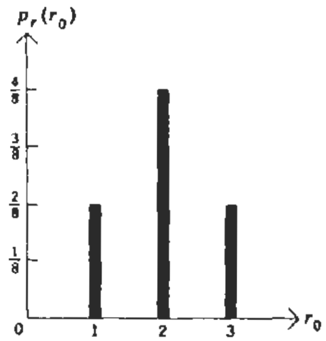
The probability of *any* event described in terms of the experimental values of random variables h and r may be found easily in this event space once $p_{h,r}(h_0,r_0)$ has been determined.

For instance, we may obtain the marginal PMF's, $p_h(h_0)$ and $p_r(r_0)$, simply by collecting the probabilities of the appropriate events in the h_0,r_0 sample space.



The reader can check that the above procedure and a similar operation for random variable h lead to the marginal PMF's





It is important to note that, in general, there is no way to go back from the marginal PMF's to determine the compound PMF.

2-4 Conditional Probability Mass Functions

Our interpretation of conditional probability in Chap. 1 noted that conditioning a sample space does one of two things to the probability of each finest-grain event. If an event does not have the attribute of the conditioning event, the conditional probability of that event is set to zero. For all finest-grain events having the attribute of the conditioning event, the conditional probability associated with each such event is equal to its original probability scaled up by a constant $[1/P(A)$, where A is the conditioning event$]$ such that the sum of the conditional probabilities in the conditional sample space is unity. We can use the same concepts in an event space, as long as we can answer either "yes" or "no" for each event point to the question "Does this event have the attribute of the conditioning event?" Difficulty would arise only when the conditioning event was of finer grain than the event space. This matter was discussed near the end of Sec. 1-4.

When we consider a discrete random variable taking on a particular experimental value as a result of a performance of an experiment, this is simply an event like those we denoted earlier as A, B, or anything else and all our notions of conditional probability are carried over to the discussion of discrete random variables. We define the conditional PMF by

$p_{x|y}(x_0 \mid y_0)$ = conditional probability that the experimental value of random variable x is x_0, given that, on the same performance of the experiment, the experimental value of random variable y is y_0

From the definition of conditional probability, there follows

$$p_{x|y}(x_0 \mid y_0) = \frac{p_{x,y}(x_0,y_0)}{p_y(y_0)} \qquad \text{and, similarly,} \qquad p_{y|x}(y_0 \mid x_0) = \frac{p_{x,y}(x_0,y_0)}{p_x(x_0)}$$

As was the case in the previous chapter, the conditional probabilities are not defined for the case where the conditioning event is of probability zero.

Writing these definitions another way, we have

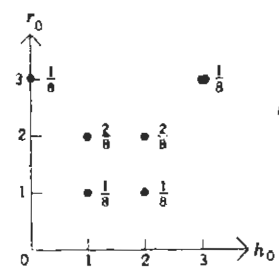
$$p_{x,y}(x_0,y_0) = p_x(x_0)p_{y|x}(y_0 \mid x_0) = p_y(y_0)p_{x|y}(x_0 \mid y_0)$$

Notice that, in general, the marginal PMF's $p_x(x_0)$ and $p_y(y_0)$ *do not* specify $p_{x,y}(x_0,y_0)$ just as, in general, $p(A)$ and $p(B)$ do not specify $p(AB)$.

Finally, we need a notation for a conditional joint probability mass function where the conditioning event is other than an observed experimental value of one of the random variables. We shall use $p_{x,y|A}(x_0,y_0 \mid A)$ to denote the conditional compound PMF for random variables x and y given event A. This is, by the definition of conditional probability,

$$p_{x,y|A}(x_0,y_0 \mid A) = \begin{cases} \dfrac{p_{x,y}(x_0,y_0)}{P(A)} & \text{if } (x_0,y_0) \text{ in } A \\ 0 & \text{if } (x_0,y_0) \text{ in } A' \end{cases}$$

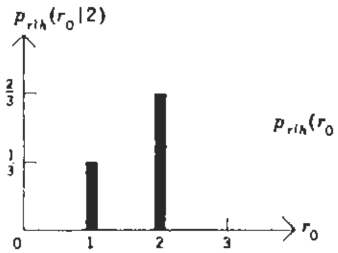
We return to the h_0,r_0 event space of the previous sections and its compound PMF in order to obtain some experience with conditional probability mass functions



$h_0\,r_0$ event space including value of joint PMF for each event point

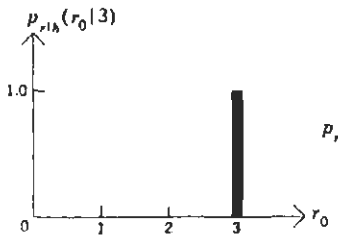
We begin by finding the conditional PMF for random variable r, the length of the longest run obtained in three flips, given that the experimental value of h, the number of heads, is equal to 2. Thus, we wish to find $p_{r|h}(r_0 \mid 2)$. Only two of the event points in the original h_0,r_0

event space have the attribute of the conditioning event ($h_0 = 2$), and the relative likelihood of these points must remain the same in the conditional event space. Either by direct use of the definition or from our reasoning in sample space, there results



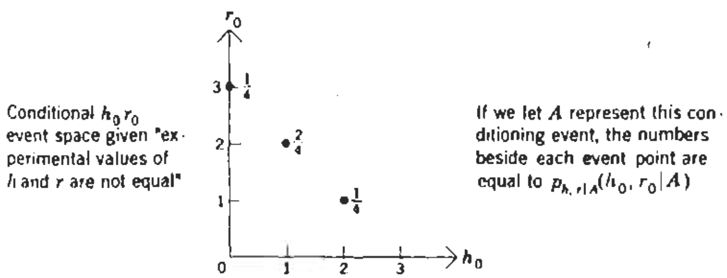
$$p_{r|h}(r_0|2) = \begin{cases} \frac{1}{3} & r_0 = 1 \\ \frac{2}{3} & r_0 = 2 \\ 0 & \text{for all other values of } r_0 \end{cases}$$

Had the conditioning event been that the experimental value of h were equal to 3, this would specify the experimental value of r for this experiment, because there is only one possible h_0,r_0 event point with $h_0 = 3$. The resulting conditional PMF would, of course, be



$$p_{r|h}(r_0|3) = \begin{cases} 1 & r_0 = 3 \\ 0 & r_0 \neq 3 \end{cases}$$

Suppose that we wish to condition the compound PMF in the h_0,r_0 event space by the event that the experimental values of h and r resulting from a performance of the experiment are not equal. In going to the appropriate conditional event space and allocation of conditional probability, we remove event points incompatible with the conditioning event and renormalize to obtain



Finally, we can note one reasonable example of a phenomenon which was mentioned earlier. We have stated that we cannot always directly condition an *event* space by an arbitrary event defined on the experiment. For our example, if we were told that the second flip resulted in heads, our simple conditioning argument cannot be applied in an h_0,r_0 event space because we can't answer uniquely "yes" or "no" as to whether or not each event point has this attribute. The conditioning requires information which appeared in the sequential *sample* space of Sec. 2-1 but which is no longer available in an h_0,r_0 *event* space.

2-5 Independence and Conditional Independence of Random Variables

In Sec. 1-6 we obtained our definition of the independence of two events by stating formally an intuitive notion of independence. For two random variables to be independent, we shall require that *no* possible experimental value of one random variable be able to give us any new information about the probability of *any* experimental value of the other random variable. A formal statement of this notion of the independence of two random variables is

Random variables x and y are defined to be independent if and only if $p_{y|x}(y_0 \mid x_0) = p_y(y_0)$ for all possible values of x_0 and y_0.

From the definition of the conditional PMF's, as long as the conditioning event is of nonzero probability, we may always write

$$p_{x,y}(x_0,y_0) = p_x(x_0)p_{y|x}(y_0 \mid x_0) = p_y(y_0)p_{x|y}(x_0 \mid y_0)$$

and, substituting the above definition of independence into this equation, we find that $p_{y|x}(y_0 \mid x_0) = p_y(y_0)$ for all x_0,y_0 requires that $p_{x|y}(x_0 \mid y_0) = p_x(x_0)$ for all x_0,y_0; thus, one equivalent definition of the independence condition would be to state that random variables x and y are independent if and only if $p_{x,y}(x_0,y_0) = p_x(x_0)p_y(y_0)$ for all x_0,y_0.

We define any number of random variables to be *mutually* independent if the compound PMF for all the random variables factors into the product of all the marginal PMF's for all arguments of the compound PMF.

It is also convenient to define the notion of conditional independence. One of several equivalent definitions is

Random variables x and y are *defined* to be conditionally independent given event A [with $P(A) \neq 0$] if and only if

$$p_{x,y|A}(x_0,y_0 \mid A) = p_{x|A}(x_0 \mid A)p_{y|A}(y_0 \mid A) \qquad \text{for all } (x_0,y_0)$$

Of course, the previous unconditional definition may be obtained by setting $A = U$ in the conditional definition. The function $p_{x|A}(x_0 \mid A)$ is referred to as the *conditional marginal* PMF for random variable x given that the experimental outcome on the performance of the experiment had attribute A.

We shall learn later that the definition of independence has implications beyond the obvious one here. In studying situations involving several random variables, it will normally be the case that, if the random variables are mutually independent, the analysis will be greatly simplified and several powerful theorems will apply.

The type of independence we have defined in this section is often referred to as *true*, or *statistical*, independence of random variables. These words are used to differentiate between this complete form of independence and a condition known as *linear* independence. The latter will be defined in Sec. 2-7.

The reader may wish to use our three-flip experiment and its h_0,r_0 event space to verify that, although h and r are clearly not independent in their original event space, they are conditionally independent given that the longest run was shorter than three flips.

2-6 Functions of Random Variables

A function of some random variables is just what the name indicates—it is a function whose experimental value is determined by the experimental values of the random variables. For instance, let h and r again be the number of heads and the length of the longest run for three flips of a fair coin. Some functions of these random variables are

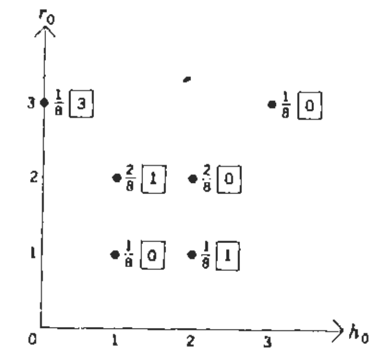
$$v(h,r) = h^2 \qquad w(h,r) = |h - r| \qquad x(h,r) = e^{-h} \log (r \cos h)$$

$$y(h,r) = \max\left(\frac{h}{r}, \frac{r}{2h}\right) \qquad z(h,r) = \begin{cases} h + r & r < 2h \\ 3h - r & r \geq 2h \end{cases}$$

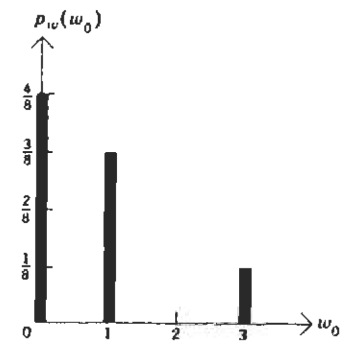
and, of course, h and r.

Functions of random variables are thus new random variables themselves. The experimental values of these new random variables may be displayed in the event space of the original random variables, for instance, by adding some additional markers beside the event points. Once this is done, it is a simple matter to assemble the PMF for a function of the original random variables.

For example, let random variable w be defined by $w = |h - r|$. We'll write the value of w assigned to each event point in a box beside the point in the h_0,r_0 event space,



and then working in the above event space, we can rapidly collect $p_w(w_0)$, to obtain



2-7 Expectation and Conditional Expectation

Let x be a random variable, and let $g(x)$ be any single-valued function of its argument. Then $g(x)$ is a function of a random variable and is itself a random variable. We define $E[g(x)]$, the *expectation*, or *expected value*, of this function of random variable x, to be

$$E[g(x)] = \sum_{x_0} g(x_0)p_x(x_0) = \overline{g(x)}$$

and we also define $E[g(x) \mid A]$, the conditional expectation of $g(x)$ given that the experimental outcome has attribute A, to be

$$E[g(x) \mid A] = \sum_{x_0} g(x_0)p_{x|A}(x_0 \mid A) = \overline{g(x \mid A)}$$

As usual, this definition for the conditional case includes the unconditional case (obtained by setting A equal to the universal event).

Consider the event-space interpretation of the definition of $E[g(x)]$ in an x_0 event space. For each event point x_0, we multiply $g(x_0)$ by the probability of the event point representing that experimental outcome, and then we sum all such products. Thus, the expected value of $g(x)$ is simply the weighted sum of all possible experimental values of $g(x)$, each weighted by its probability of occurrence on a performance of the experiment. We might anticipate a close relationship between $E[g(x)]$ and the average of the experimental values of $g(x)$ generated by many performances of the experiment. This type of relation will be studied in the last two chapters of this book.

Certain cases of $g(x)$ give rise to expectations of frequent interest and these expectations have their own names.

If $g(x) = x^n$:

$$E[g(x)] = \sum_{x_0} x_0{}^n p_x(x_0) = \overline{x^n}$$

If $g(x) = [x - E(x)]^n$:

$$E[g(x)] = \sum_{x_0} [x_0 - E(x)]^n p_x(x_0) = \overline{(x - \bar{x})^n}$$

The quantity $\overline{x^n}$ is known as the *nth moment*, and the quantity $\overline{(x - \bar{x})^n}$ is known as the *nth central moment* of random variable x.

