2 Random Variables, Densities, and Cumulative Distribution Functions

2.1 Introduction

It is natural for the outcomes of many experiments in the real world to be measured in terms of real numbers. For example, measuring the height and weight of individuals, observing the market price and quantity demanded of a commodity, measuring the yield of a new variety of wheat, or measuring the miles per gallon achievable by a new hybrid automobile all result in real-valued outcomes. The sample spaces associated with these types of experiments are subsets of the real line or, if multiple values are needed to characterize the outcome of the experiment, subsets of n-dimensional real space, $\mathbb{R}^n$.

There are also experiments whose outcomes are not inherently numbers and whose sample space is not inherently a subset of a real space. For example, observing whether a tossed coin results in heads or tails, observing whether an item selected from an assembly line is defective or nondefective, observing the type of weeds growing in a garden, and observing which engine components caused an engine failure in an automobile are not experiments characterized inherently by real-valued outcomes. It will prove to be both convenient and useful to convert such sample spaces into real-valued sample spaces by associating a real number to each outcome in the original sample space. This process can be viewed as coding the outcomes of an experiment with real numbers.

Furthermore, the outcomes of an experiment may not be of direct interest in a given problem setting; instead, real-valued functions of the outcomes may be of prime importance. For example, in a game of craps, it is not the outcome of each die that is of primary importance, but rather the sum of the dots facing up determines whether a player has won or lost. As another example, if a firm is interested in calculating the profit associated with a given operation, it is the price of the product multiplied by the quantity sold, defining revenue, that will be of primary importance in the profit calculation, and not price and quantity, per se.

All of the previous situations involve the concept of a random variable, which can be used to characterize the ultimate experimental outcomes of interest as real numbers. We now develop the concept of a random variable.

2.2 Univariate Random Variables and Density Functions

We begin with the definition of the term **random variable** appropriate for the univariate, or one-variable, case.

Definition 2.1 ***Univariate Random Variable***

Let $\{S, \Upsilon, P\}$ be a probability space. If $X : S \to \mathbb{R}$ is a real-valued function having as its domain the elements of S, then X is a random variable.

A pictorial illustration of the random variable concept is given in Figure 2.1.

The reader might find it curious, and perhaps even consider it a misnomer, for the term "random variable" to be used as a label for the concept just given. The expression *random-valued function* would seem more appropriate since it is, after all, a real-valued *function* that is at the heart of the concept presented in the definition. Nonetheless, usage of "random variable" has become standard terminology, and we will use it also.

The phrase **outcome of the random variable** refers to the particular image element in the range of the random variable, $R(X)$, that occurs as a result of

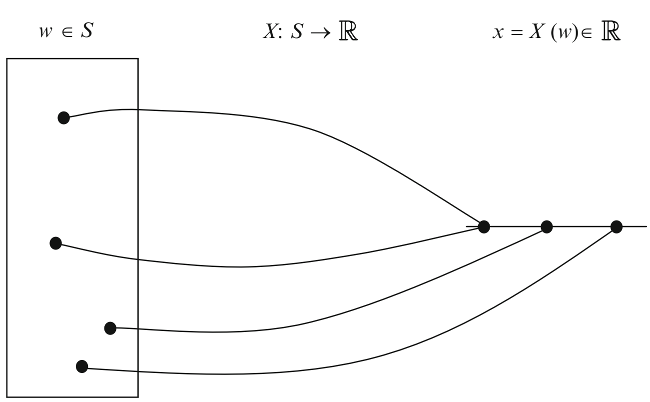


Figure 2.1
Random variable X.

observing the outcome of a given experiment, i.e., if the outcome of an experiment is $w \in S$, then the outcome of the random variable is $x = X(w)$.

Definition 2.2
Random Variable Outcome

The image $x = X(w)$ of an outcome $w \in S$ generated by a random variable X.

Henceforth, we will use upper case letters, such as X, to denote random variables and their lower case counterparts to denote an image value of the random variable, as $x = X(w)$ for $w \in S$. The letter X that we use here is arbitrary, and any other symbol could be used to denote a random variable. For the most part, we will use letters in the latter part of the alphabet for representing random variables. Letters at the beginning of the alphabet will be used to denote constants, and so the expression $x = a$ will mean that the value, x, of the random variable, X, equals the constant a. Similarly, $x \in A$ will mean that the value of X is an element of the set A.

If the outcomes of an experiment are real numbers to begin with, they are directly interpretable as outcomes of a random variable, since we can always represent the real-valued outcomes $w \in S$ as images of an identity function, e.g., $X(w) = w$. If the outcomes of an experiment are not initially in the form of real numbers, a random variable can always be defined that associates a real number with each outcome $w \in S$, as $X(w) = x$, and thus as we noted above, a random variable effectively codes the outcomes of a sample space with real numbers. Through the use of the random variable concept, all experiments with univariate outcomes can be ultimately interpreted as having sample spaces consisting of real-valued elements. In particular, the range of the random variable, $R(X) = \{x : x = X(w), w \in S\}$, represents a real-valued sample space for the experiment.

2.2.1 Probability Space Induced by a Random Variable

Given a real-valued sample space that has been defined for a given experiment via a random variable, we seek a probability space that can be used for assigning probabilities to events involving random variable outcomes. This requires that we establish how probabilities are to be assigned to subsets of the real-valued sample space $R(X)$. In so doing, we must define an appropriate *probability set function* for assigning probabilities to subsets of $R(X)$ and identify the *event space* or domain of the probability set function.

Given the probability space, $\{S, \Upsilon, P\}$, we are initially equipped to assign probabilities to events in S. What is the probability that an outcome of X resides in the set $A \subset R(X)$? Suppose an event in S can be defined, say B, that occurs *iff* the event A in $R(X)$ occurs. Then, since the two events occur only simultaneously, they must have the same probability of occurring and we can state that $P_X(A) \equiv P(B)$ when $A \Leftrightarrow B$, where $P_X(\cdot)$ is used to denote the probability set function for assigning probability to events for outcomes of X. Two events that occur only simultaneously are called **equivalent events**, where the fundamental implication of the term is that the probabilities of the events are equivalent or

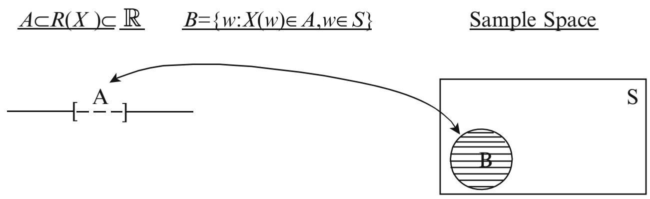


Figure 2.2 Event equivalence: event *A* and associated inverse image, *B*, for *X*.

the same. The event B in S that is equivalent to event A in $R(X)$can be defined as $B = \{w{:}X(w) \in A,\ w \in S\}$, which is the set of inverse images of the elements of A defined by the function X. By definition, $w \in B \Leftrightarrow x \in A$, and thus A and B are equivalent events (see Figure 2.2). It is clear that in order for two equivalent events to represent different sets of outcomes, they must reside in different probability spaces – if they resided in the same probability space, they could occur only simultaneously *iff* they were the same event.

Definition 2.3 ***Equivalent Events***

> Let S_1 and S_2 be different sample spaces. If $A \subset S_1$ occurs *iff* $B \subset S_2$ occurs, then A and B are said to be equivalent events.

Based on the preceding discussion, we have the following representation of probability assignments to events involving random variable outcomes:

$$P_X(A) \equiv P(B) \text{ for } B = \{w : X(w) \in A,\ w \in S\}.$$

Thus, probabilities assigned to events in S are transferred to events in $R(X)$ through the functional relationship $x = X(w)$, which relates outcomes w in S to outcomes x in $R(X)$. Note, this underscores a fundamental difference between ordinary real-valued functions and random variables, which are also real-valued functions. In particular, random variables are defined on a domain, S, that belongs to a probability space, $\{S, \Upsilon, P\}$, and thus the random variable function not only maps domain elements into image elements $x \in R(X)$, but it also maps probabilities of events in S to events in $R(X)$. An ordinary real-valued function does only the former mapping, and its domain does not reside in a probability space, and thus there is no simultaneous probability mapping.

What is the domain of $P_X(\cdot)$, i.e., what is the event space for X? It is clear from the foregoing discussion that to be able to assign probabilities to a set $A \subset R(X)$ it must be the case that its associated inverse image in S, $B = \{w : X(w) \in A,\ w \in S\}$, can be assigned probability based on the known probability space $\{S, \Upsilon, P\}$. If not, there is no basis for assigning probability to either sets B or A from knowledge of the probability space $\{S, \Upsilon, P\}$. No difficulty will arise if S is a finite or countably infinite sample space, since then the event space Υ equals the collection of *all* subsets of S, and *whatever* subset $B \subset S$ is associated with the subset $A \subset R(X)$, B can be assigned probability. Thus, *any* real-valued function defined on a discrete sample space will generate a real-valued sample space for which all subsets can be assigned probability.

Table 2.1 Relationship Between Original and X-Induced Probability Spaces

Probability space	Random variable X: $S \to \mathbb{R}$	Induced probability space
$\{S, \Upsilon, P(\cdot)\}$	$x = X(w)$	$\left\{ \begin{array}{l} R(X) = \{x : x = X(w), w \in S\} \\ \Upsilon_X = \{A : A \text{ is an event in } R(X)\} \\ P_X(A) = P(B), B = \{w : X(w) \in A, w \in S\}, \forall A \in \Upsilon_X \end{array} \right\}$

Henceforth, the event space, Υ_X, for outcomes of random variables defined on finite or countably infinite sample spaces is defined to be the set of *all* subsets of $R(X)$.

In order to avoid problems that might occur when S is uncountably infinite, one can simply restrict the types of real-valued functions that are used to define random variables to those for which the problem will not occur. To this effect, a proviso is generally added, either explicitly or implicitly, to the definition of a random variable X requiring the real-valued function defined on S to be such that for every Borel set, A, contained in $R(X)$, the set $B = \{w: X(w) \in A, w \in S\}$ is an event in S that can be assigned probability (which is to say, it is *measureable* in terms of probability). Then, since every Borel set $A \subset R(X)$ would be associated with an *event* $B \subset S$, every Borel set could be assigned a probability as $P_X(A) = P(B)$. Since the collection of Borel sets includes all intervals in $R(X)$ (and thus all points in $R(X)$), as well as all other sets that can be formed from the intervals by a countable number of union, intersection, and/or complement operations, the collection of Borel sets defines an event space sufficiently large for all real world applications.

In practice, it requires a great deal of ingenuity to define a random variable for which probability cannot be associated with each of the Borel sets in $R(X)$, and the types of functions that naturally arise when defining random variables in actual applications will generally satisfy the aforementioned proviso. Henceforth, we will assume that the event space, Υ_X, for random variable outcomes consists of all Borel sets in $R(X)$ if $R(X)$ is uncountable. We add that for all practical purposes, the reader need not even unduly worry about the latter restriction to Borel sets, since any subset of an uncountable $R(X)$ that is of practical interest will be a Borel set.

In summary, a random variable induces an alternative probability space for the experiment. The **induced probability space** takes the form $\{R(X), \Upsilon_X, P_X\}$ where the range of the random variable $R(X)$ is the real-valued sample space, Υ_X is the event space for random variable outcomes, and P_X is a probability set function defined on the events in Υ_X. The relationship between the original and induced probability spaces associated with a random variable is summarized in Table 2.1.

Example 2.1
An Induced Probability Space

Let $S = \{1, 2, 3, \ldots, 10\}$ represent the potential number of cars that a car salesperson sells in a given week, let the event space Υ be the set of all subsets of S, and let the probability set function be defined as $P(B) = (1/55)\sum_{w \in B} w$ for $B \in \Upsilon$. Suppose the salesperson's weekly pay consists of a base salary of \$100/week plus a \$100 commission for each car sold. The salesperson's weekly pay can be represented by the random variable $X(w) = 100 + 100w$, for $w \in S$. The induced probability space $\{R(X), \Upsilon_X, P_X\}$ is then characterized by $R(X) = \{200, 300, 400, \ldots, 1100\}$, $\Upsilon_X = \{A: A \subset R(X)\}$, and $P_X(A) = (1/55)\sum_{w \in B} w$ for $B = \{w: (100 + 100w) \in A, w \in S\}$ and $A \in \Upsilon_X$. Then, for example, the event that the salesperson makes $\leq$ \$300/week, $A = \{200, 300\}$, has probability $P_X(A) = (1/55)\sum_{w \in \{1,2\}} w = (3/55)$. □

A major advantage in dealing with only real-valued sample spaces is that all of the mathematical tools developed for the real number system are available when analyzing the sample spaces. In practice, once the induced probability space has been identified, the underlying probability space $\{S, \Upsilon, P\}$ is generally ignored for purposes of defining random variable events and their probabilities. In fact, we will most often choose to deal with the induced probability space $\{R(X), \Upsilon_X, P_X\}$ directly at the outset of an experiment, paying little attention to the underlying definition of the *function* having the range $R(X)$ or to the original probability space $\{S, \Upsilon, P\}$. However, we will sometimes need to return to the formal relationship between $\{S, \Upsilon, P\}$ and $\{R(X), \Upsilon_X, P_X\}$ to facilitate the proofs of certain propositions relating to random variable properties.

Note for future reference that a real-valued function of a random variable is, itself, a random variable. This follows by definition, since a real-valued function of a random variable, say Y defined by $y = Y(X(w))$ for $w \in S$, is a function of a function (i.e., a composition of functions) of the elements in a sample space S, which is then indirectly also a real-valued function of the elements in the sample space S. One might refer to such a random variable as a **composite random variable**.

2.2.2 Discrete Random Variables and Probability Density Functions

In practice, it is useful to have a representation of the probability set function, P_X that is in the form of a well-defined algebraic formula and that does not require constant reference either to events in S or to the probability set function defined on the events in S. A conceptually straightforward way of representing P_X is available when the real-valued sample space $R(X)$ contains, at most, a countable number of elements. In this case, any subset of $R(X)$ can be represented as the union of the specific elements comprising the subset, i.e., if $A \subset R(X)$ then $A = \cup_{x \in A}\{x\}$. Since the elementary events in A are clearly disjoint, we know from Axiom 1.3 that $P_X(A) = \sum_{x \in A} P_X(\{x\})$. It follows that once we know the probability of every elementary event in $R(X)$, we can assign probability to any other event in $R(X)$ by summing the probabilities of the elementary events contained in the event. This suggests that we define a *point* function $f: R(X) \rightarrow \mathbb{R}$ as $f(x) \equiv$ probability of $x = P_X(\{x\}) \, \forall x \in R(X)$. Once f is defined, then P_X can be defined for all events as $P_X(A) = \sum_{x \in A} f(x)$. Furthermore, knowledge of $f(x)$eliminates the need for any further reference to the probability space $\{S, \Upsilon, P\}$ for assigning probabilities to events in $R(X)$.