Often we desire a few simple parameters to characterize the PMF for a particular random variable. Two choices which have both intuitive appeal and physical significance are the expected value and the second central moment of the random variable. We shall discuss the intuitive interpretation of these quantities here. Their physical significance will become apparent in our later work on limit theorems and statistics.

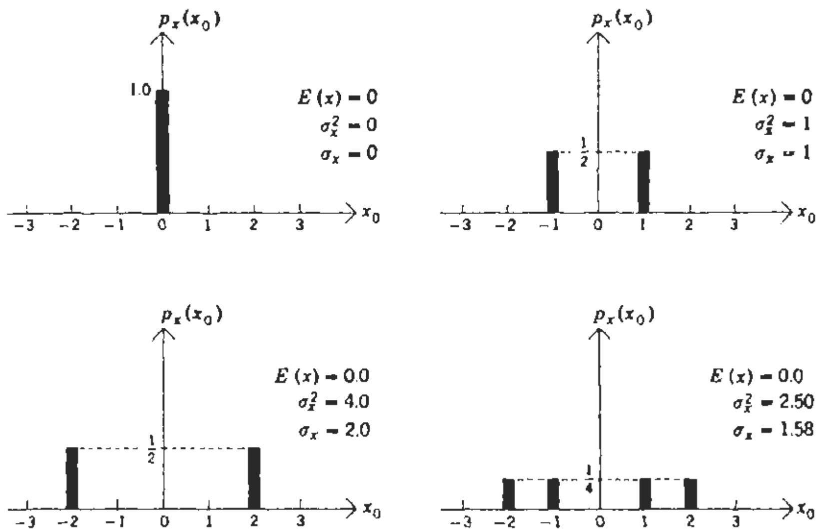
The *first moment* (or *expected value* or *mean value*) of random variable x is given by

$$E(x) = \sum_{x_0} x_0 p_x(x_0)$$

and if we picture a PMF bar graph for random variable x to be composed of broomsticks on a weightless axis, we may say that $E(x)$ specifies the location of the *center of mass* of the PMF.

The second central moment $E\{[x - E(x)]^2\}$ is a measure of the second power of the spread of the PMF for random variable x about its expected value. The second central moment of random variable x is known as its *variance* and is denoted by $\sigma_x{}^2$. The square root of the

variance, σ_x, is known as the *standard deviation* of random variable x and may be considered to be one characterization of the spread of a PMF about $E(x)$. Here are a few PMF's for random variable x, each with the same mean but different standard deviations.



A *conditional* central moment is a measure of the nth power of the spread of a conditional PMF for a random variable about its *conditional* mean. For instance, given that the experimental outcome had attribute A, the conditional variance of random variable x, $\sigma_{x|A}^2$, is given by

$$\sigma_{x|A}^2 = \sum_{x_0} [x_0 - E(x \mid A)]^2 p_{x|A}(x_0 \mid A)$$

For functions of several random variables, we again define expectation to be the weighted sum of all possible experimental values of the function, with each such value weighted by the probability of its occurrence on a performance of the experiment. Let $g(x,y)$ be a single-valued function of random variables x and y. By now we are familiar enough with the ideas involved to realize that a definition of $E[g(x,y) \mid A]$, the conditional expectation of $g(x,y)$, will include the definition for the unconditional case.

$$E[g(x,y) \mid A] = \sum_{x_0} \sum_{y_0} g(x_0,y_0) p_{x,y|A}(x_0,y_0 \mid A)$$

We should remember that in order to determine the expectation of $g(x)$ [or of $g(x,y)$ or of any function of $g(x,y)$] it is not necessary that we first determine the PMF $p_g(g_0)$. The calculation of $E[g(x,y)]$ can always be carried out directly in the x_0,y_0 event space (see Prob. 2.10).

We wish to establish some definitions and results regarding the expected values of sums and products of random variables. From the definition of the expectation of a function of several random variables, we may write

$$E(x+y) = \sum_{x_0}\sum_{y_0} (x_0+y_0)p_{x,y}(x_0,y_0)$$

$$= \underbrace{\sum_{x_0}\sum_{y_0} x_0 p_{x,y}(x_0,y_0)}_{\text{We'll sum this over } y_0 \text{ first}} + \underbrace{\sum_{x_0}\sum_{y_0} y_0 p_{x,y}(x_0,y_0)}_{\text{We'll sum this over } x_0 \text{ first}}$$

$$= \sum_{x_0} x_0 p_x(x_0) + \sum_{y_0} y_0 p_y(y_0) = E(x) + E(y) = E(x+y)$$

The expected value of the sum of two random variables is always equal to the sum of their expected values. This holds with no restrictions on the random variables, which may, of course, be functions of other random variables. The reader should, for instance, be able to use this result directly in the definition of the variance of random variable x to show

$$\sigma_x^2 = E\{[x - E(x)]^2\} = E(x^2) - [E(x)]^2$$

Now, consider the expected value of the product xy,

$$E(xy) = \sum_{x_0}\sum_{y_0} x_0 y_0 p_{x,y}(x_0,y_0)$$

In general, we can carry this operation no further without some knowledge about $p_{x,y}(x_0,y_0)$. Clearly, if x and y are independent, the above expression will factor to yield the result $E(xy) = E(x)E(y)$. Even if x and y are not independent, it is still possible that the numerical result would satisfy this condition.

If $E(xy) = E(x)E(y)$, random variables x and y are said to be *linearly independent.* (Truly independent random variables will always satisfy this condition.)

An important expectation, the *covariance* of two random variables, is introduced in Prob. 2.33. Chapter 3 will deal almost exclusively with the properties of some other useful expected values, the *transforms* of PMF's. We shall also have much more to say about expected values when we consider limit theorems and statistics in Chaps. 6 and 7.

2-8 Examples Involving Discrete Random Variables

We have dealt with only one example related to our study of discrete random variables. We now work out some more detailed examples.

example 1 A biased four-sided die is rolled once. Random variable N is defined to be the down-face value and is described by the PMF,

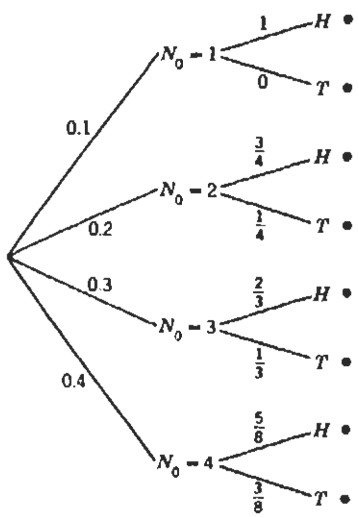
$$p_N(N_0) = \begin{cases} \dfrac{N_0}{10} & \text{for } N_0 = 1, 2, 3, 4 \\ 0 & \text{for all other values of } N_0 \end{cases}$$

Based on the outcome of this roll, a coin is supplied for which, on any flip, the probability of an outcome of heads is $(N+1)/2N$. The coin is flipped once, and the outcome of this flip completes the experiment.

Determine:

(a) The expected value and variance of discrete random variable N.
(b) The conditional PMF, conditional expected value, and conditional variance for random variable N given that the coin came up heads.
(c) If we define the events
Event A: Value of down face on die roll is either 1 or 4
Event H: Outcome of coin flip is heads
are the events A and H independent?

We'll begin by drawing a sequential sample space for the experiment, labeling all branches with the appropriate branch-traversal probabilities, and collecting some relevant information.



$P(\cdot)$	H	A	AH
$\frac{2}{20}$	✓	✓	✓
$\frac{0}{20}$		✓	
$\frac{3}{20}$	✓		
$\frac{1}{20}$			
$\frac{4}{20}$	✓		
$\frac{2}{20}$			
$\frac{5}{20}$	✓	✓	✓
$\frac{3}{20}$		✓	
$\sum P(\bullet) = 1$	$P(H) = \frac{14}{20}$	$P(A) = \frac{10}{20}$	$P(AH) = \frac{7}{20}$

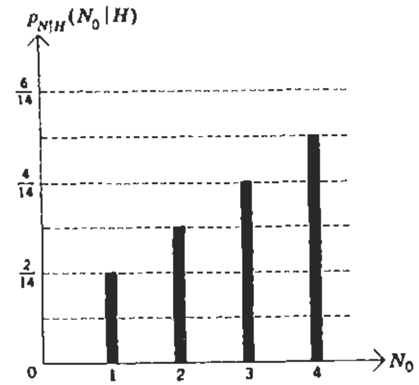
a Applying the definitions of mean and variance to random variable N, described by the PMF $p_N(N_0)$ given above, we have

$$E(N) = \sum_{N_0} N_0 p_N(N_0) = 1 \cdot \tfrac{1}{10} + 2 \cdot \tfrac{2}{10} + 3 \cdot \tfrac{3}{10} + 4 \cdot \tfrac{4}{10} = 3.0$$

$$\sigma_N{}^2 = \sum_{N_0} [N_0 - E(N)]^2 p_N(N_0)$$
$$= (-2)^2 \cdot \tfrac{1}{10} + (-1)^2 \cdot \tfrac{2}{10} + 0^2 \cdot \tfrac{3}{10} + 1^2 \cdot \tfrac{4}{10} = 1.0$$

b Given that event H did occur on this performance of the experiment, we may condition the above space by event H by removing all points with attribute T and scaling up the probabilities of all remaining events by multiplying by $1/P(H) = 10/7$. This results in the four-point conditional event space shown below. To the right of this conditional event space we present the resulting conditional PMF for random variable N, given that event H did occur.

	$P(\cdot \mid H)$
• $H, N_0 = 1$	$\frac{2}{14}$
• $H, N_0 = 2$	$\frac{3}{14}$
• $H, N_0 = 3$	$\frac{4}{14}$
• $H, N_0 = 4$	$\frac{5}{14}$



Applying the definitions of the mean and variance in this conditional event space, we have

$$E(N \mid H) = \sum_{N_0} N_0 p_{N|H}(N_0 \mid H) = 1 \cdot \tfrac{2}{14} + 2 \cdot \tfrac{3}{14} + 3 \cdot \tfrac{4}{14} + 4 \cdot \tfrac{5}{14} = \tfrac{20}{7}$$

$$\sigma_{N|A}{}^2 = \sum_{N_0} [N_0 - E(N \mid H)]^2 p_{N|H}(N_0 \mid H)$$
$$= (-\tfrac{13}{7})^2 \cdot \tfrac{2}{14} + (-\tfrac{6}{7})^2 \cdot \tfrac{3}{14} + (\tfrac{1}{7})^2 \cdot \tfrac{4}{14} + (\tfrac{8}{7})^2 \cdot \tfrac{5}{14} = \tfrac{55}{49}$$

c We wish to test $P(AH) \stackrel{?}{=} P(A)P(H)$, and we have already collected each of these three quantities in the sample space for the experiment.

$$P(A) = \tfrac{10}{20} \qquad P(H) = \tfrac{14}{20} \qquad P(AH) = \tfrac{7}{20}$$

So the events A and H are independent.

example 2 Patrolman G. R. Aft of the local constabulary starts each day by deciding how many parking tickets he will award that day. For any day, the probability that he decides to give exactly K tickets is given by the PMF

$$p_K(K_0) = \begin{cases} \dfrac{5 - K_0}{10} & \text{for } K_0 = 1, 2, 3, \text{ or } 4 \\ 0 & \text{for all other values of } K_0 \end{cases}$$

But the more tickets he gives, the less time he has to assist old ladies at street crossings. Given that he has decided to have a K-ticket day, the conditional probability that he will also help exactly L old ladies cross the street that day is given by

$$p_{L|K}(L_0 \mid K_0) = \begin{cases} \dfrac{1}{5 - K_0} & \text{if } 1 \le L_0 \le 5 - K_0 \\ 0 & \text{if } L_0 < 1 \text{ or if } L_0 > 5 - K_0 \end{cases}$$

His daily salary S is computed according to the formula

$$S = 2K + L \qquad \text{(dollars)}$$

Before we answer some questions about Officer Aft, we should be sure that we understand the above statements. For instance, on a day when Officer Aft has decided to give two tickets, the conditional PMF states that he is equally likely to help one, two, or three old ladies. Similarly, on a day when he has decided to give exactly four tickets, it is *certain* that he will help exactly one old lady cross the street.

(a) Determine the marginal PMF $p_L(L_0)$. This marginal PMF tells us the probability that Officer Aft will assist exactly L_0 old ladies on any day. Determine also the expected value of random variable L.

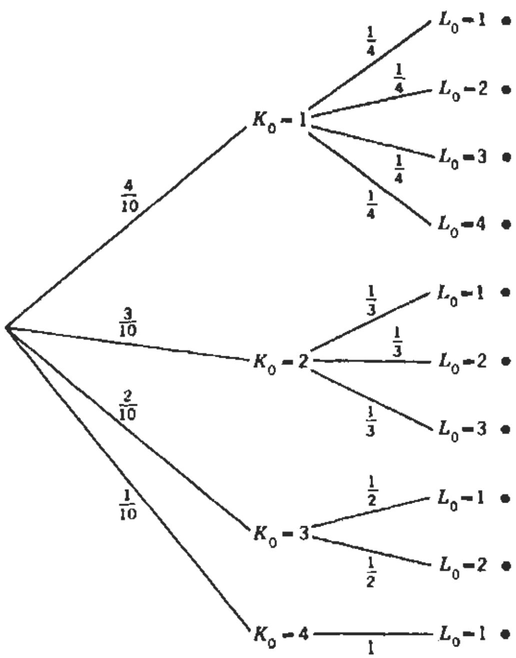
(b) Random variable S, Officer Aft's salary on any given day, is a function of random variables K and L. Determine the expected value of the quantity $S(L,K)$.

(c) Given that, on the day of interest, Officer Aft earned at least \$6, determine the conditional marginal PMF for random variable K, the number of traffic tickets he awarded on that particular day.

(d) We define

Event A: Yesterday he gave a total of one or two parking tickets.
Event B: Yesterday he assisted a total of one or two old ladies.
Determine whether or not random variables K and L are conditionally independent given event AB.

From the statement of the example we can obtain a sample space and the assignment of a priori probability measure for the experiment. We could begin with a sequential picture of the experiment such as

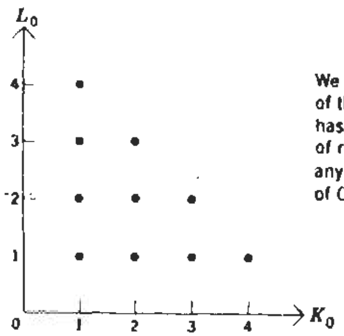


or we might work directly in a K_0,L_0 coordinate event space with the probability assignment $p_{K,L}(K_0,L_0)$ determined by

$$p_{K,L}(K_0,L_0) = p_K(K_0)p_{L|K}(L_0 \mid K_0)$$

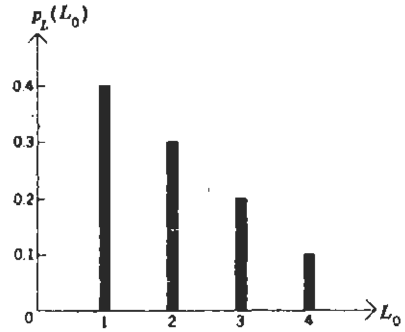
$$= \begin{cases} \dfrac{5-K_0}{10}\cdot\dfrac{1}{5-K_0} & \text{if } K_0 = 1, 2, 3, 4 \text{ and } 1 \le L_0 \le 5 - K_0 \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 0.1 & \text{if } K_0 = 1, 2, 3, 4 \text{ and } 1 \le L_0 \le 5 - K_0 \\ 0.0 & \text{otherwise} \end{cases}$$



We have established that each of the ten possible event points has an a priori probability of 0.1 of representing the outcome of any particular day in the life of Officer G. R. Aft

a The calculation $p_L(L_0) = \sum_{K_0} p_{L,K}(L_0,K_0)$ is easily performed in our event space to obtain

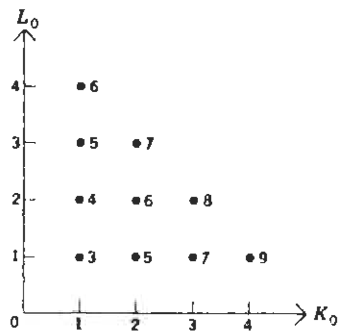


To find the expectation of L, we could multiply the experimental value of L corresponding to each sample point by the probability measure assigned to that point and sum these products. But since we already have $p_L(L_0)$, it is quicker to work in the event space for random variable L to obtain

$$E(L) = \sum_{L_0} L_0 p_L(L_0) = 1\cdot\tfrac{4}{10} + 2\cdot\tfrac{3}{10} + 3\cdot\tfrac{2}{10} + 4\cdot\tfrac{1}{10} = 2.0$$

Although it has happened in this example, we should note that there is no reason whatever why the expected value of a random variable need be equal to a possible experimental value of that random variable.

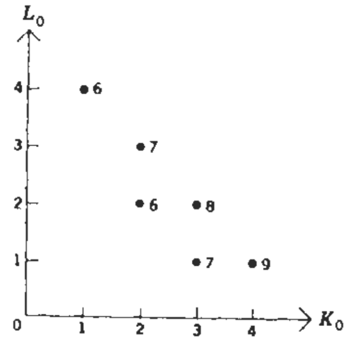
b $E(S) = \sum_{K_0}\sum_{L_0} (2K_0 + L_0)p_{K,L}(K_0,L_0)$. We can simply multiply the experimental value of S corresponding to each event point by the probability assignment of that event point. Let's write the corresponding experimental value of S beside each event point and then compute $E(S)$.



$E(S) = \frac{1}{10}(3 + 4 + 5 + 6 + 5 + 6 + 7 + 7 + 8 + 9) = \$6$

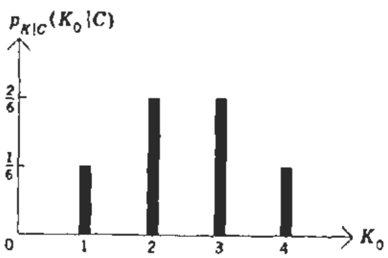
c Given that Officer Aft earned at least \$6, we can easily condition the above event space to get the conditional event space (still with the experimental values of S written beside the event points representing the remaining possible experimental outcomes)

Conditional K_0L_0 sample space given experimental outcome had attribute $S \geq 6.0$



Since the relative likelihood of all event points included in the conditioning event can't change, these remaining six event points are still equally probable

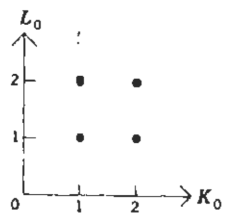
Thus, by using the notation Event C: $S \geq 6$, we have



Recall that we can use this simple interpretation of conditional probability only if the event space to which we wish to apply it is of fine enough grain to allow us to classify each event point as being wholly in C or being wholly in C'.

d There are only four (equally likely) event points in the conditional K_0,L_0 event space given that the experimental outcome has attribute AB.

Conditional K_0L_0 event space given experimental outcome has attribute AB



We wish to check

$$p_{K,L|AB}(K_0,L_0 \mid AB) \stackrel{?}{=} p_{K|AB}(K_0 \mid AB)p_{L|AB}(L_0 \mid AB) \qquad \text{for all } K_0,L_0$$

Each of these three PMF's is found from the conditional K_0,L_0 event space presented above.

$$p_{K,L|AB}(K_0,L_0 \mid AB) = \begin{cases} \frac{1}{4} & \text{if } K_0 = 1, 2 \text{ and } L_0 = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

$$p_{K|AB}(K_0 \mid AB) = \begin{cases} \frac{1}{2} & \text{if } K_0 = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

$$p_{L|AB}(L_0 \mid AB) = \begin{cases} \frac{1}{2} & \text{if } L_0 = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

The definition of conditional independence is found to be satisfied, and we conclude that random variables K and L, which were not independent in the original sample space, are conditionally independent given that the experimental outcome has attribute AB. Thus, for instance, given that AB has occurred, the conditional marginal PDF of variable L will be unaffected by the experimental value of random variable K.

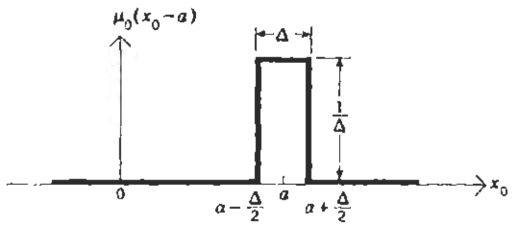
In this text, the single word *independence* applied to random variables is always used to denote *statistical* independence.

2-9 A Brief Introduction to the Unit-impulse Function

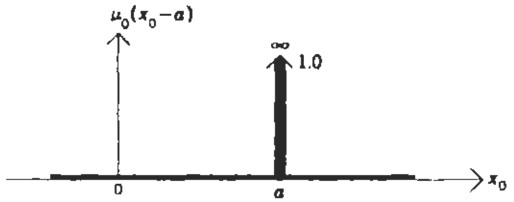
To prepare for a general discussion of continuous random variables, it is desirable that we become familiar with some properties of the unit-impulse function. Our introduction to this function, though adequate for our purposes, will lack certain details required to deal with more advanced matters.

The unit-impulse function $\mu_0(x_0 - a)$ is a function of x_0 which is equal to infinity at $x_0 = a$ and which is equal to zero for all other values of x_0. The integral of $\mu_0(x_0 - a)$ over any interval which includes the point where the unit impulse is nonzero is equal to unity.

One way to obtain the unit impulse $\mu_0(x_0 - a)$ is to consider the *limit*, as Δ goes to zero, of



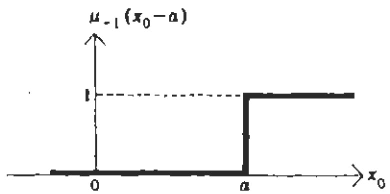
which, after the limit is taken, is normally represented as



where the scale height of the impulse is irrelevant and the total area under the impulse function is written beside the arrowhead. The integral of the unit-impulse function,

$$\int_{x_0=-\infty}^{x_0} \mu_0(x_0-a)\,dx_0$$

is a function known as the unit-step function and written as $\mu_{-1}(x_0-a)$.



Thus an impulse may be used to represent the derivative of a function at a point where the function has a vertical discontinuity.

As long as a function $g(x_0)$ does not have a discontinuity at $x_0 = a$, the integral

$$\int g(x_0)\mu_0(x_0-a)\,dx_0$$

over any interval which includes $x_0 = a$ is simply $g(a)$. This results from the fact that, over the very small range where the impulse is nonzero, $g(x_0)$ may be considered to be constant, equal to $g(a)$, and factored out of the integral.

2-10 The Probability Density Function for a Continuous Random Variable

We wish to extend our previous work to include the analysis of situations involving random variables whose experimental values may fall anywhere within continuous ranges. Some examples of such continuous random variables have already occurred in Secs. 1-2 and 2-1.

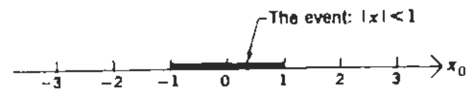
The assignment of probability measure to continuous sample and event spaces will be given by a probability *density* function (PDF). Let us begin by considering a single continuous random variable x whose

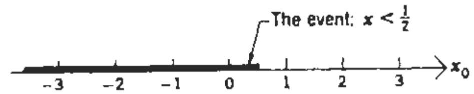
event space is the real line from $x_0 = -\infty$ to $x_0 = \infty$. We define the PDF for random variable x, $f_x(x_0)$, by

$$\text{Prob}(a < x \le b) = \int_a^b f_x(x_0)\,dx_0$$

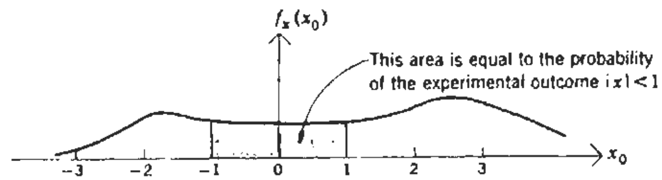
Thus, $f_x(x_0)$ is a density of probability measure on the event space (the x_0 axis) for random variable x.

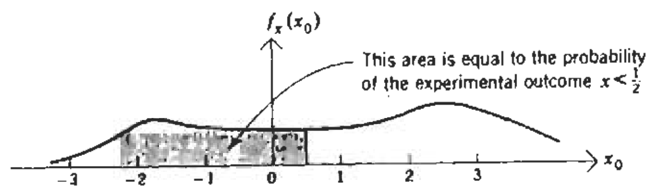
Any event can be collected by selecting those parts of the x_0 axis which have the attribute of the event. For instance,





The probability of any event is found by evaluating the integral of $f_x(x_0)$ over those parts of the event space included in the event.





Should the PDF $f_x(x_0)$ contain impulses at either a or b, the integral $\int_a^b f_x(x_0)\,dx_0$ is defined to include the area of any impulse at the upper limit but *not* the area of any impulse at the lower limit. Note that this convention is determined by our choice of the inequality signs in the definition of $f_x(x_0)$.

Based on our understanding of event space and probability measure, we note that any PDF must have the following properties:

$$\int_{-\infty}^{\infty} f_x(x_0)\,dx_0 = 1 \qquad 0 \le f_x(x_0) \le \infty$$

If we wish to reason in terms of the probability of events (of nonzero probability), it is important to realize that it is *not* the PDF itself, but rather its integral over regions of the event space, which has this interpretation. As a matter of notation, we shall always use $f_x(x_0)$ for PDF's and reserve the letter p for denoting the probability of events. This is consistent with our use of $p_x(x_0)$ for a PMF.

Note that, unless the PDF happens to have an impulse at an experimental value of a random variable, the probability assignment to any single exact experimental value of a continuous random variable is zero. [The integral of a finite $f_x(x_0)$ over an interval of zero width is equal to zero.] This doesn't mean that every particular precise experimental value is an impossible outcome, but rather that such an event of probability zero is one of an infinite number of possible outcomes.