In the following example we illustrate the specification of the point function, f.

Example 2.2 ***Assigning Probabilities with a Point Function***

Examine the experiment of rolling a pair of dice and observing the number of dots facing up on each die. Assume the dice are fair. Letting i and j represent the number of dots facing up on each die, respectively, the sample space for the experiment is $S = \{(i, j) : i \text{ and } j \in \{1, 2, 3, 4, 5, 6\}\}$. Now define the random variable $x = X((i, j)) = i + j$ *for* $(i, j) \in S$. Then the following correspondence can be set up between outcomes of X, events in S, and the probability of outcomes of X and events in S, where $w = (i, j)$:

	$X(w) = x$	$B_x = \{w : X(w) = x, w \in S\}$	$f(x) = P(B_x)$
$R(X)$	2	{(1,1)}	1/36
	3	{(1,2), (2,1)}	2/36
	4	{(1,3), (2,2), (3,1)}	3/36
	5	{(1,4), (2,3), (3,2), (4,1)}	4/36
	6	{(1,5), (2,4), (3,3), (4,2), (5,1)}	5/36
	7	{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}	6/36
	8	{(2,6), (3,5), (4,4), (5,3), (6,2)}	5/36
	9	{(3,6), (4,5), (5,4), (6,3)}	4/36
	10	{(4,6), (5,5), (6,4)}	3/36
	11	{(5,6), (6,5)}	2/36
	12	{(6,6)}	1/36

The range of the random variable is $R(X) = \{2, 3, \ldots, 12\}$, which represents the collection of images of the points $(i, j) \in S$ generated by the function $x = X((i, j)) = i + j$. Probabilities of the various outcomes of X are given by $f(x) = P(B_x)$, where B_x is the collection of inverse images of x.

If we desired the probability of the event that $x \in A = \{7, 11\}$, then $P_X(A) = \sum_{x \in A} f(x) = f(7) + f(11) = 8/36$ (which, incidentally, is the probability of winning a game of craps on the first roll of the dice). If $A = \{2\}$, the singleton set representing "snake eyes," we find that $P_X(A) = \sum_{x \in A} f(x) = f(2) = 1/36$. □

In examining the outcomes of X and their respective probabilities in Example 2.2, it is recognized that a compact algebraic specification can be suggested for $f(x)$, namely[1] $f(x) = (6 - |x - 7|)/36\, I_{\{2,3,\ldots,12\}}(x)$. It is generally desirable to express the relationship between the domain and image elements of a function

[1]Notice that the algebraic specification faithfully represents the positive values of $f(x)$ in the preceding table of values, and defines $f(x)$ to equal $0 \;\forall\, x \notin \{2, 3, \ldots, 12\}$. Thus, the domain of f is the entire real line. The reason for extending the domain of f from $R(X)$ to $\mathbb{R}$ will be discussed shortly. Note that assignments of probabilities to events as $P_X(A) = \sum_{x \in A} f(x)$ are unaffected by this domain extension.

in a compact algebraic formula whenever possible, as opposed to expressing the relationship in tabular form as in Example 2.2. This is especially true if the number of elements in $R(X)$ is large. Of course, if the number of elements in the domain is infinite, the relationship cannot be represented in tabular form and must be expressed algebraically. The reader is asked to define an appropriate point function f for representing probabilities of the elementary events in the sample space $R(X)$ of Example 2.1.

We emphasize that if the outcomes of the random variable X are the outcomes of fundamental interest in a given experimental situation, then given that a probability set function, $P_X(\mathrm{A}) = \sum_{x\in A} f(x)$, has been defined on the events in $R(X)$, the original probability space $\{S, \Upsilon, P\}$ is no longer needed for defining probabilities of events in $R(X)$. Note that in Example 2.2, given $f(x)$, the probability set function $P_X(A) = \sum_{x\in A} f(x)$ can be used to define probabilities for all events $A \subset R(X)$ without reference to $\{S, \Upsilon, P\}$.

The next example illustrates a case where an experiment is analyzed exclusively in terms of the probability space relating to random variable outcomes.

Example 2.3 ***Probability Set Function Definition via Point Function***

The Bippo Lighter Co. manufactures a Piezo gas BBQ grill lighter that has a .90 probability of lighting the grill on any given attempt to use the lighter. The probability that it lights on a given trial is independent of what occurs on any other trial. Define the probability space for the experiment of observing the number of ignition trials required to obtain the first light. What is the probability that the lighter lights the grill in three or fewer trials?

Answer: The range of the random variable, or equivalently the real-valued sample space, can be specified as $R(X) = \{1, 2, 3,\ldots\}$. Since $R(X)$ is countable, the event space Υ_X will be defined as the set of all subsets of $R(X)$. The probability that the lighter lights the grill on the first attempt is clearly .90, and so $f(1) = .90$. Using independence of events, the probability it lights for the first time on the second trial is $(.10)\,(.90) = .09$, on the third trial is $(.10)^2\,(.90) = .009$, on the fourth trial is $(.10)^3\,(.90) = .0009$, and so on. In general, the probability that it takes x trials to obtain the first light is $f(x) = (.10)^{x-1}\,.90\, I_{\{1,2,3\ldots\}}(x)$. Then the probability set function is given by $P_X(A) = \sum_{x\in A} (.10)^{x-1} .90 I_{\{1,2,3,\ldots\}}(x)$. The event that the lighter lights the grill in three trials or less is represented by $A = \{1, 2, 3\}$. Then $P_X(A) = \sum_{x=1}^{3} (.10)^{x-1} .90 = .999$. □

The preceding examples illustrate the concept of a **discrete random variable** and a **discrete probability density function**, which we formalize in the following definitions.

Definition 2.4 ***Discrete Random Variable***

A random variable is called discrete if its range consists of a countable number of elements.

Definition 2.5 ***Discrete Probability Density Function***

The discrete probability density function, f, is defined as $f(x) \equiv$ probability of x, $\forall x \in R(X)$, and $f(x) = 0$, $\forall x \notin R(X)$.

Note that in the case of discrete random variables, some authors refer to $f(x)$ as a **probability mass function** as opposed to a **discrete probability density function**. We will continue to use the latter terminology.

It should be noted that even though there is only a countable number of elements in the range of the discrete random variable, X, the probability density function (PDF) defined here has the entire (uncountable) real line for its domain. The value of f at a point x in the range of the random variable is the probability of x, while the value of f is zero at all other points on the real line. This definition is adopted for the sake of mathematical convenience – it standardizes the domain of all discrete density functions to be the real line while having no effect on the assignment of event probabilities made via the set function $P_X(A) = \sum_{x \in A} f(x)$. This convention will provides a considerable simplification in the definition of marginal and conditional density functions which we will examine ahead.

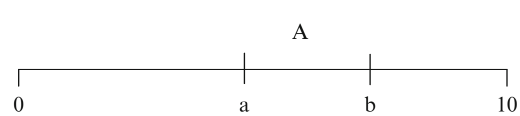
In our previous examples, the probability space for the experiment was a priori deducible under the stated assumptions of the problems. It is most often the case in practice that the probability space is not a priori deducible, and an important problem in statistical inference is the identification of the appropriate density function, $f(x)$, to use in defining the probability set function component of the probability space.

2.2.3 Continuous Random Variables and Probability Density Functions

So far, our discussion concerning the representation of P_X in terms of the point function, $f(x)$, is applicable only to those random variables that have a countable number of possible outcomes. Can P_X be similarly represented when the range of X is uncountably infinite? Given that we can have an event A defined as an uncountable subset of $R(X)$, it is clear that the summation operation over the elements of the set, (i.e., $\sum_{x \in A}$) is not generally defined. Thus, defining a probability set function on the events in $R(X)$ as $P(A) = \sum_{x \in A} f(x)$ will not be possible. However, integration over uncountable sets is possible, suggesting that the probability set function might be defined as $P(A) = \int_{x \in A} f(x)dx$ when $R(X)$ is uncountably infinite. In this case the point function $f(x)$ would be defined so that $\int_{x \in A} f(x)\, dx$ defines the probability of the event A. The following example illustrates the specification of such a point function $f(x)$ when $R(X)$ is uncountably infinite.

Example 2.4
Probabilities by Integrating a Point Function

Suppose a trucking company has observed that accidents are equally likely to occur on a certain 10-mile stretch of highway, beginning at point 0 and ending at point 10. Let $R(X) = [0,\ 10]$ define the real-valued sample space of potential accident points.



It is clear that given all points are equally likely, the probability set function should assign probabilities to intervals of highway, say A, in such a way that the probability of an accident is equal to the proportion of the total highway length represented by the stretch of highway, A, as

$$P_X(A) = \frac{\text{length of } A}{10} = \frac{b-a}{10}, \quad \text{for } A = [a, b].$$

If we wish to assign these probabilities using $P_X(A) = \int_{x \in A} f(x)dx$, we require that $\int_a^b f(x)dx \equiv \frac{b-a}{10}$ for all $0 \leq a \leq b \leq 10$. The following lemma will be useful in deriving the explicit functional form of $f(x)$:

Lemma 2.1
Fundamental Theorem of Calculus

Let $f(x)$ be a continuous function at b and a, respectively.[2] Then $\frac{\partial \int_a^b f(x)dx}{\partial b} = f(b)$ and $\frac{\partial \int_a^b f(x)dx}{\partial a} = -f(a)$.

Applying the lemma to the preceding integral identity yields

$$\frac{\partial \int_a^b f(x)dx}{\partial b} = f(b) \equiv \frac{\partial \left(\frac{b-a}{10}\right)}{\partial b} = \frac{1}{10} \quad \forall\ b \in [0, 10],$$

which implies that the function defined by $f(x) = .1\, I_{[0,10]}(x)$ can be used to define the probability set function $P_X(A) = \int_{x \in A} .1\, dx$, for $A \in \Upsilon_X$. For an example of the use of this representation, the probability that an accident occurs in the first half of the stretch of highway, i.e., the probability of the event $A = [0, 5]$, is given by $P_X(A) = \int_0^5 .1 dx = .5$. □

The preceding example illustrates the concept of a **continuous random variable** and a **continuous probability density function,** which we formalize in the next definition.

Definition 2.6
Continuous Random Variables and Continuous Probability Density Functions

A random variable is called continuous if (1) its range is uncountably infinite, and (2) there exists a nonnegative-valued function $f(x)$, defined for all $x \in (-\infty, \infty)$, such that for any event $A \subset R(X)$, $P_X(A) = \int_{x \in A} f(x)\, dx$, and $f(x) = 0\ \forall\, x \notin R(X)$. The function $f(x)$ is called a continuous probability density function.

Clarification of a number of important characteristics of continuous random variables is warranted. First of all, note that probability in the case of a

[2] See F.S. Woods (1954) *Advanced Calculus*, Boston: Ginn and Co., p. 141. Regarding **continuity** of $f(x)$, note that $f(x)$ is continuous at a point $d \in D(f)$ if, $\forall \varepsilon > 0$, $\exists$ a number $\delta(\varepsilon) > 0$ such that if $|x - d| < \delta(\varepsilon)$, then $f(x) - f(d) < \varepsilon$. The function f is continuous if it is continuous at every point in its domain. Heuristically, a function will be continuous if there are no breaks in the graph of $y = f(x)$. Put another way, if the graph of $y = f(x)$ can be completely drawn without ever lifting a pencil from the graph paper, then f is a continuous function.

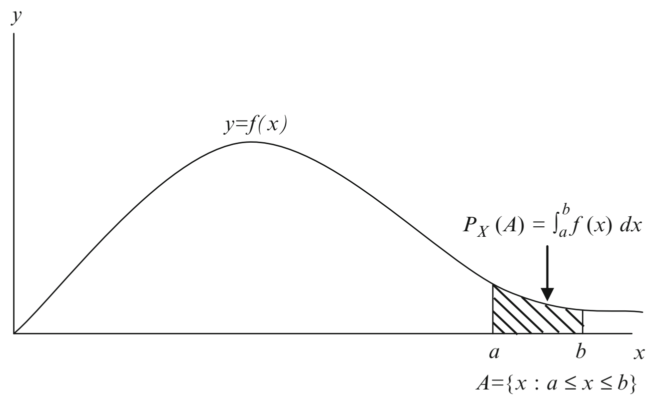


Figure 2.3 Probability represented as area.

continuous random variable is represented by the area under the graph of the density function f and above the points in the set A, as illustrated in Figure 2.3.

Of course, the event in question need not be an interval, but given our convention regarding the event space Υ_X, the event will be a Borel set for which an integral can be defined. A justification for the existence of the integral for Borel sets is beyond the scope of this text, but implementation of the integration process in these cases is both natural and straightforward.[3] The next example illustrates the procedure of determining probabilities for events more complicated than a single interval.