We next define the cumulative distribution function (CDF) for random variable x, $p_{x\le}(x_0)$, by

$$p_{x\le}(x_0) = \text{Prob}(x \le x_0) = \int_{-\infty}^{x_0} f_x(x_0)\,dx_0$$

The function $p_{x\le}(x_0)$ denotes the probability that, on any particular performance of the experiment, the resulting experimental value of random variable x will be less than or equal to x_0. Note the following properties of the CDF:

$$p_{x\le}(\infty) = 1 \qquad p_{x\le}(-\infty) = 0$$

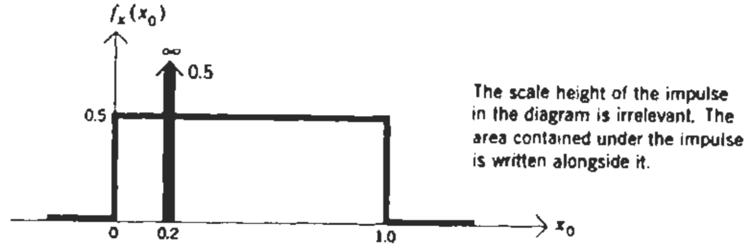
$$p_{x\le}(b) \ge p_{x\le}(a) \qquad \text{for } b \ge a$$

$$\text{Prob}(a < x \le b) = p_{x\le}(b) - p_{x\le}(a) \qquad \text{for } b \ge a$$

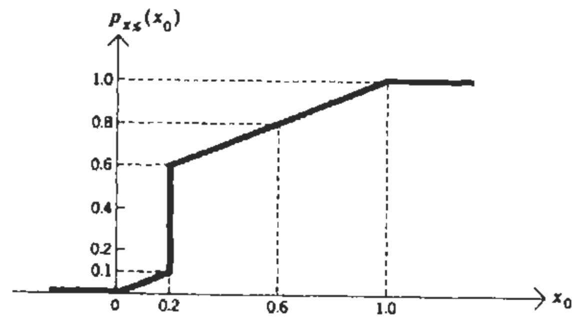
$$\frac{d}{dx_0}\,p_{x\le}(x_0) = f_x(x_0)$$

The CDF will be especially useful for some of our work with continuous random variables. For a discrete random variable, the CDF is discontinuous and therefore, to some tastes, not differentiable. For our purposes, we have defined the derivative of the CDF at such a discontinuity to be an impulse of infinite height, zero width, and area equal to the discontinuity.

Let's consider an experiment involving a random variable x defined in the following manner: A fair coin is flipped once. If the outcome is heads, the experimental value of x is to be 0.2. If the outcome is tails, the experimental value of x is obtained by one spin of a fair, infinitely finely calibrated wheel of fortune whose range is from zero to unity. This gives rise to the PDF

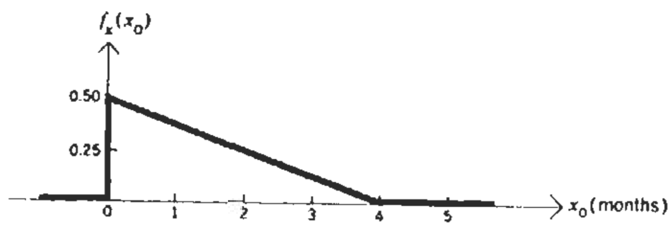


and to the CDF

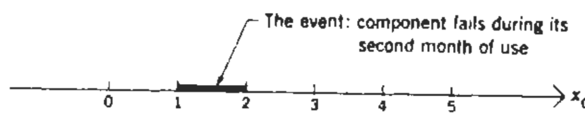


Because of the integration convention discussed earlier in this section, we can note that $p_{x\le}(x_0)$ has (in principle) its discontinuity immediately to the left of $x_0 = 0.2$.

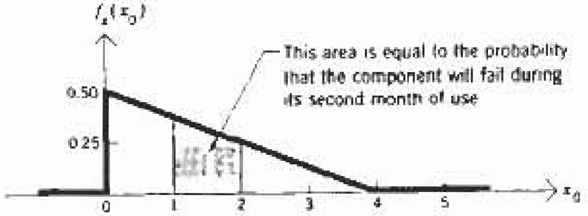
We consider one simple example which deals with a continuous random variable. Assume that the *lifetime* of a particular component is known to be a random variable described by the PDF



Let's begin by determining the a priori probability that the component fails during its second month of use. In an x_0 event space we can collect this event,



Thus, we require the quantity $\text{Prob}(1 < x \leq 2) = \int_1^2 f_x(x_0)\, dx_0 = \frac{5}{16}$, which is equal to the shaded area in the following sketch:



Since random variable x does not have a nonzero probability of taking on an experimental value of precisely 1.0 or 2.0, it makes no difference for this example whether we write $\text{Prob}(1 < x < 2)$ or $\text{Prob}(1 \leq x < 2)$ or $\text{Prob}(1 < x \leq 2)$ or $\text{Prob}(1 \leq x \leq 2)$.

Next, we ask for the conditional probability that the component will fail during its second month of use, given that it did not fail during the first month. We'll do this two ways. One approach is to define the events,

and then use the definition of conditional probability to determine the desired quantity, $P(B \mid A')$. Since it happens here that event B is included in event A', there follows

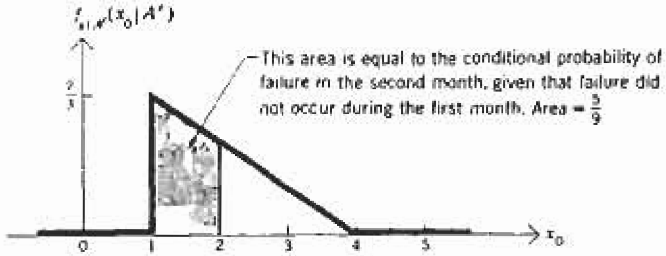
$$P(B \mid A') = \frac{P(A'B)}{P(A')} = \frac{P(B)}{P(A')}$$

$P(B) = \frac{5}{16}$ (previous result) $\qquad P(A') = \int_1^\infty f_x(x_0)\, dx_0 = \frac{9}{16}$

$$P(B \mid A') = \frac{5}{9}$$

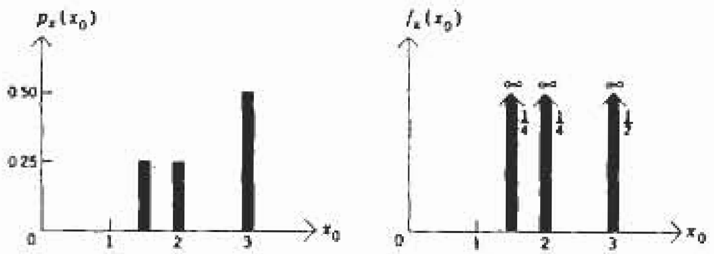
As we would expect from the nature of the physical situation in this problem, our result has the property $P(B|A') > P(B)$.

One other approach to this question would be to condition the event space for random variable x by noting that experimental values of x between 0.0 and 1.0 are now impossible. The relative probabilities of all events wholly contained within the conditioning event A' are to remain the same in our conditional space as they were in the original event space. To do this, the a priori PDF for the remaining event points must be multiplied by a constant so that the resulting conditional PDF for x will integrate to unity. This leads to the following conditional PDF $f_{x|A'}(x_0 \mid A')$:



In this latter solution, we simply worked directly with the notion which was formalized in Sec. 1-4 to obtain the definition of conditional probability.

Before closing this section, let us observe that, once we are familiar with the unit-impulse function, a PMF may be considered to represent a special type of PDF. For instance, the following probability mass function and probability density function give identical descriptions of some random variable x.

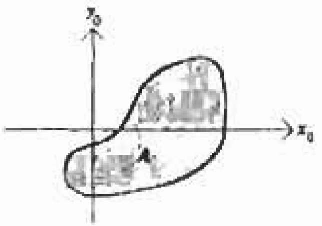


Only when there are particular advantages in doing so shall we represent the probability assignment to a purely discrete random variable by a PDF instead of a PMF.

2-11 Compound Probability Density Functions

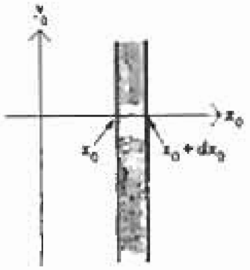
We may now consider the case where several continuous random variables are defined on the sample space of an experiment. The assignment of probability measure is then specified by a compound probability density function in an event space whose coordinates are the experimental values of the random variables.

In the two-dimensional event space for the possible experimental values of random variables x and y, the compound PDF $f_{x,y}(x_0,y_0)$ may be pictured as a surface plotted above the x_0,y_0 plane. The volume enclosed between any area in the x_0,y_0 plane and this $f_{x,y}(x_0,y_0)$ surface is equal to the probability of the experimental outcome falling within that area. For any event A defined in the x_0,y_0 plane we have



$$P(A) = \iint_A f_{x,y}(x_0,y_0)\,dx_0\,dy_0$$

The probability that the experimental value of x will fall between x_0 and $x_0 + dx_0$, which we know must be equal to $f_x(x_0)\,dx_0$, is obtained by integrating the compound PDF over the strip in the event space which represents this event.



$$\text{Prob}(x_0 < x \le x_0 + dx_0) = \int_{y_0=-\infty}^{\infty} f_{x,y}(x_0,y_0)\,dx_0\,dy_0$$

$$= dx_0 \int_{y_0=-\infty}^{\infty} f_{x,y}(x_0,y_0)\,dy_0 = f_x(x_0)\,dx_0$$

For the continuous case, we obtain the marginal PDF's by integrating over other random variables, just as we performed the same operation by summing over the other random variables in the discrete case.

$$f_x(x_0) = \int_{y_0=-\infty}^{\infty} f_{x,y}(x_0,y_0)\,dy_0 \qquad f_y(y_0) = \int_{x_0=-\infty}^{\infty} f_{x,y}(x_0,y_0)\,dx_0$$

And, in accordance with the properties of probability measure,

$$0 \le f_{x,y}(x_0,y_0) \le \infty \qquad \int_{y_0=-\infty}^{\infty}\int_{x_0=-\infty}^{\infty} f_{x,y}(x_0,y_0)\,dx_0\,dy_0 = 1$$

For convenience, we often use relations such as

$$\text{Prob}(x_0 < x \le x_0 + dx_0) = f_x(x_0)\,dx_0$$

$$\text{Prob}(x_0 < x \le x_0 + dx_0,\ y_0 < y \le y_0 + dy_0) = f_{x,y}(x_0,y_0)\,dx_0\,dy_0$$

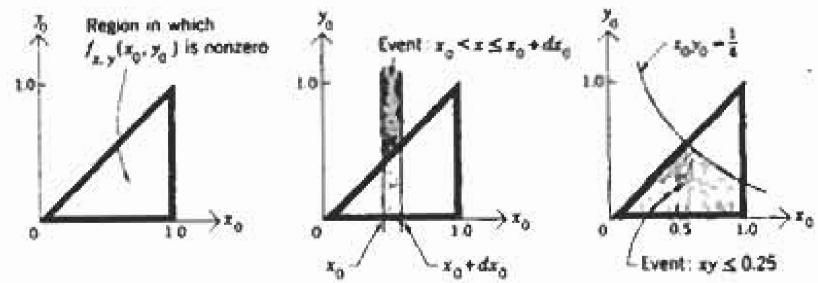
It should be remembered that such relations are not valid at points where a PDF contains impulses. *We shall not add this qualification each time we employ such incremental statements.* In any physical problem, as long as we are aware of the presence of impulses in the PDF (nonzero probability *mass* at a point), this situation will cause us no particular difficulty.

We close this section with a simple example. Suppose that the

compound PDF for random variables x and y is specified to be

$$f_{x,y}(x_0,y_0) = \begin{cases} Ax_0 & \text{if } 0 \le y_0 \le x_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and we wish to determine A, $f_x(x_0)$, and the probability that the product of the experimental values of x and y obtained on a performance of the experiment is less than or equal to 0.25. Three relevant sketches in the x_0, y_0 event space are



The value of A will be obtained from

$$\int_{x_0=-\infty}^{\infty} dx_0 \int_{y_0=-\infty}^{\infty} dy_0\, f_{x,y}(x_0,y_0) = 1$$

where our notation will be that the rightmost integration will always be performed first. Thus we have

$$\int_0^1 dx_0 \int_{y_0=0}^{x_0} dy_0\, Ax_0 = \int_0^1 Ax_0^2\,dx_0 = \frac{A}{3} = 1 \qquad A = 3$$

To determine the marginal PDF $f_x(x_0)$, we use

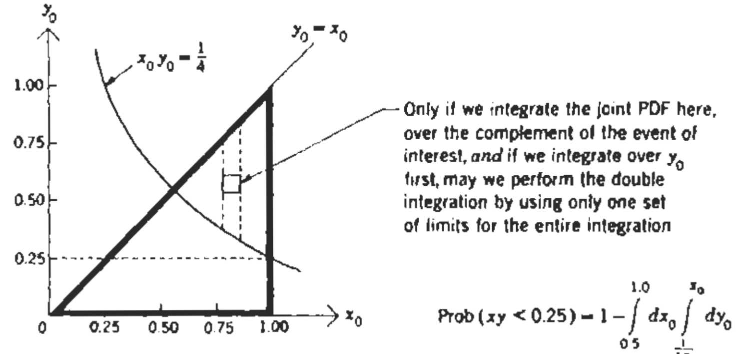
$$f_x(x_0) = \int_{y_0=-\infty}^{\infty} f_{x,y}(x_0,y_0)\,dy_0 = \begin{cases} \int_{y_0=0}^{x_0} 3x_0\,dy_0 & \text{if } 0 \le x_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Note that we must be careful to substitute the correct expression for $f_{x,y}(x_0,y_0)$ everywhere in the x_0,y_0 event space. The result simplifies to

$$f_x(x_0) = \begin{cases} 3x_0^2 & \text{if } 0 \le x_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

To determine the probability of the event $xy \le 0.25$, there are several ways to proceed. We may integrate $f_{x,y}(x_0,y_0)$ over the area representing this event in the x_0,y_0 event space. We may integrate the joint PDF over the complement of this event and subtract our result from unity. In each of these approaches we may integrate over x_0 first or over y_0 first. Note that each of these four possibilities is equivalent *but* that one of them involves far less work than the other three.

We now display this result and complete the problem by considering a more detailed sketch of the x_0,y_0 event space.

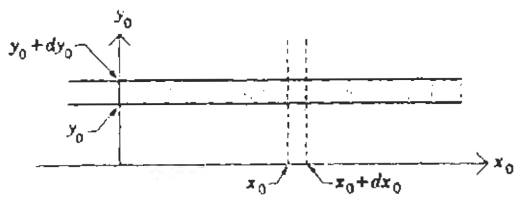


$$\text{Prob}(xy < 0.25) = 1 - \int_{0.5}^{1.0} dx_0 \int_{\frac{1}{4x_0}}^{x_0} dy_0\, 3x_0 = 0.5$$

This last discussion was a matter of elementary calculus, not probability theory. However, it is important to realize how a little forethought in planning these multiple integrations can reduce the amount of calculation and improve the probability of getting a correct result.

2-12 Conditional Probability Density Functions

Consider continuous random variables x and y, defined on the sample space of an experiment. If we are given, for a particular performance of the experiment, that the experimental value of y is between y_0 and $y_0 + dy_0$, we know that the event point representing the experimental outcome must lie within the shaded strip,



We wish to evaluate the quantity $f_{x|y}(x_0 \mid y_0)\, dx_0$, defined to be the conditional probability that the experimental value of x is between x_0 and $x_0 + dx_0$, given that the experimental value of y is between y_0 and $y_0 + dy_0$. Our procedure will be to define the incremental events of interest and substitute their probabilities into the definition of conditional probability introduced in Sec. 1-4.

Event A: $x_0 < x \le x_0 + dx_0$ Event B: $y_0 < y \le y_0 + dy_0$

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{f_{x,y}(x_0,y_0)\, dx_0\, dy_0}{f_y(y_0)\, dy_0} = \frac{f_{x,y}(x_0,y_0)}{f_y(y_0)}\, dx_0$$

Since the quantity $f_{x|y}(x_0 \mid y_0)\, dx_0$ has been defined to equal $P(A \mid B)$, we have

$$f_{x|y}(x_0 \mid y_0) = \frac{f_{x,y}(x_0,y_0)}{f_y(y_0)} \qquad \text{and, similarly,} \qquad f_{y|x}(y_0 \mid x_0) = \frac{f_{x,y}(x_0,y_0)}{f_x(x_0)}$$

The conditional PDF's are not defined when their denominators are equal to zero.

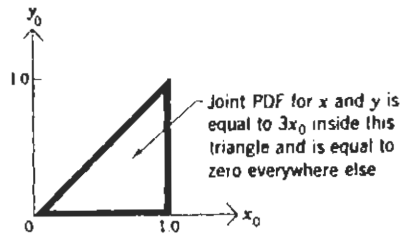
Note that the conditional PDF $f_{x|y}(x_0 \mid y_0)$, as a function of x_0 for a given y_0, is a curve of the shape obtained at the intersection of the surface $f_{x,y}(x_0,y_0)$ and a plane in three-dimensional space representing a constant value of coordinate y_0.

Using the event-space interpretation of conditional probability, we readily recognize how to condition a compound PDF given any event A of nonzero probability,

$$f_{x,y|A}(x_0,y_0 \mid A) = \begin{cases} \dfrac{f_{x,y}(x_0,y_0)}{P(A)} & \text{if } x_0,y_0 \text{ in } A \\ 0 & \text{if } x_0,y_0 \text{ in } A' \end{cases}$$

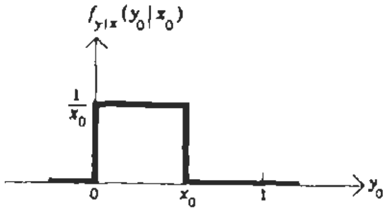
As an example of some of these concepts, let's continue with the example of the previous section, where the PDF for continuous random variables x and y is specified by

$$f_{x,y}(x_0,y_0) = \begin{cases} 3x_0 & \text{if } 0 \le y_0 \le x_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$



We'll calculate $f_{y|x}(y_0 \mid x_0)$, taking advantage of the fact that we have already determined $f_x(x_0)$.

$$f_{y|x}(y_0 \mid x_0) = \frac{f_{x,y}(x_0,y_0)}{f_x(x_0)} = \begin{cases} \dfrac{3x_0}{3x_0^2} = \dfrac{1}{x_0} & \text{if } 0 \le y_0 \le x_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$



Since the a priori PDF $f_{x,y}(x_0,y_0)$ is not a function of y_0, it is reasonable that the conditional PDF $f_{y|x}(y_0 \mid x_0)$ should be uniform over the possible experimental values of random variable y. For this example, the reader might wish to establish that

$$f_{x|y}(x_0 \mid y_0) = \begin{cases} \dfrac{2x_0}{1 - y_0^2} & \text{if } 0 \le y_0 \le x_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

2-13 Independence and Expectation for Continuous Random Variables

Two continuous random variables x and y are defined to be independent (or *statistically* independent) if and only if

$$f_{x|y}(x_0 \mid y_0) = f_x(x_0) \qquad \text{for all possible } x_0,y_0$$

and since $f_{x,y}(x_0,y_0) = f_x(x_0)f_{y|x}(y_0 \mid x_0) = f_y(y_0)f_{x|y}(x_0 \mid y_0)$ is always true by the definition of the conditional PDF's, an equivalent condition for the independence of x and y is

$$f_{x,y}(x_0,y_0) = f_x(x_0)f_y(y_0) \qquad \text{for all } x_0,y_0$$

We say that any number of random variables are mutually independent if their compound PDF factors into the product of their marginal PDF's for all possible experimental values of the random variables.

The conditional expectation of $g(x,y)$, a single-valued function of continuous random variables x and y, given that event A has occurred, is defined to be

$$E[g(x,y) \mid A] = \int_{y_0=-\infty}^{\infty} \int_{x_0=-\infty}^{\infty} g(x_0,y_0)f_{x,y|A}(x_0,y_0 \mid A)\, dx_0\, dy_0$$

All the definitions and results obtained in Sec. 2-7 carry over directly to the continuous case, with summations replaced by integrations.

2-14 Derived Probability Density Functions

We have learned that $g(x,y)$, a function of random variables x and y, is itself a new random variable. From the definition of expectation, we also know that the expected value of $g(x,y)$, or any function of $g(x,y)$,

may be found directly in the x_0,y_0 event space without ever determining the PDF $f_g(g_0)$.

However, if we have an interest only in the behavior of random variable g and we wish to answer several questions about it, we may desire to work in a g_0 event space with the PDF $f_g(g_0)$. A PDF obtained for a function of some random variables whose PDF is known is referred to as a *derived* PDF.

We shall introduce one simple method for obtaining a derived distribution by working in the event space of the random variables whose PDF is known. There may be more efficient techniques for particular classes of problems. Our method, however, will get us there and because we'll live in event space, we'll always know exactly what we are doing.

To derive the PDF for g, a function of some random variables, we need to perform only two simple steps in the event space of the original random variables:

FIRST STEP: Determine the probability of the event $g \le g_0$ for all values of g_0.

SECOND STEP: Differentiate this quantity with respect to g_0 to obtain $f_g(g_0)$.

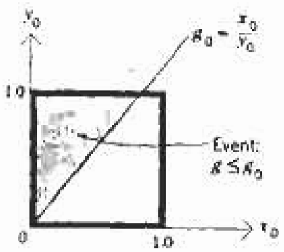
The first step requires that we calculate the cumulative probability distribution function $p_{g\le}(g_0)$. To do so, we simply integrate the given PDF for the original random variables over the appropriate region of their event space.

Consider the following example: A fair wheel of fortune, continuously calibrated from 0.00 to 1.00, is to be spun twice. The experimental values of random variables x and y are defined to be the readings on the first and second spins, respectively. [By "fair" we mean, of course, that the wheel has no memory (different spins are independent events) and that any intervals of equal arc are equally likely to include the experimental outcome.] We wish to determine the PDF $f_g(g_0)$ for the case where random variable g is defined by $g(x,y) = \dfrac{x}{y}$.

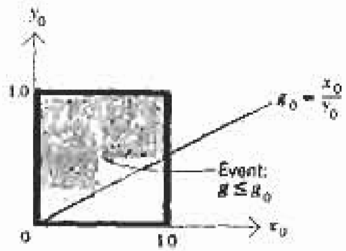
The example states that the spins are independent. Therefore we may obtain the joint PDF

$$f_{x,y}(x_0,y_0) = f_x(x_0)f_y(y_0) = \begin{cases} 1.0 & \text{if } 0 < x_0 \le 1,\ 0 < y_0 \le 1 \\ 0.0 & \text{otherwise} \end{cases}$$

Next, in the x_0,y_0 event space, we collect the event $g \le g_0$, For the case $0 \le g_0 \le 1$:



For the case $1 \le g_0 \le \infty$:



Two sketches are given to show that the boundaries of the event of interest are different for the cases $g_0 < 1$ and $g_0 > 1$. For our particular example, because of the very special fact that $f_{x,y}(x_0,y_0)$ is everywhere equal to either zero or unity, we can replace the integration

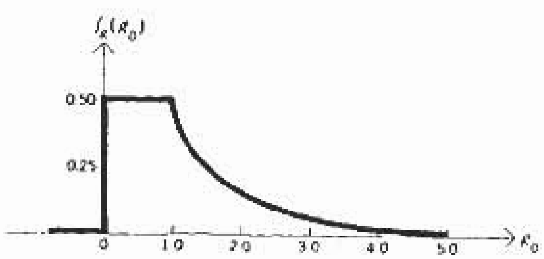
$$p_{g\le}(g_0) = \iint f_{x,y}(x_0,y_0)\,dx_0\,dy_0$$

by a simple calculation of areas to obtain, for the first step of our two-step procedure,

$$p_{g\le}(g_0) = \begin{cases} 0 & g_0 \le 0 \\ \dfrac{g_0}{2} & 0 \le g_0 \le 1 \\ 1 - \dfrac{1}{2g_0} & 1 \le g_0 \le \infty \end{cases}$$

At this point, of course, we may check that this CDF is a monotonically increasing function of x_0 and that it increases from zero at $g_0 = -\infty$ to unity at $g_0 = +\infty$. Now, for step 2, we differentiate $p_{g\le}(g_0)$ to find

$$f_g(g_0) = \begin{cases} 0 & g_0 \le 0 \\ 0.5 & 0 \le g_0 \le 1 \\ 0.5g_0^{-2} & 1 \le g_0 \le \infty \end{cases}$$



If we wish to derive a joint PDF, say, for $g(x,y)$ and $h(x,y)$, then in step 1 we collect the probability

$p_{g\le,h\le}(g_0,h_0)$ for all values of g_0 and h_0

which represents the joint CDF for random variables g and h. For step 2 we would perform the differentiation

$$\frac{\partial^2 p_{g\le,h\le}(g_0,h_0)}{\partial g_0\,\partial h_0} = f_{g,h}(g_0,h_0)$$

As the number of derived random variables increases, our method becomes unreasonably cumbersome; but more efficient techniques exist for particular types of problems.

One further detail is relevant to the mechanics of the work involved in carrying out our two-step method for derived distributions. Our method involves an integration (step 1) followed by a differentiation (step 2). Although the integrations and differentiations are generally with respect to different variables, we may wish to differentiate first before formally performing the integration. For this purpose, it is useful to remember one very useful formula and the picture from which it is obtained.

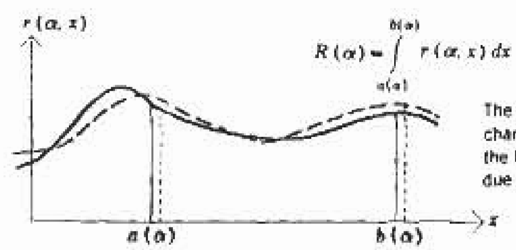
In working with a relation of the form

$$R(\alpha) = \int_{a(\alpha)}^{b(\alpha)} r(\alpha,x)\,dx$$

we have an integral over x whose lower limit, upper limit, and integrand are all functions of α. If we desire to obtain the derivative $\dfrac{dR(\alpha)}{d\alpha}$ it is more efficient to use the following formula directly than to first perform the integration with respect to x and then to differentiate with respect to α.

$$\frac{dR(\alpha)}{d\alpha} = -r[\alpha,a(\alpha)]\,\frac{da(\alpha)}{d\alpha} + r[\alpha,b(\alpha)]\,\frac{db(\alpha)}{d\alpha} + \int_{a(\alpha)}^{b(\alpha)} \frac{\partial r(\alpha,x)}{\partial\alpha}\,dx$$

This relation is easy to remember if we keep in mind the picture from which it may be obtained



The reader will have many opportunities to benefit from the above expression for $\frac{dR(\alpha)}{d\alpha}$ in obtaining derived distributions. For instance, in Prob. 2.30 at the end of this chapter, this expression is used in order to obtain the desired PDF in a useful form.