Example 2.5 ***Probabilities for Non-Continuous Events***

Reexamine the highway accident example (Example 2.4) where $R(X) = [0,10]$ and $f(x) = .1\, I_{[0,10]}(x)$.

a. What is the probability of $A = [1,2] \cup [7,9]$? The probability of A is given by the area above the points in A and below the graph of f, i.e.,

$$P_X(A) = \int_{x\in A} f(x)dx = \int_1^2 \left(\frac{1}{10}\right)dx + \int_7^9 \left(\frac{1}{10}\right)dx = .1 + .2 = .3$$

b. Given A defined above, what is the probability of $\bar{A} = [0,1) \cup (2,7) \cup (9,10]$? The area representing the probability in question is calculated as

$$P_X(\bar{A}) = \int_{x\in\bar{A}} f(x)dx = \int_0^1 \left(\frac{1}{10}\right)dx + \int_2^7 \left(\frac{1}{10}\right)dx + \int_9^{10} \left(\frac{1}{10}\right)dx$$
$$= .1 + .5 + .1 = .7.$$
□

A consequence of the definition of the probability set function P_X in Definition 2.6 is that, for a continuous random variable, the probability of any elementary event is zero, i.e., if $A = \{a\}$, then $P_X(A) = \int_a^a f(x)dx = 0$. Note this certainly does

[3] It can be shown that Borel sets are representable as the union of a collection of disjoint intervals, some of which may be single points. The collective area in question can then be defined as the sum of the areas lying above the various intervals and below the graph of f.

not imply that every outcome of X in $R(X)$ is impossible, since some elementary event in $R(X)$ will occur as a result of a given experiment. Instead, $P_X(\{x\}) = 0\ \forall x \in R(X)$ suggests that zero probability is not synonymous with impossibility. In cases where an event, say A, can occur, but the probability set function assigns the event the value zero, we say **event A occurs with probability zero**. The reader might intuitively interpret this to mean that event A is relatively impossible, i.e., relative to the other outcomes that can occur $(R(X) - A)$, the likelihood that A would occur is essentially nil. Note that the above argument together with Theorem 1.1 then suggest that if $P_X(A) = 1$ for a continuous random variable, it does *not* follow that event A is *certain* to occur. In the spirit of our preceding discussion, if event A is assigned a probability of 1, we say **event A occurs with probability 1**, and if in addition $A \neq R(X)$, we might interpret this to mean that event A is relatively certain.

Note that an important implication of the preceding property for continuous random variables, which has already been utilized in Example 2.5b, is that the sets $[a,b]$, $(a,b]$, $[a,b)$, and (a,b) are all assigned the same probability value $\int_a^b f(x)dx$ since adding or removing a finite number of elementary events to another event will be adding or removing a collection of outcomes that occur with probability zero. That is, since $[a,b] = (a,b] \cup \{a\} = [a,b) \cup \{b\} = (a,b) \cup \{a\} \cup \{b\}$, and since $P_X(\{a\}) = P_X(\{b\}) = 0$, Axiom 1.3 implies that $P_X([a,b]) = P_X((a,b]) = P_X([a,b)) = P_X((a,b))$, so that the integral $\int_a^b f(x)dx$ suffices to assign the appropriate probability to all four interval events.

There is a fundamental difference in the interpretation of the image value $f(x)$ depending on whether f is a discrete or continuous PDF. In particular, while $f(x)$ is the probability of the outcome x in the discrete case, $f(x)$ is *not* the probability of x in the continuous case. To motivate this latter point, recognize that if $f(x)$ were the probability of outcome x in the continuous case, then by our argument above, $f(x) = 0\ \ \forall\, x \in R(X)$ since the probability of elementary events are zero. But this would imply that for every event A, including the certain event $R(X)$, $P_X(A) = \int_{x\in A} f(x)dx = \int_{x\in A} 0dx = 0$, since having an integrand of 0 ensures that the integral has a zero value. The preceding property would contradict the interpretation of P_X as a probability set function, and so $f(x)$ is clearly not interpretable as a probability. It is interpretable as a density function value, but nothing more – the continuous PDF must be integrated to define probabilities.

As in the discrete case, a continuous PDF has the entire real line for its domain. Again, this convention is adopted for the sake of mathematical convenience, as it standardizes the domain of all continuous density functions while leaving probabilities of events unaffected. It also simplifies the definition of marginal and conditional probability density functions, which we will examine ahead. We now provide another example of a continuous random variable together with its density function, where the latter, we will assume, has been discovered by your personnel department.

Example 2.6
Probabilities of Lower and Upper Bounded Events

Examine the experiment of observing the amount of time that passes between employee work-related injuries at a metal fabricating plant. Let $R(X) = \{x : x \geq 0\}$ represent the potential outcomes of the experiment measured in hours, and let the density of the continuous random variable be given by

$$f(x) = \frac{1}{100}\, e^{-x/100}\, I_{(0,\infty)}\,(x).$$

(a) What is the probability of the event that 100 or more hours pass between work related injuries? Letting $A = \{x{:}x \geq 100\}$ represent the event in question,

$$P(A) = \int_{100}^{\infty} \frac{1}{100}\, e^{-x/100}\, dx = -\, e^{-x/100}\, |_{100}^{\infty} = -\, e^{-\infty/100} + e^{-1} = e^{-1} = .37.$$

(b) What is the probability that an injury occurs within 50 hours of the previous injury? Letting $B = \{x : 0 \leq x \leq 50\}$ represent the event in question,

$$P(B) = \int_{0}^{50} \frac{1}{100}\, e^{-x/100}\, dx = -\, e^{-x/100}\, |_{0}^{50} = -\, e^{-50/100} + e^{-0} = 1 - .61 = .39.$$

□

2.2.4 Classes of Discrete and Continuous PDFs

In our later study of statistical inference, we will generally identify an appropriate range for a random variable based on the characteristics of a particular experiment being analyzed and have as an objective the identification of an appropriate $f(x)$ with which to complete the specification of the probability space. The fact that for all events $A \subset R(X)$ the values generated by $\sum_{x \in A} f(x)$ or $\int_{x \in A} f(x)dx$ must adhere to the probability axioms places some general restrictions on the types of functions that can be used as density functions, regardless of the specific characteristics of a given experiment. These general restrictions on the admissible choices of $f(x)$ are identified in the following definition.

Definition 2.7
The Classes of Discrete and Continuous Probability Density Functions: Univariate Case

a. Class of Discrete Density Functions. The function $f : \mathbb{R} \rightarrow \mathbb{R}$ is a member of the class of discrete density functions *iff* (1) the set $C = \{x{:}f(x) > 0, x \in \mathbb{R}\}$ (i.e., the subset of points in $\mathbb{R}$ having a positive image under f) is countable, (2) $f(x) = 0$ for $x \in \overline{C}$, and (3) $\sum_{x \in C} f(x) = 1$.

b. Class of Continuous Density Functions. The function $f : \mathbb{R} \rightarrow \mathbb{R}$ is a member of the class of continuous density functions *iff* (1) $f(x) \geq 0$ for $x \in (-\infty, \infty)$, and (2) $\int_{-\infty}^{\infty} f(x)dx = 1$.

Note for future reference that the set of outcomes for which the PDF of a random variable assigns *positive* density weightings, i.e., $\{x : f(x){>}0, x \in \mathbb{R}\}$, is called **the support of the random variable**.

Definition 2.8
Support of a Random Variable

The set $\{x : f(x){>}0, x \in \mathbb{R}\}$ is called the support of the random variable.

Thus, in Definition 2.7, the set C is the support of the discrete random variable X when $f(x)$ is the PDF of X. We will henceforth adopt the convention that the range of a random variable is synonymous with its support. This simply implies that any value of X for which $f(x) = 0$(probability zero in the discrete case, and probability density equal to zero in the continuous case) is not part of the range $R(X)$, and is thus not considered a relevant outcome of the random variable.[4] We formalize this equivalence in the following definition.

Definition 2.9
Support and Range Equivalence

$$R(X) \equiv \{x : f(x){>}0 \textit{ for } x \in \mathbb{R}\}$$

Some clarifying remarks concerning Definition 2.7 are warranted. First, it should be noted that the definition simply identifies the respective classes of function specifications that are *candidates* for use as PDFs. The *specific* functional form of the density function appropriate for a real-world experimental situation depends on the particular characteristics of the process generating the outcomes of the experiment.

A second observation concerns the fact that the definitions focus exclusively on real-valued functions having the entire real line for their domains. As we discussed earlier, this is a convention adopted as a matter of mathematical convenience. To ensure that subsets of points outside of the range of X are properly assigned zero probability, all one needs to do is to extend the domain of f to the remaining points $\mathbb{R} - R(X)$ on the real line by assigning to each point a zero density weighting, i.e., $f(x) = 0$ if $x \in \overline{R(X)}$.

A final remark concerns the rationale in support of the properties that are required for a function f to be considered a PDF. The properties are imposed on f: $\mathbb{R} \rightarrow \mathbb{R}$ to ensure that the set functions constructed from f, i.e., $P_X(A) = \sum_{x \in A} f(x)$ or $P_X(A) = \int_{x \in A} f(x)\mathrm{dx}$, are in fact *probability* set functions, which of course requires that the probability assignments adhere to the axioms of probability. To motivate the *sufficiency* of these conditions, first examine the discrete case. Since $f(x) \geq 0\ \forall x$, $P_X(A) = \sum_{x \in A} f(x) \geq 0$ for any event, A, and Axiom 1.1 is satisfied. Letting $R(X)$ equal the set C defined in Definition 2.7.a, it follows that $\mathrm{P_X}(R(X)) = \sum_{x \in R(X)} f(x) = 1$, satisfying Axiom 1.2. Finally, if $\cup_{i \in I} A_i$ is the union of a collection of disjoint events indexed by the index set I, then summing over all of the elementary events in $A = \cup_{i \in I} A_i$ obtains $P_X(\cup_{i \in I} A_i) = \sum_{x \in A} f(x) = \sum_{i \in I}\left(\sum_{x \in A_i} f(x)\right) = \sum_{i \in I} P_X(A_i)$. Satisfying Axiom 1.3. Thus, the three probability axioms are satisfied, and P_X is a probability set function.

To motivate *sufficiency* in the continuous case, first note that Axiom 1.1 is satisfied since if $f(x) \geq 0\ \forall x$, then $P_X(A) = \int_{x \in A} f(x)dx \geq 0$ because integrating a

[4]Note that in the discrete case, it is conceptually possible to define a random variable that has an outcome that occurs with zero probability. For example, if $f(y) = I_{[0,1]}(y)$is the density function of the continuous random variable Y, then $X = I_{[0,1)}(Y)$is a discrete random variable that takes the value 1 with probability 1 and the value 0 with probability zero. Such random variables have little use in applications, and for simplicity, we suppress this possibility in making the range of the random variable synonymous with its support.

nonnegative integrand over any interval (or Borel) set, A, results in a nonnegative number. Furthermore, since $\int_{-\infty}^{\infty} f(x)dx = 1$, there exists at least one event $A \subset (-\infty,\infty)$ such that $\int_{x \in A} f(x)dx = 1$ (the event can be $(-\infty,\infty)$ itself, or else there may be some other partition of $(-\infty,\infty)$ into $A \cup B$ such that $\int_{x \in A} f(x)dx = 1$ and $\int_{x \in B} f(x)dx = 0$). Letting $R(X) = A$, we have that $P_X(R(X)) = \int_{x \in R(X)} f(x)dx = 1$ and Axiom 1.2 is satisfied. Finally, if $D = \cup_{i \in I} A_i$ is the union of a collection of disjoint events indexed by the index set I, then by the additive property of integrals, $P_X(D) = \int_{x \in D} f(x)dx = \sum_{i \in I}\left(\int_{x \in A_i} f(x)dx\right) = \sum_{i \in I} P_X(A_i)$ satisfying Axiom 1.3. Thus, the three probability axioms are satisfied, and P_X is a probability set function.

It can also be shown that the function properties presented in Definition 2.7 are actually *necessary* for the discrete case and *practically necessary* in the continuous case. For the discrete case, first recall that $f(x)$ is directly interpretable as the probability of the outcome x, and this requires that $f(x) \geq 0 \; \forall x \in \mathbb{R}$ (or else we would be assigning negative probabilities to some x's). Second, the number of outcomes that can receive positive probability must be countable in the discrete case since $R(X)$ is countable, leading to the requirement that $C = \{x : f(x) > 0,\; x \in R\}$ is countable. Finally, $\sum_{x \in C} f(x) = 1$ is required if the probability assigned to the certain event is to equal one.

In the continuous case, it is necessary that $\int_{-\infty}^{\infty} f(x)dx = 1$. To see this, first note that $R(X)$ is the certain event, implying $P(R(X)) = 1$. Now since $R(X)$ and $\overline{R(X)}$ are disjoint, we have that $P_X(R(X) \cup \overline{R(X)}) = P_X(R(X)) + P_X(\overline{R(X)}) = 1 + P_X(\overline{R(X)})$, which implies $P_X(\overline{R(X)}) = 0$ since probabilities cannot exceed 1. But, since $R(X) \cup \overline{R(X)} = \mathbb{R}$ by definition, then $P(R(X)) \cup (\overline{R(X)}) = \int_{-\infty}^{\infty} f(x)dx = 1$. Regarding the requirement that $f(x) \geq 0$ for $x \in (-\infty,\infty)$, note that the condition is technically *not necessary*. It is known from the properties of integrals that the value of $\int_a^b f(x)dx$ is invariant to changes in the value of $f(x)$ at a finite number of isolated points, and thus $f(x)$ could technically be negative for such a finite number of x values without affecting the values of the probability set function. As others do, we will ignore this technical anomaly since its practical significance in defining PDFs is nil. We thus insist, as a practical matter, on the nonnegativity of $f(x)$.

Example 2.7 ***Verifying Probability Density Functions***

In each case below, determine whether the stated function can serve as a PDF:

a. $f(x) = 1/2 I_{[0,2]}(x)$.
Answer: The function can serve as a continuous probability density function since $f(x) \geq 0 \; \forall x \in (-\infty,\infty)$ (note $f(x) = 1/2 > 0 \; \forall x \in [0,2]$ and $f(x) = 0$ for $x \notin [0,2]$), and $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} (1/2)I_{[0,2]}(x)dx = \int_0^2 1/2 dx = x/2|_0^2 = 1$.

b. $f(x) = (.3)^x(.7)^{1-x} I_{\{0,1\}}(x)$.
Answer: The function can serve as a discrete probability density function, since $f(x) > 0$ on the countable set $\{0, 1\}$, $\sum_{x=0}^{1} f(x) = 1$, and $f(x) = 0 \; \forall x \notin \{0, 1\}$.

c. $f(x) = (x^2 + 1)I_{[-1,1]}(x)$.
Answer: The function *cannot* serve as a PDF. While $f(x) \geq 0 \; \forall x \in (-\infty,\infty)$, the function does not integrate to 1:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} (x^2+1)\, I_{[-1,1]}\,(x)dx$$
$$= \int_{-1}^{1} (x^2+1)dx$$
$$= \frac{x^3}{3} + x\,|_{-1}^{1} = \frac{8}{3} \neq 1$$

d. $f(x) = (3/8)(x^2+1)I_{[-1,1]}(x)$.
Answer: The reader should demonstrate that this function *can* serve as a continuous probability density function. Note its relationship to the function in part *c*. □

2.2.5 Mixed Discrete-Continuous Random Variables

The categories of *discrete* and *continuous* random variables do not exhaust the possible types of random variables. There is a category of random variable called *mixed discrete-continuous* which exhibits the characteristics of a discrete random variable for some events and the characteristics of a continuous random variable for other events. In particular, a mixed discrete-continuous random variable is such that a countable subset of the elementary events are assigned positive probabilities, as in the case of a discrete random variable, *except* the sum of the probabilities over the countable set does *not* equal 1. The remaining probability is attributable to an uncountable collection of elementary events, each elementary event being assigned zero probability, as in the case of a continuous random variable. The following example illustrates the concept of a mixed discrete-continuous random variable.