2-15 Examples Involving Continuous Random Variables

However simple the concepts may seem, no reader should assume that he has a working knowledge of the material in this chapter until he has successfully mastered many problems like those at the end of the chapter. Two examples with solutions follow, but we must realize that these examples are necessarily a narrow representation of a very large class of problems. Our first example gives a straightforward drill on notation and procedures. The second example is more physically motivated.

example 1 Random variables x, y, and z are described by the compound probability density function,

$$f_{x,y,z}(x_0,y_0,z_0) = \begin{cases} x_0z_0 + 3y_0z_0 & \text{if } 0 \le x_0 \le 1, 0 \le y_0 \le 1, 0 \le z_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Determine the quantities:

(a) $p_{x\le}(3)$ (b) $f_{x,y}(x_0,y_0)$
(c) $p_{x\le,y\le,z\le}(1,2,z_0)$ (d) $f_x(x_0)$
(e) $E(xy)$ (f) $E(y \mid x)$

a From the statement of the compound PDF, note that the experimental value of random variable x can never be greater than unity. Since we are asked to determine the probability that this experimental value is less than or equal to 3.0, we can immediately answer

$$p_{x\le}(3) = 1.0$$

b When determining expressions for PDF's and CDF's we must always remember that the proper expressions must be listed for all values of the arguments.

$$f_{x,y}(x_0,y_0) = \begin{cases} \int_{z_0=0}^{1} dz_0(x_0z_0 + 3y_0z_0) & \text{if } 0 \le x_0 \le 1, 0 \le y_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

which simplifies to

$$f_{x,y}(x_0,y_0) = \begin{cases} \frac{1}{2}(x_0 + 3y_0) & \text{if } 0 \le x_0 \le 1, 0 \le y_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

c Because of the given ranges of possible experimental values of x and y, we note that

$$p_{x\le,y\le,z\le}(1,2,z_0) = p_{z\le}(z_0)$$

$$p_{z\le}(z_0) = \begin{cases} 0 & z_0 \le 0 \\ \int_{z_0=0}^{z_0} dz_0 \int_{x_0=0}^{1} dx_0 \int_{y_0=0}^{1} dy_0(x_0z_0 + 3y_0z_0) & 0 \le z_0 \le 1 \\ 1 & 1 \le z_0 \end{cases}$$

which simplifies to

$$p_{z\le}(z_0) = \begin{cases} 0 & z_0 \le 0 \\ z_0^2 & 0 \le z_0 \le 1 \\ 1 & 1 \le z_0 \end{cases}$$

which has all the essential properties of a CDF.

d Since we have already found $f_{x,y}(x_0,y_0)$, we can determine the marginal PDF $f_x(x_0)$ by integrating over y_0. For $0 \le x_0 \le 1$, we have

$$f_x(x_0) = \int_{y_0=0}^{1} dy_0 \tfrac{1}{2}(x_0 + 3y_0) = (\tfrac{1}{2}x_0 + \tfrac{3}{4})$$

and we know that $f_x(x_0)$ is zero elsewhere.

$$f_x(x_0) = \begin{cases} \frac{1}{2}x_0 + \frac{3}{4} & \text{if } 0 \le x_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Whenever possible, we perform simple checks on our answers to see whether or not they make any sense. For instance, here we'd check to see that $\int_{x_0=-\infty}^{\infty} f_x(x_0)\, dx_0$ is unity. Happily, it is.

e $E(xy) = \int_{x_0=0}^{1} dx_0 \int_{y_0=0}^{1} dy_0\, x_0y_0f_{x,y}(x_0,y_0) = \frac{1}{3}$

This result is at least compatible with reason, since xy is always between zero and unity.

f $E(y \mid x) = \int_{y_0=0}^{1} y_0f_{y|x}(y_0 \mid x_0)\, dy_0$

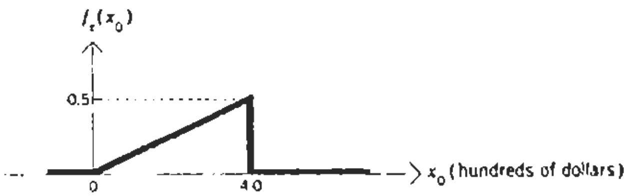
$$= \int_{y_0=0}^{1} y_0 \frac{f_{x,y}(x_0,y_0)}{f_x(x_0)}\, dy_0 \qquad \text{for all possible } x_0$$

$$= \int_{y_0=0}^{1} y_0 \frac{\frac{1}{2}(x_0 + 3y_0)}{\frac{1}{2}x_0 + \frac{3}{4}}\, dy_0 = \frac{\frac{1}{2}}{\frac{1}{2}x_0 + \frac{3}{4}} \int_{y_0=0}^{1} (x_0y_0 + 3y_0^2)\, dy_0$$

$$= \frac{x_0 + 2}{2x_0 + 3}$$

For any possible value of x_0, our $E(y \mid x)$ does result in a conditional expectation for y which is always between the smallest and largest possible experimental values of random variable y.

example 2 Each day as he leaves home for the local casino, Oscar spins a biased wheel of fortune to determine how much money to take with him. He takes exactly x hundred dollars with him, where x is a continuous random variable described by the probability density function



As a matter of convenience, we are assuming that the currency is infinitely divisible. (Rounding off to the nearest penny wouldn't matter much.)

Oscar has a lot of experience at this. All of it is bad. Decades of experience have shown that, over the course of an evening, Oscar never wins. In fact, the amount with which he returns home on any particular night is uniformly distributed between zero and the amount with which he started out.

Let random variable y represent the amount (in hundreds of dollars) Oscar brings home on any particular night.

(a) Determine $f_{x,y}(x_0,y_0)$, the joint PDF for his original wealth x and his terminal wealth y on any evening.

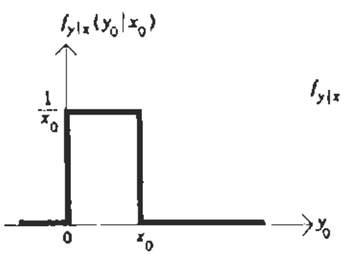
(b) Determine $f_y(y_0)$, the marginal probability density function for the amount Oscar will bring home on a randomly selected night.

(c) Determine the expected value of Oscar's loss on any particular night.

(d) On one particular night, we learn that Oscar returned home with less than \$200. For that night, determine the conditional probability of each of the following events:

(i) He started out for the casino with less than \$200.
(ii) His loss was less than \$100.
(iii) His loss that night was exactly \$75.

a From the example statement we obtain

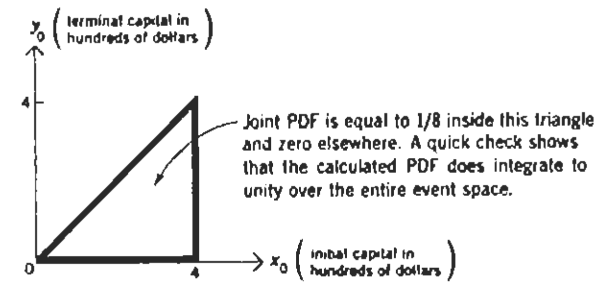


$$f_{y|x}(y_0 \mid x_0) = \begin{cases} \frac{1}{x_0} & 0 \le y_0 \le x_0 \le 4 \\ 0 & \text{otherwise} \end{cases}$$

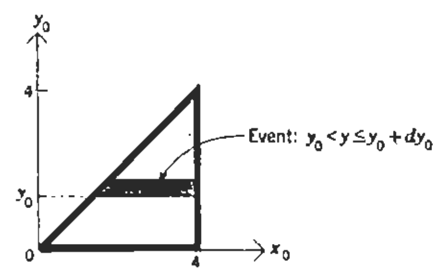
The definition of conditional probability is used with the given $f_x(x_0)$ to determine

$$f_{x,y}(x_0,y_0) = f_x(x_0) f_{y|x}(y_0 \mid x_0) = \begin{cases} \frac{x_0}{8} \cdot \frac{1}{x_0} = \frac{1}{8} & \text{if } 0 \le y_0 \le x_0 \le 4 \\ 0 & \text{otherwise} \end{cases}$$

and this result may be displayed in an x_0,y_0 event space,



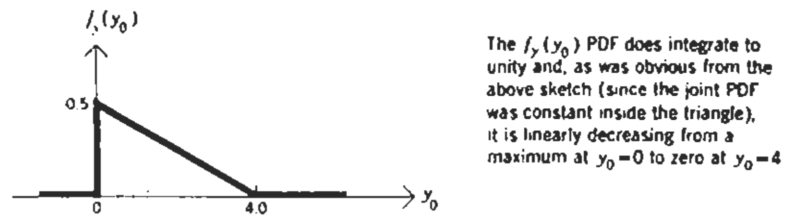
b



For $0 \le y_0 \le 4$,

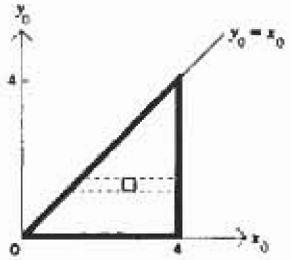
$$f_y(y_0) = \int_{-\infty}^{\infty} dx_0\, f_{x,y}(x_0,y_0) = \int_{x_0=y_0}^{4} dx_0\, \frac{1}{8} = \frac{1}{8}(4 - y_0)$$

And we can sketch this PDF as



c $E(x - y) = \int_{y_0} \int_{x_0} (x_0 - y_0) f_{x,y}(x_0,y_0)\, dx_0\, dy_0$

We must always be careful of the limits on the integrals when we substitute actual expressions for the compound PDF. We'll integrate over x_0 first.



$$E(x - y) = \int_{y_0=0}^{4} dy_0 \int_{x_0=y_0}^{4} dx_0 \, \tfrac{1}{8}(x_0 - y_0) = \$133.33$$

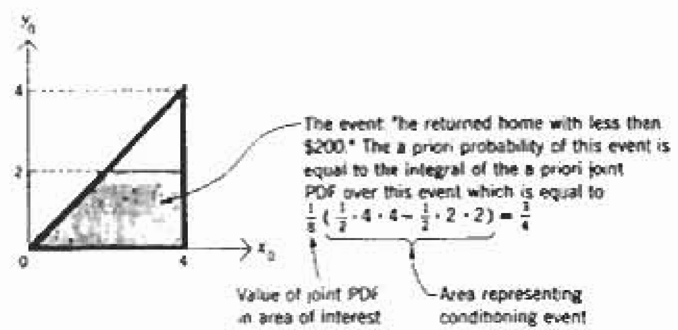
where again we are using the convention that the successive integrals are to be performed in order from right to left. By changing a sign in the proof in Sec. 2-7, we may prove the relation

$$E(x - y) = E(x) - E(y)$$

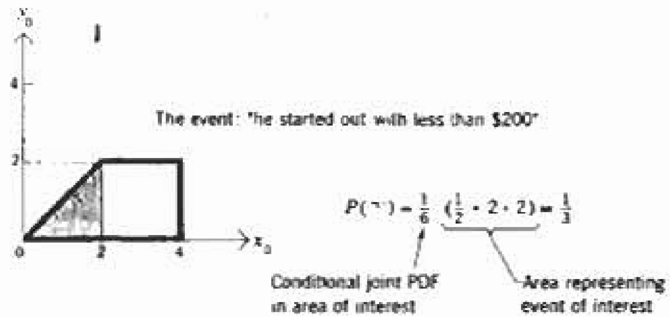
Since we already have the marginal PDF's, this relation allows us to obtain $E(x - y)$ by another route,

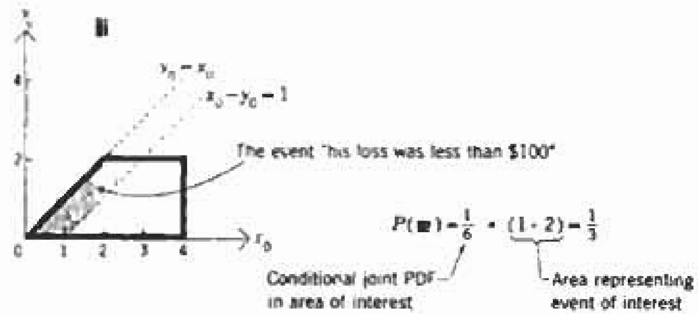
$$E(x) - E(y) = \int_{x_0=-\infty}^{\infty} x_0 f_x(x_0)\,dx_0 - \int_{y_0=-\infty}^{\infty} y_0 f_y(y_0)\,dy_0$$
$$= \int_0^4 x_0 \cdot \frac{x_0}{8}\,dx_0 - \int_0^4 y_0 \frac{4 - y_0}{8}\,dy_0 = \$133.33$$

d Given that Oscar returned home with less than \$200, we work in the appropriate conditional sample space. At all points consistent with the conditioning event, the conditional PDF is equal to the original joint PDF scaled up by the reciprocal of the a priori probability of the conditioning event.