Example 2.8
Operating Life as a Mixed Discrete-Continuous RV

Let X be a random variable representing the length of time, measured in units of one hundred thousand hours, that a LCD color screen for a laptop computer operates properly until failure. Assume the probability set function associated with the random variable is $P_X(A) = .25\ I_A(0) + .75 \int_{x\in A} e^{-x} I_{(0,\infty)}(x)dx$ for every event A (i.e., Borel set) contained in $R(X) = [0,\infty)$.

a. What is the probability that the color screen is defective, i.e., it does not function properly at the outset?
Answer: The event in question is $A = \{0\}$. Using P_X, we calculate the probability to be $P_X(\{0\}) = .25\ I_{\{0\}}(0) + .75 \int_{x\in\phi} e^{-x}dx = .25$. (Note: By definition, $\int_{x\in\phi} f(x)\mathrm{d}x = 0$).

b. What is the probability that the color screen operates satisfactorily for less than 100,000 hours?
Answer: Here, $A = [0,1)$. Using P_X, we calculate $P_X([0,1)) = .25\ I_{[0,1)}\ (0) + .75 \int_0^1 e^{-x}\ dx = .25 + .474 = .724$.

c. What is the probability that the color screen operates satisfactorily for at least 50,000 hours?

Answer: The event in question is $A = [.5,\infty)$. The probability assigned to this event is given by $P_X([.5,\infty)) = .25 \; I_{[.5,\infty)}(0) + .75 \int_{.5}^{\infty} e^{-x} dx = 0 + .4549 = .4549.$ □

We formalize the concept of a mixed discrete-continuous random variable in the following definition.

Definition 2.10 ***Mixed Discrete-Continuous Random Variables***

A random variable is called mixed discrete-continuous *iff*

a. Its range is uncountably infinite;

b. There exists a countable set, C, of outcomes of X such that $P_X(\{x\}) = f_d(x) > 0 \; \forall x \in C$, $f_d(x) = 0 \; \forall x \notin C$, and $\sum_{x \in C} f_d(x) < 1$, where the function f_d is referred to as the **discrete density function component** of the probability set function of X;

c. There exists a nonnegative-valued function, f_c, defined for all $x \in (-\infty,\infty)$ such that for every event $B \subset R(X) - C$, $P_X(B) = \int_{x \in B} f_c(x)dx$, $f(x) = 0 \; \forall x \in \mathbb{R} - R(X)$, and $\int_{-\infty}^{\infty} f_c(x)dx = 1 - \sum_{x \in C} f_d(x)$, where the function f_c is referred to as the **continuous density function component** of the probability set function of X; and

d. The probability set function for X is given by combining or *mixing* the discrete and continuous density function components in (b) and (c) above, as $P_X(A) = \sum_{x \in A \cap C} f_d(x) + \int_{x \in A} f_c(x)dx$ for every event A.

To see how the definition applies to a specific experimental situation, recall Example 2.8. If we substitute $f_d(x) = .25 \, I_{\{0\}}(x)$, $C = \{0\}$, and $f_c(x) = .75e^{-x} I_{(0,\infty)}(x)$ into the definition of P_X given in Definition 2.10.d, we obtain

$$P_X(A) = \sum_{x \in A \cap \{0\}} (.25 I_{\{0\}}(x)) + .75 \int_{x \in A} e^{-x} I_{(0,\infty)}(x)dx = .25 I_A(0) + .75 \int_{x \in A} e^{-x} I_{(0,\infty)}(x)dx,$$

which is identical to the probability set function defined in Example 2.8.

As the reader may have concluded from examining the definition, the concept of a mixed discrete-continuous random variable is more complicated than either the discrete or continuous random variable case, since there is no single PDF that can either be summed or integrated to define probabilities of events.[5] On the other hand, once the discrete and continuous random variable concepts are understood, the notion of a mixed discrete-continuous random variable is a rather straightforward conceptual extension. Note that the definition of the probability set function in Definition 2.10.d essentially implies that the probability of an event A is equivalent to adding together the probabilities of the

[5] In a more advanced treatment of the subject, we could resort to more general integration methods, in which case a single *integral* could once again be used to define P_X. On Stieltjes integration, see R.G. Bartle (1976) *The Elements of Real Analysis*, 2nd ed., New York: John Wiley, and Section 3.2 of the next chapter.

discrete event $A \cap C$ and the continuous event A. Assigning probability to the event $A \cap C$ is done in a way that emulates the discrete random variable case – a real-valued function (the discrete density component) is summed over the points in the event $A \cap C$. The probability of the event A is calculated in a way that emulates the continuous random variable case – a real-valued function (the continuous density component) is integrated over the points in the event A. Adding together the results obtained for the discrete event $A \cap C$ and the continuous event A defines the probability of the "mixed" event A. Note that the overlap of discrete points $A \cap C$ in the event A is immaterial when the probability of A is assigned via the continuous PDF component since $\int_{x \in A-(A \cap C)} f_c(x)dx = \int_{x \in A} f_c(x)dx$, i.e., the integral over the countable points in $A \cap C$ will be zero.[6]

2.3 Univariate Cumulative Distribution Functions

Situations arise in practice that require finding the probability that the outcome of a random variable is less than or equal to some real number, i.e., the event in question is $\{x\colon x \leq b,\ x \in R(X)\}$ for some real number b. These types of probabilities are provided by the **cumulative distribution function** (CDF), which we introduce in this section.

Henceforth, we will eliminate the random variable subscript used heretofore in our probability set function notation; we will now write $P(A)$ rather than $P_X(A)$ whenever the context makes it clear to which probability space the event A refers. Thus, the notation $P(A)$ will be used to represent the probability of either an event $A \subset S$ or an event $A \subset R(X)$. To economize on notation further, we introduce an **abbreviated set definition** for representing events.

Definition 2.11
Abbreviated Set Definition for Events

For an event $\{x\colon$ set defining conditions, $x \in R(X)\}$ and associated probability represented by $P(\{x\colon$ set defining conditions, $x \in R(X)\})$, the **abbreviated set definition** for the event and associated probability are respectively {set-defining conditions} and P(set-defining conditions), the condition $x \in R(X)$ always being tacitly assumed. Alternatively, S may appear in place of $R(X)$.

For an example of an abbreviated set definition that is particularly relevant to our current discussion of CDFs, note that $\{x \leq b\}$ will be used to represent $\{x\colon x \leq b,\ x \in R(X)\}$, and $P(x \leq b)$ will be used to represent $P(\{x\colon x \leq b,\ x \in R(X)\})$.[7]

---

[6]There are still other types of random variables besides those we have examined, but they are rarely utilized in applied work. See T.S. Chow and H. Teicher (1978) *Probability Theory*, New York: Springer-Verlag, pp. 247–248.

[7]Alternative shorthand notation that is often used in the literature is respectively $\{X \leq b\}$ and $P(X \leq b)$. Our notation establishes a distinction between the function X and a value of the function x.

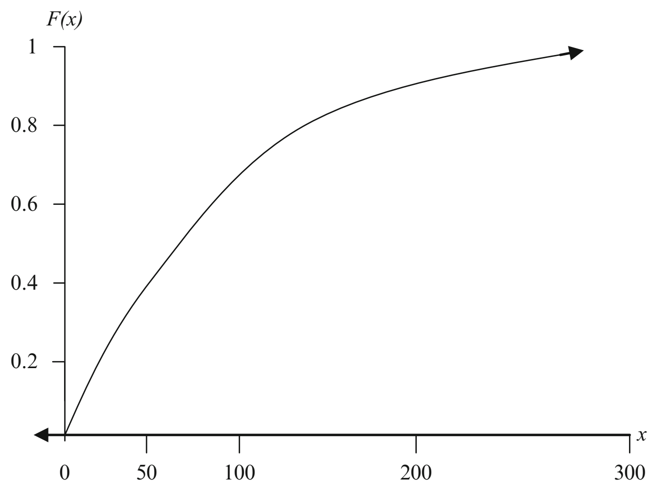


Figure 2.4
A CDF for a continuous X.

The formal definition of the cumulative distribution function, and its particular algebraic representations in the discrete, continuous, and mixed discrete-continuous cases, are given next.

Definition 2.12
Univariate Cumulative Distribution Function

The cumulative distribution function of a random variable X is defined by $F(b) \equiv P(x \leq b) \ \forall b \in (-\infty,\infty)$. The functional representation of $F(b)$ in particular cases is as follows:

a. *Discrete*: $F(b) = \sum_{x \leq b, f(x)>0} f(x), \ b \in (-\infty,\infty)$

b. *Continuous*: $F(b) = \int_{-\infty}^{b} f(x)dx, \ b \in (-\infty,\infty)$

c. *Mixed discrete-continuous*: $F(b) = \sum_{x \leq b, f_d(x)>0} f_d(x) + \int_{-\infty}^{b} f_c(x)dx$, $b \in (-\infty,\infty)$.

Example 2.9
CDF for Continuous RV

Reexamine Example 2.6, where the amount of time that passes between work-related injuries is observed. We can define the cumulative distribution function for X as

$$F(b) = \int_{-\infty}^{b} \frac{1}{100} e^{-x/100} \, I_{(0,\infty)}(x)dx = \left[1 - e^{-b/100}\right] I_{(0,\infty)}(b).$$

If one were interested in the event that an injury occurs within 50 hours of the previous injury, the probability would be given by

$F(50) = [1 - \mathrm{e}^{-50/100}] I_{(0,\infty)}(50) = 1 - .61 = .39.$

A graph of the cumulative distribution function is given in Figure 2.4. □

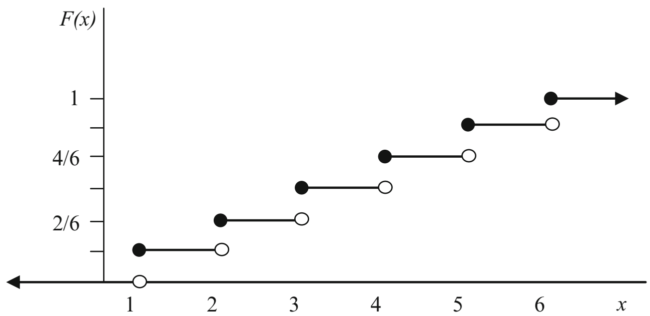


Figure 2.5
A CDF for a discrete X.

Example 2.10
CDF for Discrete RV

Examine the experiment of rolling a fair die and observing the number of dots facing up. Let the random variable X represent the possible outcomes of the experiment, so that $R(X) = \{1, 2, 3, 4, 5, 6\}$ and $f(x) = 1/6\ I_{\{1,2,3,4,5,6\}}(x)$. The cumulative distribution function for X can be defined as

$$F(b) = \sum_{x \le b,\ f(x)>0} \frac{1}{6} I_{\{1,2,3,4,5,6\}}(x) = \frac{1}{6}\text{trunc}(b) I_{[0,6]}(b) + I_{(6,\infty)}(b),$$

where trunc(b) is the **truncation function** defined by assigning to any domain element b the number that results after truncating the decimal part of b. For example, trunc(5.97) = 5, or trunc(−2.12) = −2. If we were interested in the probability of tossing a 3 or less, the probability would be given by

$$F(3) = \frac{1}{6}\ \text{trunc}(3) I_{[0,6]}(3) + I_{(6,\infty)}(3) = \frac{1}{2} + 0 = \frac{1}{2}.$$

A graph of the cumulative distribution function is given in Figure 2.5. □

Example 2.11
CDF for a Mixed Discrete Continuous RV

Recall Example 2.8, where color screen lifetimes were represented by a mixed discrete-continuous random variable. The cumulative distribution for X is given by

$$\begin{aligned} F(b) &= .25 I_{[0,\infty)}(b) + .75 \int_{-\infty}^{b} e^{-x} I_{(0,\infty)}(x)dx \\ &= .25 I_{[0,\infty)}(b) + .75\left[1 - e^{-b}\right] I_{(0,\infty)}(b). \end{aligned}$$

If one were interested in the probability that the color screen functioned for 100,000 hours or less, the probability would be given by

$$\begin{aligned} F(1) &= .25 I_{[0,\infty)}(1) + .75\left[1 - e^{-1}\right] I_{(0,\infty)}(1) \\ &= .25 + .474 = .724. \end{aligned}$$ □

A graph of the cumulative distribution function is given in Figure 2.6.