Thus, the conditional joint PDF is equal to $\frac{1}{8}/\frac{3}{4} = \frac{1}{6}$ in the region where it is nonzero. Now we may answer all questions in this conditional space.





We realize that, in general, we would have to integrate the conditional PDF over the appropriate events to obtain their probabilities. Only because the conditional PDF is a constant have we been able to reduce the integrations to simple multiplications.

iii As long as we allow the currency to be infinitely divisible, the conditional probability measure associated with the event $x - y = 75$ is equal to zero. The integral of the compound PDF $f_{x,y}(x_0,y_0)$ over the *line* representing this event in the x_0,y_0 event space is equal to zero.

PROBLEMS

2.01 The *geometric* PMF for discrete random variable K is defined to be

$$p_K(K_0) = \begin{cases} C(1 - P)^{K_0 - 1} & \text{if } K_0 = 1, 2, 3, \ldots \text{ and } 0 < P < 1 \\ 0 & \text{for all other values of } K_0 \end{cases}$$

a Determine the value of C.

b Let N be a positive integer. Determine the probability that an experimental value of K will be greater than N.

c Given that an experimental value of random variable K is greater than integer N, what is the conditional probability that it is also larger than $2N$? (We shall discuss this special result in Chap. 4.)

d What is the probability that an experimental value of K is equal to an integer multiple of 3?

2.02 The probability that any particular bulb will burn out during its Kth month of use is given by the PMF for K,

$$p_K(K_0) = \tfrac{1}{5}(\tfrac{4}{5})^{K_0-1} \qquad K_0 = 1, 2, 3, \ldots$$

Four bulbs are life-tested simultaneously. Determine the probability that

a None of the four bulbs fails during its first month of use.
b Exactly two bulbs have failed by the end of the third month.
c Exactly one bulb fails during each of the first three months.
d Exactly one bulb has failed by the end of the second month, and exactly two bulbs are still working at the start of the fifth month.

2.03 The *Poisson* PMF for random variable K is defined to be

$$p_K(K_0) = \begin{cases} \dfrac{\mu^{K_0}e^{-\mu}}{K_0!} & \text{if } K_0 = 0, 1, 2, \ldots \text{ (and } \mu > 0) \\ 0 & \text{for all other values of } K_0 \end{cases}$$

a Show that this PMF sums to unity.
b Discrete random variables R and S are defined on the sample spaces of two different, unrelated experiments, and these random variables have the PMF's

$$p_R(R_0) = \frac{\mu^{R_0}e^{-\mu}}{R_0!} \qquad R_0 = 0, 1, 2, \ldots$$

$$p_S(S_0) = \frac{\lambda^{S_0}e^{-\lambda}}{S_0!} \qquad S_0 = 0, 1, 2, \ldots$$

Use an R_0,S_0 sample space to determine the PMF $p_T(T_0)$, where discrete random variable T is defined by $T = R + S$.
c Random variable W is defined by $W = cR$, where c is a known nonzero constant. Determine the PMF $p_W(W_0)$ and the expected value of W. How will the nth central moment of W change if c is doubled?

2.04 Discrete random variable x is described by the PMF

$$p_x(x_0) = \begin{cases} K - \dfrac{x_0}{12} & \text{if } x_0 = 0, 1, 2 \\ 0 & \text{for all other values of } x_0 \end{cases}$$

Let $d_1, d_2, \ldots, d_N$ represent N successive independent experimental values of random variable x.
a Determine the numerical value of K.

b Determine the probability that $d_1 > d_2$.
c Determine the probability that $d_1 + d_2 + \cdots + d_N \leq 1.0$
d Define $r = \max(d_1,d_2)$ and $s = \min(d_1,d_2)$. Determine the following PMF's for all values of their arguments:
i $p_s(s_0)$ **ii** $p_{r|s}(r_0 \mid 0)$
iii $p_{r,s}(r_0,s_0)$ **iv** $p_t(t_0)$, with $t = (1 + d_1)/(1 + s)$
e Determine the expected value and variance of random variable s defined above.
f Given $d_1 + d_2 \leq 2.0$, determine the conditional expected value and conditional variance of random variable s defined above.

2.05 Discrete random variable x is described by the PMF $p_x(x_0)$. Before an experiment is performed, we are required to guess a value d. After an experimental value of x is obtained, we shall then be paid an amount $A - B(x - d)^2$ dollars.
a What value of d should we use to maximize the expected value of our financial gain?
b Determine the value of A such that the expected value of the gain is zero dollars.

2.06 Consider an experiment in which a fair four-sided die (with faces labeled 0, 1, 2, 3) is thrown once to determine how many times a fair coin is to be flipped. In the sample space of this experiment, random variables n and k are defined by

n = down-face value on the throw of the tetrahedral die
k = total number of heads resulting from the coin flips

Determine and sketch each of the following functions for all values of their arguments:
a $p_n(n_0)$ **b** $p_{k|n}(k_0 \mid 2)$ **c** $p_{n|k}(n_0 \mid 2)$ **d** $p_k(k_0)$
e Also determine the conditional PMF for random variable n, given that the experimental value of k is an odd number.

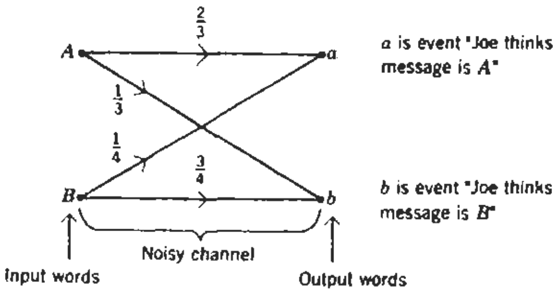
2.07 Joe and Helen each know that the a priori probability that her mother will be home on any given night is 0.6. However, Helen can determine her mother's plans for the night at 6 P.M., and then, at 6:15 P.M., she has only one chance each evening to shout one of two code words across the river to Joe. He will visit her with probability 1.0 if he thinks Helen's message means "Ma will be away," and he will stay home with probability 1.0 if he thinks the message means "Ma will be home."

But Helen has a meek voice, and the river is channeled for heavy barge traffic. Thus she is faced with the problem of coding for a noisy channel. She has decided to use a code containing only the code words A and B.

The channel is described by

$$P(a \mid A) = \tfrac{2}{3} \qquad P(a \mid B) = \tfrac{1}{4} \qquad P(b \mid A) = \tfrac{1}{3} \qquad P(b \mid B) = \tfrac{3}{4}$$

and these events are defined in the following sketch:



a In order to minimize the probability of error between transmitted and received messages, should Helen and Joe agree to use code I or code II?

Code I	Code II
A = Ma away	A = Ma home
B = Ma home	B = Ma away

b Helen and Joe put the following cash values (in dollars) on all possible outcomes of a day:

Ma home and Joe comes	−30
Ma home and Joe doesn't come	0
Ma away and Joe comes	+30
Ma away and Joe doesn't come	−5

Joe and Helen make all their plans with the objective of maximizing the expected value of each day of their continuing romance. Which of the above codes will maximize the expected cash value per day of this romance?

c Clara isn't quite so attractive as Helen, but at least she lives on the same side of the river. What would be the lower limit of Clara's expected value per day which would make Joe decide to give up Helen?

d What would be the maximum rate which Joe would pay the phone company for a noiseless wire to Helen's house which he could use once per day at 6:15 P.M.?

e How much is it worth to Joe and Helen to double her mother's probability of being away from home? Would this be a better or worse investment than spending the same amount of money for a telephone line (to be used once a day at 6:15 P.M.) with the following properties:

$$P(a \mid A) = P(b \mid B) = 0.9 \qquad P(b \mid A) = P(a \mid B) = 0.1$$

2.08 A frazzle is equally likely to contain zero, one, two, or three defects. No frazzle has more than three defects. The cash price of each frazzle is set at \$$(10 - K^2)$, where K is the number of defects in it. Gummed labels, each representing \$1, are placed on each frazzle to indicate its cash value (one label for a \$1 frazzle, two labels for a \$2 frazzle, etc.).

What is the probability that a randomly selected label (chosen from the pile of labels at the printing plant) will end up on a frazzle which has exactly two defects?

2.09 A pair of fair four-sided dice is thrown once. Each die has faces labeled 1, 2, 3, and 4. Discrete random variable x is defined to be the product of the down-face values. Determine the conditional variance of x^2 given that the sum of the down-face values is greater than the product of the down-face values.

2.10 Discrete random variables x and y are defined on the sample space of an experiment, and $g(x,y)$ is a single valued function of its argument. Use an event-space argument to establish that

$$E[g(x,y)] = \sum_{x_0} \sum_{y_0} g(x_0,y_0) p_{x,y}(x_0,y_0)$$

and

$$E(g) = \sum_{g_0} g_0 p_g(g_0)$$

are equivalent expressions for the expected value of $g(x,y)$.

2.11 At a particular point on a busy one-way single-lane road, a study is made of the distribution of the interarrival period T between successive car arrivals. A reasonable quantization of the data for a chain of 10,001 cars results in the following tabulation:

T, seconds	2	4	6	8	12
Number of occurrences	1,000	2,000	4,000	2,000	1,000

(Consider the cars to be as wide as the road, but very short.)

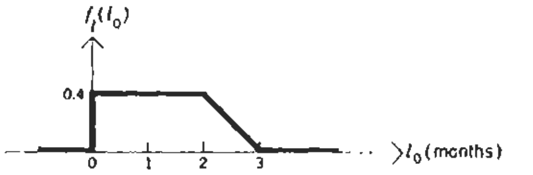
a A young wombat, who never turns back, requires five seconds to cross the street. Determine the probability that he survives if:

i He starts immediately after a car has passed.

ii He starts at a random time, selected without any dependence on the state of the traffic.

b Is the safer method of the above problem always safer, no matter what data are given?

2.12 The life span of a particular mechanical part is a random variable described by the following PDF:



If three such parts are put into service independently at $t = 0$, determine

a The probability that the first failure will not have occurred before time t_0 $(0 \leq t_0 \leq \infty)$

b $E(l)$

c A simple expression for the expected value of the time until the majority of the parts will have failed.

2.13 Continuous random variables w, x, y, and z are described by the compound PDF $f_{w,x,y,z}(w_0,x_0,y_0,z_0)$. Determine a simple expression for the probability of the event $x = \max(w,x,y,z)$.

2.14 Random variables x and y described by the PDF

$$f_{x,y}(x_0,y_0) = \begin{cases} K & \text{if } x_0 + y_0 \leq 1 \text{ and } x_0 > 0 \text{ and } y_0 > 0 \\ 0 & \text{otherwise} \end{cases}$$

a Are x and y independent random variables?

b Are they conditionally independent given $\max(x,y) \leq 0.5$?

c Determine the expected value of random variable r, defined by $r = xy$.

d If we define events A and B by

Event $A: 2(y - x) \geq y + x$ $\qquad$ Event $B: y > \frac{3}{4}$

obtain the numerical values of $P(A)$, $P(B)$, $P(A'B)$, $P[(A'B')']$, and determine and plot the conditional probability density function $f_{x|A'B'}(x_0 \mid A'B')$.

2.15 One of two wheels of fortune, A and B, is selected by the flip of a fair coin, and the wheel chosen is spun once to determine an experimental value of random variable x. Random variable y, the reading obtained with wheel A, and random variable w, the reading obtained with wheel B, are described by the PDF's

$$f_y(y_0) = \begin{cases} 1 & \text{if } 0 < y_0 \leq 1 \\ 0 & \text{otherwise} \end{cases} \qquad f_w(w_0) = \begin{cases} 3 & \text{if } 0 < w_0 \leq \frac{1}{3} \\ 0 & \text{otherwise} \end{cases}$$

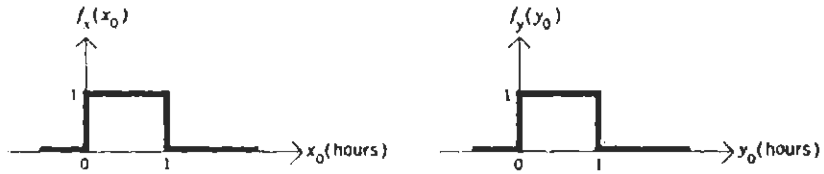
If we are told the experimental value of x was less than $\frac{1}{4}$, what is the conditional probability that wheel A was the one selected?

2.16 Four random variables are described by the probability density function

$$f_{w,x,y,z}(w_0,x_0,y_0,z_0) = \begin{cases} A\left(\dfrac{w_0x_0}{y_0z_0}\right)^2 \log \dfrac{w_0}{y_0} & \begin{aligned} &\text{if } 1 \leq w_0 \leq 2 \text{ and} \\ &1 \leq x_0 \leq 2 \text{ and} \\ &1 \leq y_0 \leq w_0 \leq 2 \text{ and} \\ &1 \leq z_0 \leq 2 \end{aligned} \\ 0 & \text{otherwise} \end{cases}$$

Determine and discuss the conditional probability density function $f_{x|w,y,z}(x_0 \mid w_0,y_0,z_0)$.