2.3.1 CDF Properties

The graphs in the preceding examples illustrate some general properties of CDFs. First, CDFs have the entire real line for their domain, while their range is contained in the interval [0, 1]. Secondly, the CDF exhibits limits as

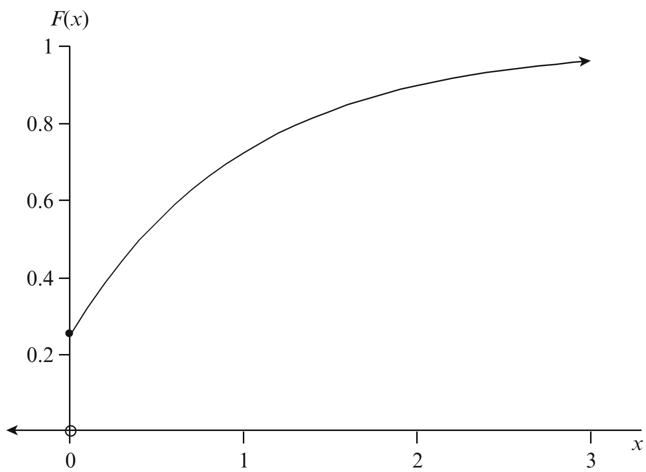


Figure 2.6
A CDF for a mixed discrete-continuous *X*.

$$\lim_{b\to-\infty} F(b) = \lim_{b\to-\infty} P(x \le b) = P(\emptyset) = 0$$

and

$$\lim_{b\to\infty} F(b) = \lim_{b\to\infty} P(x \le b) = P(R(X)) = 1.$$

It is also true that if $a < b$, then necessarily $F(a) = P(x \le a) \le P(x \le b) = F(b)$, which is the defining property for F to be an **increasing function**, i.e., if $\forall x_i$ and x_j for which $x_i < x_j$, $F(x_i) \le F(x_j)$, F is an increasing function.[8]

The CDFs of discrete, continuous, and mixed discrete-continuous random variables can be distinguished by their continuity properties and by the behavior of $F(b)$ on sets of domain elements for which F is continuous. The CDF of a continuous random variable must be a continuous function on the entire real line, as illustrated in Figure 2.4, for suppose the contrary that there existed a discontinuous "jumping up" point at a point d. Then $P(x = d) = \lim_{b\to d^-} P(b < x \le d) = F(d) - \lim_{b\to d^-} F(b) > 0$ because of the discontinuity (see Figure 2.7), contradicting that $P(x = d) = 0\ \forall d$ if X is continuous.[9]

[8] For those readers whose recollection of the limit concept from calculus courses is not clear, it suffices here to appeal to intuition and interpret the limit of $F(b)$ as "the real number to which $F(b)$ becomes and remains infinitesimally close to as b increases without bound (or as b decreases without bound)." We will examine the limit concept in more detail in Chapter 5.

[9] $\lim_{b\to d^-}$ indicates that we are examining the limit as b approaches d from below (also called a left-hand limit). $\lim_{b\to d^+}$ would indicate the limit as b approached d from above (also called a right-hand limit). For now, it will suffice for the reader to appeal to intuition and interpret $\lim_{b\to d^-} F(b)$ as "the real number to which $F(b)$ becomes and remains infinitesimally close to as b increases and becomes infinitesimally close to d."

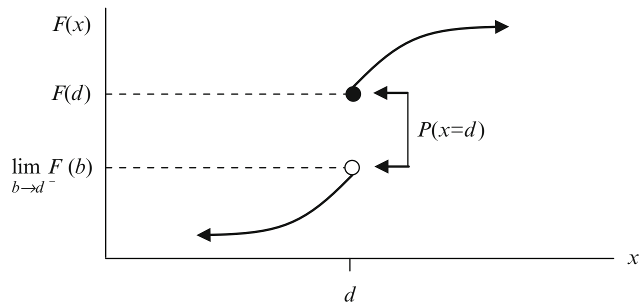


Figure 2.7
Discontinuity in a CDF.

The CDFs for both discrete and mixed discrete-continuous random variables exhibit a countable number of discontinuities at "jumping up" points, representing the assignments of positive probabilities to a countable number of elementary events (recall Figures 2.5 and 2.6). The discrete case is distinguished from the mixed case by the property that the CDF in the former case is a constant function on all intervals for which F is continuous. The mixed case will have a CDF that is an increasing function of x on one or more interval subsets of the real line.[10]

2.3.2 Duality Between CDFs and PDFs

A CDF can be used to derive a PDF as well as discrete and continuous density components in the mixed discrete-continuous random variable case.

Theorem 2.1 ***Discrete PDFs from CDFs***

Let $x_1 < x_2 < x_3 < \ldots$, be the countable collection of outcomes in the range of the discrete random variable X. Then the discrete PDF for X can be defined as

$$f(x_1) = F(x_1),$$
$$f(x_i) = F(x_i) - F(x_{i-1}), \quad i = 2, 3, \ldots,$$
$$f(x) = 0 \text{ for } x \notin R(X).$$

Proof The proof follows directly from the definition of the CDF, and is left to the reader.■

Note, in a large number of empirical applications of discrete random variables, the range of the random variable exhibits an identifiable smallest value, x_1, as in Theorem 2.1. In cases where the range of the random variable does not have a finite smallest value, the Theorem can be restated simply as $f(x_i) = F(x_i) - F(x_{i-1})$, for $x_i > x_{i-1}$ and $f(x) = 0$ for $x \notin R(X)$.

[10] A *strictly* increasing function has $F(x_i) < F(x_j)$ when $X_i < X_j$.

Theorem 2.2 ***Continuous PDFs from CDFs*** *Let $f(x)$ and $F(x)$ represent the PDF and CDF for the continuous random variable X. The density function for X can be defined as $f(x) = dF(x)/dx$ wherever $f(x)$ is continuous, and $f(x) = 0$ (or any nonnegative number) elsewhere.*

Proof By the fundamental theorem of the calculus (recall Lemma 2.1), it follows that

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

wherever $f(x)$ is continuous, so the first part of the theorem is demonstrated. Now, since X is a continuous random variable, then $P(x \leq b) = F(b) = \int_{-\infty}^{b} f(x)dx$ exists $\forall b$ by definition. Changing the value of the nonnegative integrand at points of discontinuity will have no effect on the value of $F(b) = \int_{-\infty}^{b} f(x)dx$,[11] so that *f(x)* can be defined arbitrarily at the points of discontinuity. ■

Theorem 2.3 ***Density Components of a Mixed Discrete-Continuous Random Variable from CDFs*** *Let X be a mixed discrete-continuous random variable with a CDF, F. Let $x_1 < x_2 < x_3 < \ldots$ be the countable collection of outcomes of X for which $F(x)$ is discontinuous. Then the discrete density component of X can be defined as $f_d(x_i) = F(x_i) - \lim_{b \to x_i^-} F(b)$ for $i = 1, 2, 3, \ldots$; and $f_d(x) = 0$* (for any nonnegative numbers) *elsewhere.*

The continuous density component of X can be defined as $f_c(x) = dF(x)/dx$ wherever $f(x)$ is continuous, and $f(x) = 0$ (or any nonnegative number) elsewhere.

Proof The proof is a combination of the arguments used in the proofs of the preceding two theorems and is left to the reader. ■

Given Theorems 2.1–2.3, it follows that there is a complete **duality between CDFs and PDFs** whereby either function can be derived from the other. We illustrate Theorems 2.1–2.3 in the following examples.

Example 2.12 ***Deriving Discrete PDF via Duality*** Recall Example 2.10, where the outcome of rolling a fair die is observed. We can define the discrete density function for X using the CDF for X as follows:

$$f(1) = F(1) = \frac{1}{6}$$

$$f(x) = \begin{cases} F(x) - F(x-1) = \dfrac{x}{6} - \dfrac{x-1}{6} = 1/6 \text{ for } x = 2,3,4,5,6, \\ 0 \quad \text{elsewhere} \end{cases}.$$

[11] This can be rigorously justified by the fact that under the conditions stated: (1) the (improper) Riemann integral is equivalent to a Lebesque integral; (2) the largest set of points for which $f(x)$ can be discontinuous and still have the integral $\int_{-\infty}^{b} f(x)dx$ defined $\forall b$ has "measure zero;" and (3) the values of the integrals are unaffected by changing the values of the integrand on a set of points having "measure zero." This result applies to multivariate integrals as well. See C.W. Burill, 1972, *Measure, Integration, and Probability*, New York: McGraw-Hill, pp. 106–109, for further details.

A more compact representation of $f(x)$ can be given as $f(x) = 1/6\ I_{\{1,2,3,4,5,6\}}(x)$, which we know to be the appropriate discrete density function for the case at hand. □

Example 2.13
Deriving Continuous PDF* via *Duality

Recall Example 2.9, where the time that passes between work-related injuries is observed. We can define the continuous density function for X using the stated CDF for X as follows:

$$f(x) = \begin{cases} \dfrac{dF(x)}{dx} = \dfrac{d\left(1 - e^{-x/100}\right) I_{(0,\infty)}(x)}{dx} = \dfrac{1}{100}\, e^{-x/100} & \text{for } x \in (0,\infty) \\ 0 & \text{for } x \in (-\infty, 0) \end{cases}$$

The derivative of $F(x)$ does not exist at the point $x = 0$ (recall Figure 2.4), which is a reflection of the fact that $f(x)$ is discontinuous at $x = 0$. We arbitrarily assign $f(x) = 0$ when $x = 0$ so that the density function of x is ultimately defined by $f(x) = 1/100\ \mathrm{e}^{-x/100}\ I_{(0,\infty)}(x)$, which we know to be an appropriate continuous density function for the case at hand. □

Example 2.14
Deriving Mixed Discrete-Continuous PDF* via *Duality

Recall Example 2.11, where the operating lives of notebook color screens are observed. The CDF of the mixed discrete-continuous random variable X is discontinuous only at the point $x = 0$ (recall Figure 2.6). Then the discrete density component of X is given by
$f_d(0) = F(0) - \lim_{b \to 0^-} F(b) = .25 - 0 = .25$ and $f_d(x) = 0$, $x \neq 0$, or alternatively,
$f_d(x) = .25 I_{\{0\}}(x)$,

which we know to be the appropriate discrete density function component in this case.

The continuous density function component can be defined as

$$f_c = \begin{cases} \dfrac{dF(x)}{dx} = .75e^{-x} & \text{for } x \in (0,\infty), \\ 0 & \text{for } x \in (-\infty, 0), \end{cases}$$

but the derivative of $F(x)$ does not exist at the point $x = 0$ (recall Figure 2.6). We arbitrarily assign $f_c(x) = 0$ when $x = 0$, so that the continuous density function component of X is finally representable as $f_c(x) = .75\, e^{-x}\, I_{(0,\infty)}(x)$, which we know to be an appropriate continuous density function component in this case. □

2.4 Multivariate Random Variables, PDFs, and CDFs

In the preceding sections of this chapter, we have examined the concept of a univariate random variable, where only one real-valued function was defined on the elements of a sample space. The concept of a multivariate random variable is an extension of the univariate case, where two or more real-valued functions are concurrently defined on the elements of a given sample space. Underlying the concept of a multivariate random variable is the notion of a **real-valued vector function**, which we define now.

Definition 2.13
Real-Valued Vector Function

Let $g_i : A \to \mathbb{R}$, $i = 1,\ldots,n$, be a collection of n real-valued functions, where each function is defined on the domain A. Then the function $\mathbf{g} : A \to \mathbb{R}^n$ defined by

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} g_1(w) \\ \cdot \\ \cdot \\ \cdot \\ g_n(w) \end{bmatrix} = \mathbf{g}(w), \text{ for } w \in A,$$

is called an **(n-dimensional) real-valued vector function**. The real-valued functions $g_1,\ldots,g_n$ are called **coordinate functions** of the vector *function* **g**.

Note that the real-valued vector function $\mathbf{g} : A \to \mathbb{R}^n$ is distinguished from the scalar function $g : A \to \mathbb{R}$ by the fact that its range elements are *n-dimensional vectors* of real numbers as opposed to scalar real numbers. The range of the real-valued vector function is given by $R(\mathbf{g}) = \{(y_1,\ldots,y_n) : y_i = g_i(w), i = 1,\ldots,n; w \in A\}$. We now provide a formal definition of the notion of a multivariate random variable.

Definition 2.14
Multivariate (n-variate) Random Variable

Let $\{S, \Upsilon, P\}$ be a probability space. If $\mathbf{X}: S \to \mathbb{R}^n$ is a real-valued vector function having as its domain the elements of S, then $\mathbf{X}$ is called a **multivariate (n-variate) random variable**.

Since the multivariate random variable is defined by

$$\underset{(n\times 1)}{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} X_1(w) \\ X_2(w) \\ \cdot \\ \cdot \\ \cdot \\ X_n(w) \end{bmatrix} = \underset{(n\times 1)}{\mathbf{X}(w)} \text{ for } w \in S,$$

it is admissible to interpret $\mathbf{X}$ as a collection of n univariate random variables, each defined on the same probability space $\{S, \Upsilon, P\}$. The range of the n-variate random variable is given by $R(\mathbf{X}) = \{(x_1,\ldots,x_n) : x_i = X_i(w), i = 1,\ldots,n; w \in S\}$.

The multivariate random variable concept applies to any real world experiment in which more than one characteristic is observed for each outcome of the experiment. For example, upon making an observation concerning a futures trade on the Chicago Mercantile Exchange, one could record the price, quantity, delivery date, and commodity grade associated with the trade. Upon conducting a poll of registered voters, one could record various political preferences and a myriad of sociodemographic data associated with each randomly chosen interviewee. Upon making a sale, a car dealership will record the price, model, year, color, and the selections from the options list that were made by the buyer.

Definitions for the concept of *discrete* and *continuous* multivariate random variables and their associated density functions are as follows:

Definition 2.15
Discrete Multivariate Random Variables and Probability Density Functions

A multivariate random variable is called *discrete* if its range consists of a countable number of elements. The **discrete joint PDF**, f, for a discrete multivariate random variable $\mathbf{X} = (X_1,\ldots,X_n)$ is defined as $f(x_1,\ldots,x_n) \equiv$ {probability of $(x_1,\ldots,x_n)$} if $(x_1,\ldots,x_n) \in R(\mathbf{X})$, $f(x_1,\ldots,x_n) = 0$ otherwise.

Definition 2.16
Continuous Multivariate Random Variables and Probability Density Functions

A multivariate random variable is called *continuous* if its range is uncountably infinite and there exists a nonnegative-valued function $f(x_1,\ldots,x_n)$, defined for all $(x_1,\ldots,x_n)\in\mathbb{R}^n$, such that $P(A) = \int_{(x_1,\ldots,x_n)\in A} f(x_1,\ldots,x_n)\, dx_1 \ldots dx_n$ for any event $A \subset R(\mathbf{X})$, and $f(x_1,\ldots,x_n) = 0 \quad \forall\, (x_1,\ldots,x_n) \notin R(\mathbf{X})$. The function $f(x_1,\ldots,x_n)$ is called a **continuous joint PDF**.