2.17 Random variables x and y are independent and are described by the probability density functions $f_x(x_0)$ and $f_y(y_0)$,



Stations A and B are connected by two *parallel* message channels. A message from A to B is sent over both channels at the same time. Random variables x and y represent the message delays over parallel channels I and II, respectively.

A message is considered "received" as soon as it arrives on any one channel, and it is considered "verified" as soon as it has arrived over both channels.

a Determine the probability that a message is received within 15 minutes after it is sent.

b Determine the probability that the message is received but not verified within 15 minutes after it is sent.

c Let μ represent the time (in hours) between transmission at A and verification at B. Determine the cumulative distribution function $p_{\mu\leq}(\mu_0)$, and then differentiate it to obtain the PDF $f_\mu(\mu_0)$.

d If the attendant at B goes home 15 minutes after the message is received, what is the probability that he is present when the message should be verified?

e If the attendant at B leaves for a 15-minute coffee break right after the message is received, what is the probability that he is present at the proper time for verification?

f The management wishes to have the maximum probability of having the attendant present for *both* reception and verification. Would they do better to let him take his coffee break as described above or simply allow him to go home 45 minutes after transmission?

2.18 The data for an experiment that could be performed only once consist of experimental values of random variables x and y, which are described by the a priori probability density function:

$$f_{x,y}(x_0,y_0) = \begin{cases} Ax_0^2y_0 & \text{if } 0 \le (x_0 + y_0) \le 1, \quad x_0 \ge 0, \quad y_0 \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

But the experimental data were lost. All the experimenter remembers is that the experimental value of x was either 0.4 or 0.6, and he decides that he observed one or the other of these values with equal probability.

Based on the above information, determine the experimenter's probability density function for his experimental value of random variable y.

2.19 The *exponential* PDF for random variable x is given by

$f_x(x_0) = \lambda e^{-\lambda x_0}$ for $x_0 \ge 0$.

a Determine the probability that an experimental value of x will be greater than $E(x)$.

b Suppose the lifetime of a bulb is given by the above PDF and we are told that the bulb has already been on for T units of time; determine the PDF for the remaining lifetime $y = x - T$ of the bulb. (This special result will be discussed in Chap. 4.)

c Assume that each bulb is replaced as soon as it burns out. Over a very long period, determine the fraction of this interval for which illumination is supplied by those bulbs whose lifetimes are longer than $E(x)$.

2.20 **a** For three spins of a fair wheel of fortune, what is the probability that none of the resulting experimental values is within $\pm 30°$ of any other experimental value?

b What is the smallest number of spins for which the probability that at least one other reading is within $\pm 30°$ of the first reading is at least 0.9?

2.21 Random variable x is described by a PDF which is constant between $x_0 = 0$ and $x_0 = 1$ and which is zero elsewhere. K independent successive experimental values of this random variable are labeled $d_1, d_2, \ldots, d_K$. Define the random variables

r = second largest of $d_1, d_2, \ldots, d_K$

s = second smallest of $d_1, d_2, \ldots, d_K$

and determine the joint probability density function $f_{r,s}(r_0,s_0)$ for all values of r_0 and s_0.

2.22 The probability density function for continuous random variable x is a constant in the range $a < x \le b$ and zero elsewhere.

a Determine σ_x, the standard deviation of random variable x.

b Determine the conditional standard deviation of x, given that $|x - E(x)| > \sigma_x$.

c If $y = cx + d$, determine $E(y)$ and σ_y in terms of $E(x)$ and σ_x. Do your results depend on the form of the PDF for random variable x?

2.23 Random variable x is described by the PDF

$$f_x(x_0) = \begin{cases} 0.1 & \text{if } 0 \le x_0 \le 10.0 \\ 0 & \text{otherwise} \end{cases}$$

Another random variable, y, is defined by $y = -\ln x$. Determine the PDF $f_y(y_0)$.

2.24 Random variables x and y are distributed according to the joint probability density function

$$f_{x,y}(x_0,y_0) = \begin{cases} Ax_0 & \text{if } 1 \le x_0 \le y_0 \le 2 \\ 0.0 & \text{otherwise} \end{cases}$$

a Evaluate the constant A.

b Determine the marginal probability density function $f_y(y_0)$.

c Determine the expected value of $1/x$, given that $y = \frac{3}{2}$.

d Random variable z is defined by $z = y - x$. Determine the probability density function $f_z(z_0)$.

2.25 Random variables x and y are described by the joint density function

$$f_{x,y}(x_0,y_0) = \begin{cases} K & \text{if } 0 \le y_0 \le x_0 \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Random variable z is defined by

$z = \max(x,2y)$

Determine and sketch $f_z(z_0)$, the probability density function for random variable z.

2.26 Melvin Fooch, a student of probability theory, has found that the hours he spends working (w) and sleeping (s) in preparation for a final exam are random variables described by

$$f_{w,s}(w_0,s_0) = \begin{cases} K & \text{if } 10 \le w_0 + s_0 \le 20, \quad 0 \le w_0, \quad 0 \le s_0 \\ 0 & \text{otherwise} \end{cases}$$

What poor Melvin doesn't know, and even his best friends won't tell him, is that working only furthers his confusion and that his grade, g, can be described by

$$g = 2.50(s - w) + 50.0$$

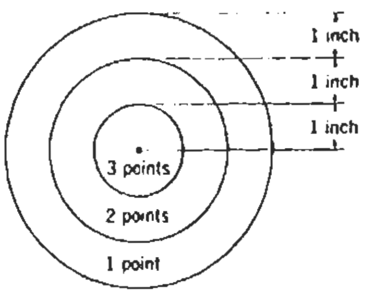
a Evaluate constant K.

b The instructor has decided to pass Melvin if, on the exam, he achieves $g \ge 75.0$. What is the probability that this will occur?

c Make a neat and fully labeled sketch of the probability density function $f_g(g_0)$.

d Melvin, true to form, got a grade of exactly 75.0 on the exam. Determine the conditional probability that he spent less than one hour working in preparation for this exam.

2.27 Each day Wyatt Uyrp shoots one "game" by firing at a target with the following dimensions and scores for each shot:



The score on any shot depends only on its distance from the center of the target

His pellet supply isn't too predictable, and the number of shots for any day's game is equally likely to be one, two, or three. Furthermore, Wyatt tires rapidly with each shot. Given that it is the kth pellet in a particular game, the value of r (distance from target center to point of impact) for a pellet is a random variable with probability density function

$$f_{r|k}(r_0 \mid k_0) = \begin{cases} \dfrac{1}{k_0} & \text{if } 0 \le r_0 \le k_0 \\ 0 & \text{otherwise} \end{cases}$$

a Determine and plot the probability mass function for random variable s_3, where s_3 is Mr. Uyrp's score on a three-shot game.

b Given only that a particular pellet was used during a three-shot game, determine and sketch the probability density function for r, the distance from the target center to where it hit.

c Given that Wyatt scored a total of exactly six points on a game, determine the probability that this was a two-shot game.

d We learn that, in a randomly selected game, there was at least one shot which scored exactly two points. Determine the conditional expected value of Wyatt's total score for that game.

e A particular pellet, marked at the factory, was used eventually by Wyatt. Determine the PMF for the number of points he scored on the shot which consumed this pellet.

2.28 Random variables x and y are described by the joint PDF

$$f_{x,y}(x_0,y_0) = \begin{cases} 1 & \text{if } 0 \le x_0 \le 1, \quad 0 \le y_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and random variable z is defined by $z = xy$.

Determine the conditional second moment of z, given that the equation $r^2 + xr + y = 0$ has real roots for r.

2.29 A target is located at the origin of an x,y cartesian coordinate system. One missile is fired at the target, and we assume that x and y, the coordinates of the missile impact point, are independent random variables each described by the *unit normal* PDF,

$$f_x(x_0) = f_y(x_0) = \frac{1}{\sqrt{2\pi}} e^{-x_0^2/2} \qquad -\infty \le x_0 \le \infty$$

Determine the PDF for random variable r, the distance from the target to the point of impact. Your answer should be an example of the *Rayleigh* PDF,

$$f_r(r_0) = \frac{r_0}{a^2} e^{-r_0^2/2a^2} \qquad r_0 \ge 0$$

2.30 Consider independent random variables x and y with the marginal PDF's

$$f_x(x_0) = f_y(x_0) = \frac{1}{\sqrt{2\pi}} e^{-x_0^2/2} \qquad -\infty \le x_0 \le \infty$$

Determine the PDF for random variable q, defined by $q = \dfrac{y}{x}$. Your answer should be a simple case of the *Cauchy* PDF,

$$f_q(q_0) = \frac{a}{\pi[a^2 + (q_0 - b)^2]} \qquad -\infty \le q_0 \le \infty$$

2.31 **a** Let x be a random variable with PDF $f_x(x_0)$. Determine the transformation $y = g(x)$ such that y will have the uniform PDF

$$f_y(y_0) = \begin{cases} 1 & \text{if } 0 \le y_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

b How could you use a set of experimental values of a uniformly distributed random variable to obtain experimental values described by an arbitrary PDF $f_x(x_0)$?

2.32 Is it generally true that $E[g(x)]$ is the same as $g[E(x)]$? For instance, is $E\left(\frac{1}{x}\right)$ the same as $\frac{1}{E(x)}$? Please remember your result and avoid one of the most common errors in probabilistic reasoning.

2.33 **a** Variable x^*, the *standardized* random variable for random variable x, is given by $x^* = [x - E(x)]/\sigma_x$. Determine the expected value and variance of x^*.

b The *correlation coefficient* ρ, or *normalized covariance*, for two random variables x and y is defined to be

$$\rho_{xy} = E(x^*y^*) = E\left[\left(\frac{x - E(x)}{\sigma_x}\right)\left(\frac{y - E(y)}{\sigma_y}\right)\right]$$

Determine the numerical value of ρ_{xy} if:

i $x = ay$.
ii $x = -ay$.
iii x and y are linearly independent.
iv x and y are statistically independent.
v $x = ay + b$.

c For each performance of the experiment, the experimental value of random variable y^* is to be approximated by cx^*. Prove that the value of constant c which minimizes the expected *mean square error*, $E[(y^* - cx^*)^2]$, for this approximation is given by $c = \rho_{xy}$.

2.34 Al and Bo are the only participants in a race, and their elapsed times may be considered to be the random variables x and y, respectively.

$$f_x(x_0) = \begin{cases} 0.0 & x_0 < 1 \\ 1.0 & 1 \le x_0 \le 2 \\ 0.0 & 2 < x_0 \end{cases} \qquad f_y(y_0) = \begin{cases} 0.0 & y_0 < 1 \\ 0.5 & 1 \le y_0 \le 3 \\ 0.0 & 3 < y_0 \end{cases}$$

Let A be the event "Al won the race."

a Determine the conditional probability density function $f_{x|A}(x_0 \mid A)$.
b Let $w = y - x$. Determine $E(w)$ and $E(w \mid A)$.
c What is the minimum number of races which Bo must agree to enter such that the a priori probability that he will win at least one of the races is at least 0.99?

2.35 By observing the histogram of a particular random variable x, it is noted that, if we define $y = \ln(x - a)$, the behavior of y may be approximated by a Gaussian density function,

$$f_y(y_0) = \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-(y_0-m)^2/2\sigma_y^2}$$

Determine the probability density function $f_x(x_0)$.

2.36 Oscar has lost his dog in either forest A (with a priori probability 1/3) or forest B (with a priori probability 2/3). The probability that the dog will survive any particular night in forest A is 4/5 and in forest B is 3/5.

If the dog is in A (either dead or alive) and Oscar spends a day searching for him in A, the probability that he will find the dog that day is 1/2. The similar detection probability for a day of search in forest B is 2/5.

The dog cannot go from one forest to the other. Oscar can search only in the daytime and can travel from one forest to the other only at night.

Coolheaded Oscar has established the following values (in dollars):

Finding dog alive	+60
Each day (or part thereof) of search	−3
Finding dog dead	0
Not finding dog	−10
Additional cost if Oscar must actually search in both forests	−3

Oscar is incapable of figuring it all out; so he decides that he will search for just two days—looking in B on the first day and, if necessary, looking in A on the second day.

a Determine the expected value of this policy.
b Given that Oscar fails to find the dog on the first day, is the second day of search a worthwhile investment? Explain.
c If only Oscar were a thinker, he would at least have considered the following list of policies for possible two-day search efforts:

S_1: Search in A on 1st day; in B on 2d day if necessary.

S_2: Search in B on 1st day; in A on 2d day if necessary.
S_3: Search in A on 1st day; in A on 2d day if necessary.
S_4: Search in B on 1st day; in B on 2d day if necessary.
S_5: Search in A on 1st day; don't search on 2d day.
S_6: Search in B on 1st day; don't search on 2d day.
S_7: Don't search at all!

Oscar would still have to choose his own decision criteria—in fact he has already decided no dog is worth more than two days of searching in mosquito-infested forests. However, to help Oscar quantify his thinking, determine which of the above policies would:

i Maximize his expected gain (graduate student)
ii Minimize his maximum possible loss (coward)
iii Maximize his maximum possible gain (hero)
iv Maximize the probability that he will find his dog alive (idealist)