2.4.1 Multivariate Random Variable Properties and Classes of PDFs

A number of properties of discrete and continuous multivariate random variables, and their joint probability densities, can be identified through analogy with the univariate case. In particular, the multivariate random variable induces a new probability space, $\{R(\mathbf{X}), \Upsilon_{\mathbf{X}}, P_{\mathbf{X}}\}$, for the experiment. The rationale underlying the transition from the probability space $\{S, \Upsilon, P\}$ to the induced probability space $\{R(\mathbf{X}), \Upsilon_{\mathbf{X}}, P_{\mathbf{X}}\}$ is precisely the same as in the univariate case, except for the increased dimensionality of the elements in $R(\mathbf{X})$ in the multivariate case. The probability set function defined on the events in the event space is represented in terms of multiple summation of a PDF in the discrete case, and multiple integration of a PDF in the continuous case. In the discrete case, $f(x_1,\ldots,x_n)$ is directly interpretable as the probability of the outcome $(x_1,\ldots,x_n)$; in the continuous case the probability of each elementary event is zero and $f(x_1,\ldots,x_n)$ is not interpretable as a probability. As a matter of mathematical convenience, both density functions are defined to have the entire n-dimensional real space for their domains, so that $f(x_1,\ldots,x_n) = 0 \;\forall\, \mathbf{x} \notin R(\mathbf{X})$.

Regarding the classes of functions that can be used as discrete or continuous joint density functions, we provide the following generalization of Definition 2.5:

Definition 2.17
The Classes of Discrete and Continuous Joint Probability Density Functions

a. Class of discrete joint density functions. A function $f : \mathbb{R}^n \to \mathbb{R}$ is a member of the class of discrete joint density functions *iff*:

1. the set $C = \{(x_1,\ldots,x_n): f(x_1,\ldots,x_n) > 0,\ (x_1,\ldots,x_n) \in \mathbb{R}^n\}$ is countable;
2. $f(x_1,\ldots,x_n) = 0$ for $\mathbf{x} \in \overline{C}$; and
3. $\sum_{(x_1,\ldots,x_n)\in C} f(x_1,\ldots,x_n) = 1$.

b. Class of continuous joint density functions. A function $f : \mathbb{R}^n \to \mathbb{R}$ is a member of the class of continuous joint density functions *iff*:

1. $f(x_1,\ldots,x_n) \geq 0 \;\forall (x_1,\ldots,x_n) \in \mathbb{R}^n$; and
2. $\int_{\mathbf{x}\in\mathbb{R}^n} f(x_1,\ldots,x_n)\,dx_1\ldots dx_n = 1$.

The reader can generalize the arguments used in the univariate case to demonstrate that the properties stated in Definition 2.17 are sufficient, as well as necessary in the discrete case and "almost necessary" in the continuous case, for set functions defined as

$$P(A) = \begin{cases} \sum_{(x_1,\ldots,x_n)\in A} f(x_1,\ldots,x_n) & \text{(discrete case)}, \\ \int_{(x_1,\ldots,x_n)\in A} f(x_1,\ldots,x_n)dx_1\ldots dx_n & \text{(continuous case)} \end{cases}$$

to satisfy the probability axioms $\forall A \in \Upsilon_{\mathbf{X}}$.

Similar to the univariate case, we define the support of a multivariate random variable, and the equivalence of the range and support as follows.

Definition 2.18
Support of a Multivariate Random Variable

The set $\{\mathbf{x} : f(\mathbf{x})>0, \mathbf{x} \in \mathbb{R}^n\}$ is called the support of the $n \times 1$ random variable $\mathbf{X}$.

Definition 2.19
Support and Range Equivalence of Multivariate Random Variables

$R(\mathbf{X}) \equiv \{\mathbf{x} : f(\mathbf{x})>0 \text{ for } \mathbf{x} \in \mathbb{R}^n\}$

The following is an example of the specification of the probability space for a bivariate discrete random variable.

Example 2.15
Probability Space for a Bivariate Discrete RV

For the experiment of rolling a pair of dice in Example 2.2, distinguish the two die by letting the first die be "red" and the second "green." Thus an outcome (i,j) refers to i dots on the red die and j dots on the green die. Define the following two random variables: $x_1 = X_1(w) = i$, and $x_2 = X_2(w) = i + j$

The range of the bivariate random variable (X_1, X_2) is given by $R(\mathbf{X}) = \{(x_1, x_2)$: $x_1 = i$, $x_2 = i + j$, i and $j \in \{1,\ldots,6\}\}$. The event space is $\Upsilon_{\mathbf{X}} = \{A\text{: } A \subset R(\mathbf{X})\}$. The correspondence between elementary events in $R(\mathbf{X})$ and elementary events in S is displayed as follows:

	x_1						
		1	2	3	4	5	6
x_2	2	(1,1)					
	3	(1,2)	(2,1)				
	4	(1,3)	(2,2)	(3,1)			
	5	(1,4)	(2,3)	(3,2)	(4,1)		
	6	(1,5)	(2,4)	(3,3)	(4,2)	(5,1)	
	7	(1,6)	(2,5)	(3,4)	(4,3)	(5,2)	(6,1)
	8		(2,6)	(3,5)	(4,4)	(5,3)	(6,2)
	9			(3,6)	(4,5)	(5,4)	(6,3)
	10				(4,6)	(5,5)	(6,4)
	11					(5,6)	(6,5)
	12						(6,6)

} Elementary events in S

It follows immediately from the correspondence with the probability space $\{S, \Upsilon, P\}$ that the discrete density function for the bivariate random variable (X_1, X_2) can be represented as

$$f(x_1, x_2) = \frac{1}{36} I_{\{1,\ldots,6\}}(x_1)\, I_{\{1,\ldots,6\}}(x_2 - x_1),$$

and the probability set function defined on the events in $R(\mathbf{X})$ is then

$$P(A) = \sum_{(x_1,x_2)\in A} f(x_1, x_2) \text{ for } A \in \Upsilon_{\mathbf{X}}.$$

Let $A = \{(x_1,x_2): 1 \le x_1 \le 2,\ 2 \le x_2 \le 5,\ (x_1, x_2) \in R(\mathbf{X})\}$, which is the event of rolling 2 or less on the red die and a total of 5 or less on the pair of dice. Then the probability of this event is given by

$$P(A) = \sum_{(x_1,x_2)\in A} f(x_1, x_2) = \sum_{x_1=1}^{2} \sum_{x_2=x_1+1}^{5} f(x_1, x_2) = \frac{7}{36}. \qquad \square$$

The preceding example illustrates two general characteristics of the multivariate random variable concept that should be noted. First of all, even though a multivariate random variable can be viewed as a collection of univariate random variables, it is *not* necessarily the case that the range of the multivariate random variable $\mathbf{X}$ equals the Cartesian product of the ranges of the univariate random variable defining $\mathbf{X}$. Depending on the definition of the $\mathbf{X}_i$'s, either $R(\mathbf{X}) \neq \times_{i=1}^{n} R(X_i)$ and $R(\mathbf{X}) \subset \times_{i=1}^{n} R(X_i)$, or $R(\mathbf{X}) = \times_{i=1}^{n} R(X_i)$ is possible. Example 2.15 is an example of the former case, where a number of scalar outcomes that are *individually* possible for the univariate random variables X_1 and X_2 are not *simultaneously* possible as outcomes for the bivariate random variable (X_1, X_2). Secondly, note that our convention of defining $f(x_1, x_2) = 0\ \forall\ (x_1, x_2) \notin R(\mathbf{X})$ allows an alternative summation expression for defining the probability of event A in Example 2.15:

$$P(A) = \sum_{x_1=1}^{2} \sum_{x_2=2}^{5} f(x_1, x_2) = \frac{7}{36}.$$

We have included the point (2,2) in the summation above, which is an impossible event – we cannot roll a 2 on the red die and a total of 2 on the pair of dice, so that $(2,2) \notin R(\mathbf{X})$. Nonetheless, the probability assigned to A is correct since $f(2,2) = 0$ by definition. In general, when defining the probability of an event A for an n-dimensional discrete random variable $\mathbf{X}$, $f(x_1,\ldots,x_n)$ can be summed over the points identified in the set-defining conditions for A *without* regard for the condition that $\mathbf{x} \in R(\mathbf{X})$, since any $\mathbf{x} \notin R(\mathbf{X})$ will be such that $f(x_1,\ldots,x_n) = 0$, and the value of the summation will be left unaltered. This approach can be especially convenient if set A is defined by individual, independent set-defining conditions applied to each X_i in an n-dimensional random variable $(X_1,\ldots,X_n)$, as in the preceding example. An analogous argument applies to the continuous case, with integration replacing summation.

We now present an example of the specification of the probability space for a bivariate continuous random variable.

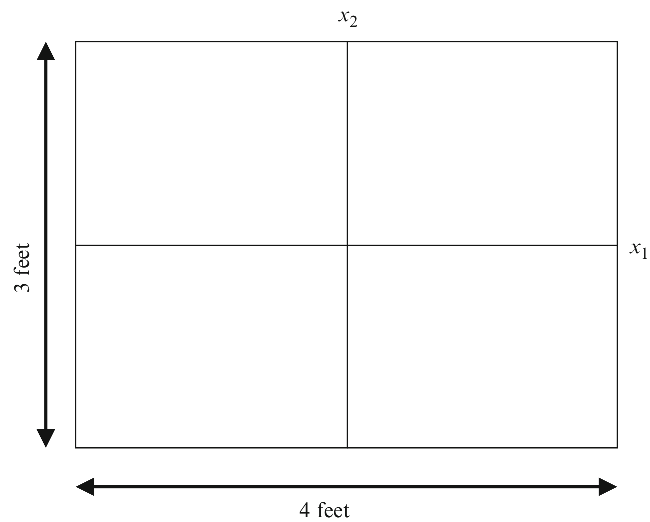


Figure 2.8
Television screen.

Example 2.16
Probability Space for a Bivariate Continuous RV

Your company manufactures big-screen television sets. The screens are 3 ft high by 4 ft wide rectangles that must be coated with a metallic reflective coating (see Figure 2.8). The machine that is coating the screens begins to randomly produce a coating flaw at a point on the screen surface, where all points on the screen are equally likely to be the point of the flaw. Letting (0,0) be the center of the screen, we represent the collection of potential flaw points as $R(\boldsymbol{X}) = \{(x_1,x_2): x_1 \in [-2,2], x_2 \in [-1.5, 1.5]\}$ □.

Clearly, the total area of the screen is $3 \cdot 4 = 12$ ft^2, and any closed rectangle on the screen having width W and height H contains the proportion $WH/12$, of the total area of the screen. Since all of the points are equally likely, the probability set function defined on the events in $R(\mathbf{X})$ should assign to each closed rectangle of points a probability equal to $WH/12$ where W and H are, respectively, the width and height of the rectangular event. We thus seek a function $f(x_1, x_2)$ such that

$$\int_c^d \int_a^b f(x_1, x_2)\, dx_1\; dx_2 \equiv \frac{(b-a)(d-c)}{12}$$

$\forall\, a, b, c$, and d such that $-2 \le a \le b \le 2$ and $-1.5 \le c \le d \le 1.5$. Differentiating the iterated integral above, first with respect to d and then with respect to b, yields $f(b,d) = 1/12 \;\; \forall\, b \in [-2,2]$ and $\forall\, d \in [-1.5, 1.5]$.[12] The form of the continuous PDF is then defined by the following:

$$f(x_1, x_2) = 1/12\; I_{[-2,2]}(x_1) I_{[-1.5,1.5]}(x_2).$$

[12]The differentiation is accomplished by applying Lemma 2.1 twice: once to the integral $\int_c^d \left[\int_a^b f(x_1,x_2)\,dx_1\right] dx_2$, differentiating with respect to d to yield $\int_a^b f(x_1, d)\, dx_1$, and then differentiating the latter integral with respect to b to obtain $f(b,d)$. In summary, $(\partial^2/\partial b \partial d) \int_c^d \int_a^b f(x_1,x_2)\, dx_1 dx_2 = f(b,d)$.

The probability set function is thus defined as $P(A) = \int_{(x_1,x_2 \in A)} (1/12) dx_1 dx_2$. Then, for example, the probability that the flaw occurs in the upper left quarter of the screen is given by

$$P(-2 \leq x_1 \leq 0, 0 \leq x_2 \leq 1.5) = \int_0^{1.5} \int_{-2}^{0} \frac{1}{12} \, dx_1 \; dx_2 = \int_0^{1.5} 1/6 \, dx_2 = .25.$$

2.4.2 Multivariate CDFs and Duality with PDFs

The CDF concept can be generalized to the multivariate case as follows:

Definition 2.20 ***Multivariate Cumulative Distribution Function***

The cumulative distribution function of an n-dimensional random variable $\mathbf{X}$ is defined by

$F(b_1, ..., b_n) = P(x_i \leq b_i, i = 1, ..., n) \; \forall (b_1, ..., b_n) \in \mathbb{R}^n.$

The algebraic representation of $F(b_1, \ldots, b_n)$ in the discrete and continuous cases can be given as follows:

a. Discrete *X*: $F(b_1, \ldots, b_n) = \sum_{x_1 \leq b_1} \cdots \sum_{\substack{x_n \leq b_n \\ f(x_1,\ldots,x_n)>0}} f(x_1, \ldots, x_n)$ for $(b_1, \ldots, b_n) \in \mathbb{R}^n$.

b. Continuous *X*: $F(b_1, ..., b_n) = \int_{-\infty}^{b_n} \cdots \int_{-\infty}^{b_1} f(x_1, \ldots, x_n) dx_1, \ldots, dx_n$ for $(b_1, \ldots, b_n) \in \mathbb{R}^n$.

Some general properties of the joint cumulative distribution function include:

1. $\lim_{b_i \to -\infty} F(b_1, \ldots, b_n) = P(\emptyset) = 0, i = 1, \ldots, n;$
2. $\lim_{b_i \to \infty \forall i} F(b_1, \ldots, b_n) = P(R(X)) = 1;$
3. $F(\mathbf{a}) \leq F(\mathbf{b})$ for $\mathbf{a} < \mathbf{b}$, where

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} < \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \mathbf{b}$$

The *vector* inequality above is taken in the usual sense to mean $a_i \leq b_i \; \forall i$, and $a_i < b_i$ for at least one i. The reader should convince herself that these properties follow directly from the definition of the multivariate cumulative distribution function and the probabilities of the events identified by the appropriate event-defining conditions.

Similar to the univariate case, the joint CDF can be used to derive joint discrete and continuous probability densities. For the discrete case, we discuss the result for bivariate random variables only. For multivariate random variables of three dimensions or higher, the large number of terms required in the density-defining procedure makes its use somewhat cumbersome. We state the generalization in a footnote.[13]

Theorem 2.4 ***Discrete Bivariate PDFs from Bivariate CDFs***

Let (X, Y) be a discrete bivariate random variable with joint cumulative distribution function F(x,y), and let $x_1 < x_2 < x_3 \ldots$, *and* $y_1 < y_2 < y_3 < \ldots$, represent the possible outcomes *of X and Y. Then*

a. $f(x_1, y_1) = F(x_1, y_1)$;
b. $f(x_1, y_j) = F(x_1, y_j) - F(x_1, y_{j-1}),\ j \geq 2$;
c. $f(x_i, y_1) = F(x_i, y_1) - F(x_{i-1}, y_1),\ i \geq 2$; and
d. $f(x_i, y_j) = F(x_i, y_j) - F(x_i, y_{j-1}) - F(x_{i-1}, y_j) + F(x_{i-1}, y_{j-1}),\ i$ and $j \geq 2$.

Proof The proof is left to the reader. ■

As we remarked in the univariate case, if the range of the random variable is such that a lowest ordered outcome does not exist, then the definition simplifies to $f(x_i, y_j) = F(x_i, y_j) - F(x_i, y_{j-1}) - F(x_{i-1}, y_j) + F(x_{i-1}, y_{j-1}),\ \forall i$ and j.

Theorem 2.5 ***Continuous Multivariate PDFs from CDFs***

Let $F(x_1,\ldots,x_n)$ *and* $f(x_1,\ldots,x_n)$ *represent the CDF and PDF for the continuous multivariate random variable* $\mathbf{X} = (X_1,\ldots,X_n)$. *The PDF of* $\mathbf{X}$ *can be defined as*

$$f(x_1, \ldots, x_n) = \begin{cases} \dfrac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n} & \text{where } f(\cdot) \text{is continuous} \\ 0 \text{ (or any nonnegative number) elsewhere.} \end{cases}$$

Proof The first part of the definition follows directly from an n-fold application of Lemma 2.1 for differentiating the iterated integral defining the joint CDF. In particular,

$$\frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n} = \frac{\partial^n \int_{-\infty}^{x_n} \ldots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n)\, dt_1 \ldots dt_n}{\partial x_1 \ldots \partial x_n} = f(x_1, \ldots, x_n)$$

wherever $f(\cdot)$ is continuous.

Regarding the second part of the definition, as long as the integral exists, arbitrarily changing the values of the nonnegative integrand at the points of discontinuity will not affect the value of $F(b_1, \ldots, b_n) = \int_{-\infty}^{b_n} \ldots \int_{-\infty}^{b_1} f(x_1, \ldots, x_n)$ $dx_1, \ldots, dx_n$ (recall footnote 11). ■

Example 2.17 ***Piecewise Definition of Discrete Bivariate CDF***

Examine the experiment of tossing two fair coins independently and observing whether heads (H) or tails (T) occurs on each toss, so that $S = \{(H,H), (H,T), (T,H), (T,T)\}$ with all elementary events in S being equally likely. Define a bivariate

[13] In the discrete m-dimensional case, the PDF can be defined as $f(\mathbf{x}) = F(\mathbf{x}) + \lim_{\delta \to 0_+} \left(\sum_{i=1}^{m} (-1)^i \sum_{\mathbf{v} \in S_i} F(\mathbf{x} - \delta \mathbf{v}) \right)$ where S_i is the set of all of the different $(m \times 1)$ vectors that can be constructed using i 1's and m-i 0's.

random variable on the elements of S by letting x represent the total number of heads and y represent the total number of tails resulting from the two tosses. The joint density function for the bivariate random variable (X,Y) is then defined by $f(x,y) = 1/4\ I_{\{(0,2),\ (2,0)\}}(x,y) + 1/2\ I_{\{(1,1)\}}(x,y)$.

It follows from Definition 2.14 that the joint CDF for (X,Y) can be represented as

$$\begin{aligned} F(b_1, b_2) = {} & \tfrac{1}{4} I_{[2,\infty)}(b_1)\, I_{(-\infty,1)}(b_2) + \tfrac{1}{4} I_{(-\infty,1)}(b_1)\, I_{[2,\infty)}(b_2) \\ & + \tfrac{1}{2} I_{[1,2)}(b_1)\, I_{[1,2)}(b_2) + \tfrac{3}{4} I_{[2,\infty)}(b_1)\, I_{[1,2)}(b_2) \\ & + \tfrac{3}{4} I_{[1,2)}(b_1)\, I_{[2,\infty)}(b_2) + I_{[2,\infty)}(b_1)\, I_{[2,\infty)}(b_2). \end{aligned}$$

The CDF no doubt appears to be somewhat "pieced together", making the definition of F a rather complicated expression. Unfortunately, such piecewise functional definitions often arise when specifying joint CDFs in the discrete case, even for seemingly simple experiments such as the one at hand. To understand more clearly the underlying rationale for the preceding definition of F, it is useful to partition $\mathbb{R}^2$ into subsets that correspond to the events in S. In particular, we are interested in defining the collection of elements $w \in S$ for which $X(w) \leq b_1$ and $Y(w) \leq b_2$ is true for the various values of $(b_1,b_2) \in \mathbb{R}^2$. Examine the following table:

b_1	b_2	$A = \{w: X(w) \leq b_1,\ Y(w) \leq b_2,\ w \in S\}$	$P(A)$
$b_1 < 1$	$b_2 < 1$	$\emptyset$	0
$1 \leq b_1 < 2$	$b_2 < 1$	$\emptyset$	0
$b_1 < 1$	$1 \leq b_2 < 2$	$\emptyset$	0
$b_1 \geq 2$	$b_2 < 1$	$\{(H,H)\}$	1/4
$b_1 < 1$	$b_2 \geq 2$	$\{(T,T)\}$	1/4
$1 \leq b_1 < 2$	$1 \leq b_2 < 2$	$\{(H,T),\ (T,H)\}$	1/2
$b_1 \geq 2$	$1 \leq b_2 < 2$	$\{(H,T),\ (T,H),\ (H,H)\}$	3/4
$1 \leq b_1 < 2$	$b_2 \geq 2$	$\{(H,T),\ (T,H),\ (T,T)\}$	3/4
$b_1 \geq 2$	$b_2 \geq 2$	S	1

The reader should convince herself using a graphical representation of $\mathbb{R}^2$ that the conditions defined on (b_1, b_2) can be used to define nine disjoint subsets of $\mathbb{R}^2$ that exhaustively partition $\mathbb{R}^2$ (i.e., the union of the disjoint sets $= \mathbb{R}^2$). The reader will notice that the indicator functions used in the definition of F were based on the latter six sets of conditions on (b_1, b_2) exhibited in the preceding table. If one were interested in the probability $P(x \leq 1,\ y \leq 1) = F(1,1)$, for example, the joint CDF indicates that 1/2 is the number we seek. □

Example 2.18
Piecewise Definition of Continuous Bivariate CDF

Reexamine the projection television screen example, Example 2.16. The joint CDF for the bivariate random variable (X_1, X_2), whose outcome represents the location of the flaw point, is given by

$$
\begin{aligned}
F(b_1,b_2) &= \int_{-\infty}^{b_2}\int_{-\infty}^{b_1}\frac{1}{12}\,I_{[-2,2]}(x_1)\,I_{[-1.5,1.5]}(x_2)\,dx_1\;dx_2\\
&= \frac{(b_1+2)(b_2+1.5)}{12}\,I_{[-2,2]}(b_1)\,I_{[-1.5,1.5]}(b_2)\\
&\quad+\frac{4(b_2+1.5)}{12}\,I_{(2,\infty]}(b_1)\,I_{[-1.5,1.5]}(b_2)\\
&\quad+\frac{3(b_1+2)}{12}\,I_{[-2,2]}(b_1)\,I_{(1.5,\infty)}(b_2)\\
&\quad+I_{(2,\infty)}(b_1)\,I_{(1.5,\infty)}(b_2).
\end{aligned}
$$

It is seen that piecewise functional definitions of joint CDFs occur in the continuous case as well. To understand the rationale for the piecewise definition, first note that if $b_1 < -2$ and/or $b_2 < -1.5$, then we are integrating over a set of (x_1, x_2) points $\{(x_1, x_2): x_1 < b_1, x_2 < b_2\}$ for which the integrand has a zero value, resulting in a zero value for the definite integral. Thus, $F(b_1,b_2) = 0$ if $b_1 < -2$ and/or $b_2 < -1.5$. If $b_1 \in[-2,2]$ and $b_2 \in[-1.5, 1.5]$, then taking the effect of the indicator functions into account, the integral defining F can be represented as

$$F(b_1,b_2)=\int_{-1.5}^{b_2}\int_{-2}^{b_1}\frac{1}{12}\,dx_1\;dx_2=\frac{(b_1+2)(b_2+1.5)}{12}$$

which is represented by the first term in the preceding definition of F. If $b_1 > 2$, but $b_2 \in[-1.5, 1.5]$, then since the integrand is zero for all values of $x_1 > 2$, we can represent the integral defining F as

$$F(b_1,b_2)=\int_{-1.5}^{b_2}\int_{-2}^{2}\frac{1}{12}\,dx_1\;dx_2=\frac{4(b_2+1.5)}{12}$$

which is represented by the second term in our definition of F. If $b_2 > 1.5$, but $b_1 \in[-2,2]$, then since the integrand is zero for all values of $x_2 > 1.5$, we have that

$$F(b_1,b_2)=\int_{-1.5}^{1.5}\int_{-2}^{b_1}\frac{1}{12}\,dx_1\;dx_2=\frac{3(b_1+2)}{12}$$

which is represented by the third term in our definition of F. Finally, if both $b_1 > 2$ and $b_2 > 1.5$, then since the integrand is zero for all values of $x_1 > 2$ and/or $x_2 > 1.5$, the integral defining F can be written as

$$F(b_1,b_2)=\int_{-1.5}^{1.5}\int_{-2}^{2}\frac{1}{12}\,dx_1\;dx_2=1$$

which justifies the final term in our definition of F. The reader should convince herself that the preceding conditions on (b_1, b_2) collectively exhaust the possible values of $(b_1, b_2) \in \mathbb{R}^2$.

If one were interested in the probability $P(x_1 \leq 1, x_2 \leq 1)$, the "relevant piece" in the definition of F would be the first term, and thus $F(1,1)=\frac{(3)(2.5)}{12}=.625$. Alternatively, the probability $P(x_1 \leq 1, x_2 \leq 10)$ would be assigned using the third term in the definition of F, yielding $F(1,10) = .75$. □

2.4.3 Multivariate Mixed Discrete-Continuous and Composite Random Variables

A discussion of multivariate random variables in the mixed discrete-continuous case could be presented here. However, we choose not to do so. In fact, we will not examine the mixed case any further in this text. We are content with having introduced the mixed case in the univariate context. The problem is that in the multivariate case, representations of the relevant probability set functions – especially when dealing with the concepts of marginal and conditional densities which will be discussed subsequently – become extremely tedious and cumbersome unless one allows a more general notion of integration than that of Riemann, which would then require us to venture beyond the intended scope of this text. We thus leave further study of mixed discrete-continuous random variables to a more advanced course. Note, however, that since elements of both the discrete and continuous random variable concepts are involved in the mixed case, our continued study of the discrete and continuous cases will provide the necessary foundation on which to base further study of the mixed case.

As a final remark concerning our general discussion of multivariate random variables, note that a function (or vector function) of a multivariate random variable is also a random variable (or multivariate random variable). This follows from the same composition of functions argument that was noted in the univariate case. That is, $\mathbf{y} = \mathbf{Y}(\mathbf{X}(w))$, or $\mathbf{y} = \mathbf{Y}(X_1(w), \ldots, X_n(w))$, or

$$\underset{m\times 1}{\mathbf{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} Y_1\left(X_1\left(w\right), \ldots, X_n\left(w\right)\right) \\ \vdots \\ Y_m\left(X_1\left(w\right), \ldots, X_n\left(w\right)\right) \end{bmatrix} = \underset{m\times 1}{\mathbf{Y}(\mathbf{X}(w))}$$

are all in the context of "functions of functions," so that ultimately $\mathbf{Y}$ is a function of the elements $w \in S$, and is therefore a random variable.[14] One might refer to these as **composite random variables**.

2.5 Marginal Probability Density Functions and CDFs

Suppose that we have knowledge of the probability space corresponding to an experiment involving outcomes of the n-dimensional random variable $\mathbf{X}_{(n)} = (X_1,\ldots,X_m, X_{m+1},\ldots,X_n)$, but our real interest lies in assigning probabilities to events involving only the m-dimensional random variable $\mathbf{X}_{(m)} = (X_1,\ldots,X_m)$, $m < n$. In practical terms, this relates to an experiment in which n different

[14]The reader is reminded that we are suppressing the technical requirement that for every Borel set of $\mathbf{y}$ values, the associated collection of w values in S must constitute an *event* in S for the function $\mathbf{Y}$ to be called a random variable. As we have remarked previously, this technical difficulty does not cause a problem in applied work.

characteristics were recorded for each outcome but we are specifically interested in analyzing only a subset of the characteristics. We will now examine the concept of a **marginal probability density function (MPDF)** for $\mathbf{X}_{(m)}$, which will be derived from knowledge of the joint density function for $\mathbf{X}_{(n)}$. Once defined, the MPDF can be used to identify the appropriate probability space only for the portion of the experiment characterized by the outcomes of $(X_1,\ldots,X_m)$, and we will be able to use the MPDF in the usual way (summation in the discrete case, integration in the continuous case) to assign probabilities to events concerning $(X_1,\ldots,X_m)$.

The key to understanding the definition of a marginal probability density is to establish the equivalence between events of the form $(x_1,\ldots,x_m) \in B$ in the probability space for $(X_1,\ldots,X_m)$ and events of the form $(x_1,\ldots,x_n) \in A$ in the probability space for $(X_1,\ldots,X_n)$ since it is the latter events to which we can assign probabilities knowing $f(x_1,\ldots,x_n)$.

2.5.1 Bivariate Case

Let $f(x_1, x_2)$ be the joint density function and $R(\mathbf{X})$ be the range of the bivariate random variable (X_1, X_2). Suppose we want to assign probability to the event $x_1 \in B$. Which event for the *bivariate* random variable is equivalent to event B occurring for the *univariate* random variable X_1? By definition, this event is given by $A = \{(x_1, x_2)\colon x_1 \in B, (x_1, x_2) \in R(\mathbf{X})\}$, i.e., the event B occurs for x_1 *iff* the outcome of (X_1, X_2) is in A so that $x_1 \in B$. Then since B and A are *equivalent events*, the probability that we will observe $x_1 \in B$ is identically equal to the probability that we will observe $(x_1, x_2) \in A$ (recall the discussion of equivalent events in Section 2.2).

For the discrete case, the foregoing probability correspondence implies that

$$P_{X_1}(B) = P(x_1 \in B) = P(A) = \sum_{(x_1,x_2)\in A} f(x_1, x_2).$$

Our convention of defining $f(x_1, x_2) = 0\ \forall (x_1, x_2) \notin R(\mathbf{X})$ allows the following alternative representation of $P_{X_1}(B)$:

$$P_{X_1}(B) = \sum_{x_1\in B}\ \sum_{x_2\in R(X_2)} f(x_1, x_2)$$

The equivalence of the two representations of $P_{X_1}(B)$ follows from the fact that the set of elementary events being summed over the latter case, $C = \{(x_1,x_2)\colon x_1 \in B, x_2 \in R(X_2)\}$ is such that $A \subset C$, and $f(x_1,x_2) = 0\ \forall (x_1,x_2) \in C-A$. The latter representation of $P_{X_1}(B)$ leads to the following definition of the **marginal probability density** of X_1:

$$f_1(x_1) = \sum_{x_2\in R(X_2)} f(x_1, x_2).$$

This function, when summed over the points comprising the event $x_1 \in B$, yields the probability that $x_1 \in B$, i.e.,

$$P_{X_1}(B) = \sum_{x_1 \in B} f_1(x_1) = \sum_{x_1 \in B} \sum_{x_2 \in R(X_2)} f(x_1, x_2).$$

Heuristically, one can think of the marginal density of X_1 as having been defined by "summing out" the values of x_2 in the bivariate PDF for (X_1, X_2). Having defined $f_1(x_1)$, the probability space for the portion of the experiment involving only X_1, can then be defined as $\{R(X_1), \Upsilon_{X_1}, P_{X_1}\}$ where $P_{X_1}(B) = \sum_{X_1 \in B} f_1(x_1)$ for $B \in \Upsilon_{X_1}$. Note that the order in which the random variables are originally listed is immaterial to the approach taken above, and the marginal density function and probability space for X_2 could be defined in an analogous manner by simply reversing the roles of X_1 and X_2 in the preceding arguments. The MPDF for X_2 would be defined as

$$f_2(x_2) = \sum_{x_1 \in R(X_1)} f(x_1, x_2),$$

with the probability space for X_2 defined accordingly.

Example 2.19
Marginal PDFs in a Discrete Bivariate Case

Reexamine Example 1.16, in which an individual was to be drawn randomly from the work force of the Excelsior Corporation to receive a monthly "loyalty award." Define the bivariate random variable (X_1,X_2) as

$$x_1 = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} \text{if} \begin{Bmatrix} \text{male} \\ \text{female} \end{Bmatrix} \text{is drawn,}$$

$$x_2 = \begin{Bmatrix} 0 \\ 1 \\ 2 \end{Bmatrix} \text{if} \begin{Bmatrix} \text{sales} \\ \text{clerical} \\ \text{production} \end{Bmatrix} \text{worker is drawn,}$$

so that the bivariate random variable is measuring two characteristics of the outcome of the experiment: gender and type of worker. The joint density of the bivariate random variable is represented in tabular form below, where the nonzero values of $f(x_1,x_2)$ are given in the cells formed by intersecting an x_1-row with a x_2-column.

		$R(X_2)$			
		0	1	2	$f_1(x_1)$
$R(X_1)$	0	.165	.135	.150	.450
	1	.335	.165	.050	.550
	$f_2(x_2)$	.500	.300	.200	

The nonzero values of the marginal density of X_2 are given in the bottom *margin* of the table, being the definition of the marginal density

$$f_2(x_2) = \sum_{x_1 \in R(X_1)} f(x_1, x_2) = \sum_{x_1=0}^{1} f(x_1, x_2) = .5I_{\{0\}}(x_2) + .3I_{\{1\}}(x_2) + .2I_{\{2\}}(x_2).$$

The probability space for X_2 is thus $\{R(X_2),\Upsilon_{X_2},P_{X_2}\}$ with $\Upsilon_{X_2} = \{A\colon A \subset R(X_2)\}$ and $P_{X_2}(A) = \sum_{x_2 \in A} f_2(x_2)$. If one were interested in the probability that the individual chosen was a sales or clerical worker, i.e., the event $A = \{0,1\}$, then $P_{X_2}(A) = \sum_{x_2=0}^{1} f_{x_2}(x_2) = .5 + .3 = .8$.

The nonzero values of the marginal density for X_1 are given in the right-hand *margin* of the table, the definition of the density being

$$f_1(x_1) = \sum_{x_2 \in R(X_2)} f(x_1, x_2) = \sum_{x_2=0}^{2} f(x_1, x_2) = .45 I_{(0)}(x_1) + .55 I_{(1)}(x_1).$$

The probability space for X_1 is thus $\{R(X_1), \Upsilon_{X_1}, P_{X_1}\}$ with $\Upsilon_{X_1} = \{A\colon A \subset R(X_1)\}$ and $P_{X_1}(A) = \sum_{x_1 \in A} f_1(x_1)$. If one were interested in the probability that the individual chosen was male, i.e., the event $A = \{0\}$, then $P_{X_1}(A) = \sum_{x_1=0}^{0} f_{x_1}(x_1) = .45$.□

The preceding example provides a heuristic justification for the term *marginal* in the bivariate case and reflects the historical basis for the name *marginal density function*. In particular, by summing across the rows or columns of a tabular representation of the joint PDF $f(x_1,x_2)$ one can calculate the marginal densities of X_1 and X_2 in the *margins* of the table.

We now examine the marginal density function concept for continuous random variables. Recall that the probability of event B occurring for the univariate random variable X_1 is **identical** to the probability that the event $A = \{(x_1,x_2){:}x_1 \in B,\ (x_1,x_2) \in R(X)\}$ occurs for the bivariate random variable $\mathbf{X} = (X_1,X_2)$. Then

$$P_{X_1}(B) = P(x_1 \in B) = P(A) = \int_{(x_1,x_2)\in A} f(x_1,x_2)dx_1dx_2.$$

Our convention of defining $f(x_1, x_2) = 0\ \forall (x_1, x_2) \notin R(\mathbf{X})$ allows an alternative representation of $P_{X_1}(B)$ to be given by

$$P_{X_1}(B) = \int_{x_1 \in B} \int_{-\infty}^{\infty} f(x_1,x_2)dx_2dx_1.$$

The equivalence of the two representations follows from the fact that the set of elementary events being integrated over in the latter case, $C = \{(x_1, x_2){:}\ x_1 \in B,\ x_2 \in (-\infty,\infty)\}$, is such that $A \subset C$, and $f(x_1, x_2) = 0\ \forall\ (x_1, x_2) \in C{-}A$. The latter representation of $P_{X_1}(B)$ leads to the definition of the marginal density of X_1 as

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1,x_2)dx_2.$$

This function, when integrated over the elementary events comprising the event $x_1 \in B$, yields the probability that $x_1 \in B$, i.e.,

$$P_{X_1}(B) = \int_{x_1 \in B} f_1(x_1)\,dx_1 = \int_{x_1 \in B} \int_{-\infty}^{\infty} f(x_1,x_2)dx_2dx_1.$$

Heuristically, one might think of the marginal density of X_1 as having been defined by "integrating out" the values of X_2 in the bivariate density function for (X_1, X_2).

Having defined $f_1(x_1)$, the probability space for the portion of the experiment involving only X_1 can then be defined as $\{R(X_1), \Upsilon_{X_1}, P_{X_1}\}$ where $P_{X_1}(A) = \int_{x_1 \in A} f_1(x_1)dx_1$ for $A \in \Upsilon_{X_1}$. Since the order in which the random variables were originally listed is immaterial, the marginal density function and probability space for X_2 can be defined in an analogous manner by simply reversing the roles of X_1 and X_2 in the preceding arguments. The MPDF for X_2 would be defined as

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1$$

with the probability space for X_2 defined accordingly.

Example 2.20 ***Marginal PDFs in a Continuous Bivariate Case***

The Seafresh Fish Processing Company operates two fish processing plants. The proportion of processing capacity at which each of the plants operates on any given day is the outcome of a bivariate random variable having joint density function $f(x_1, x_2) = (x_1 + x_2)\, I_{[0,1]}(x_1)\, I_{[0,1]}(x_2)$. The marginal density function for the proportion of processing capacity at which plant 1 operates can be defined by integrating out x_2 from $f(x_1, x_2)$ as

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 = \int_{-\infty}^{\infty} (x_1 + x_2)\, I_{[0,1]}(x_1)\, I_{[0,1]}(x_2)\, dx_2 \\ &= \int_0^1 (x_1 + x_2)\, I_{[0,1]}(x_1)\, dx_2 = \left(x_1\ x_2 + \frac{x_2^2}{2}\right) I_{[0,1]}(x_1) \Bigg|_0^1 \\ &= (x_1 + 1/2)\, I_{[0,1]}(x_1). \end{aligned}$$

The probability space for plant 1 outcomes is given by $\{R(X_1), \Upsilon_{X_1}, P_{X_1}\}$, where $R(X_1) = [0,1]$, $\Upsilon_{X_1} = \{A\text{: } A \text{ is a Borel set} \subset R(X_1)\}$, and $P_{X_1}(A) = \int_{x_1 \in A} f_1(x_1)dx_1$, $\forall A \in \Upsilon_{X_1}$. If one were interested in the probability that plant 1 will operate at less than half of capacity on a given day, i.e., the event $A = [0, .5)$, then

$$P_{X_1}(x_1 \leq .5) = \int_0^{.5} \left(x_1 + \frac{1}{2}\right) I_{[0,1]}(x_1)dx_1 = \frac{x_1^2}{2} + \frac{x_1}{2} \Bigg|_0^{.5} = .375. \qquad \square$$

Regarding other properties of marginal density functions, note that the significance of the term *marginal* is only to indicate the context in which the density was derived, i.e., the marginal density of X_1 is deduced from the joint density for (X_1, X_2). Otherwise, the MPDF has no special properties that differ from the basic properties of any other PDF.

2.5.2 *n*-Variate Case

The concept of a discrete MPDF can be straightforwardly generalized to the n-variate case, in which case the marginal densities may themselves be joint density functions. For example, if we have the density function $f(x_1, x_2, x_3)$ for the trivariate random variable (X_1, X_2, X_3), then we may conceive of six marginal density functions: $f_1(x_1), f_2(x_2), f_3(x_3)$, $f_{12}(x_1, x_2)$, $f_{13}(x_1, x_3)$, $f_{23}(x_2, x_3)$. In general, for an n-variate random variable, there are $(2^n - 2)$ possible MPDFs that can be defined from knowledge of $f(x_1,\ldots,x_n)$. We present the n-variate generalization in the following definition. We use the notation $f_{j_1 \ldots j_m}(x_{j_1}, \ldots, x_{j_m})$ to represent the MPDF of the m-variate random variable $(X_{j_1}, \ldots, X_{j_m})$ with the j_i's being the indices that identify the particular random vector of interest. The motivation for the definition is analogous to the argument in the bivariate case upon identifying the equivalent events $(x_{j_1}, \ldots, x_{j_m}) \in B$ and $A = \{\mathbf{x} : (x_{j_1}, \ldots, x_{j_m}) \in B, \mathbf{x} \in R(\mathbf{X})\}$ is left to the reader as an exercise.

Definition 2.21 ***Discrete Marginal Probability Density Functions***

Let $f(x_1,\ldots,x_n)$ be the joint discrete PDF for the n-dimensional random variable $(X_1,\ldots,X_n)$. Let $J = \{j_1, j_2, \ldots, j_m\}$, $1 \le m < n$, be a set of indices selected from the index set $I = \{1, 2, \ldots, n\}$. Then the marginal density function for the m-dimensional discrete random variable $(X_{j_1},\ldots,X_{j_m})$ is given by

$$f_{j_1 \ldots j_m}(x_{j_1}, \ldots, x_{j_m}) = \underset{(x_i \in R(X_i),\ i \in I-J)}{\sum \cdots \sum} f(x_1, \ldots, x_n).$$

In other words, to define a MPDF in the general discrete case, we simply "sum out" the variables that are not of interest in the joint density function. We are left with the marginal density function for the random variable in which we are interested. For example, if $n = 3$, so that $I = \{1,2,3\}$, and if $J = \{j_1, j_2\} = \{1,3\}$ so that I-$J = \{2\}$, then Definition 2.21 indicates that the MPDF of the random variable (X_1, X_3) is given by

$$f_{13}(x_1, x_3) = \sum_{x_2 \in R(X_2)} f(x_1, x_2, x_3).$$

Similarly, the marginal density for x_1 would be defined by

$$f_1(x_1) = \sum_{x_2 \in R(X_2)} \sum_{x_3 \in R(X_3)} f(x_1, x_2, x_3).$$

The concept of a continuous MPDF can be generalized to the n-variate case as follows:

Definition 2.22 ***Continuous Marginal Probability Density Functions***

Let $f(x_1,\ldots,x_n)$ be the joint continuous PDF for the n-variate random variable $(X_1,\ldots,X_n)$. Let $J = \{j_1, j_2, \ldots, j_m\}$, $1 \le m < n$, be a set of indices selected from the index set $I = \{1, 2, \ldots, n\}$. Then the marginal density function for the m-variate continuous random variable $(X_{j1},\ldots, X_{jm})$ is given by

$$f_{j_1 \ldots j_m}(x_{j_1}, \ldots, x_{j_m}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) \prod_{i \in I-J} dx_i.$$

In other words, to define a MPDF function in the general continuous case, we simply "integrate out" the variables in the joint density function that are not of interest. We are left with the marginal density function for the random variables in which we are interested. An example of marginal densities in the context of a trivariate random variable will be presented in Section 2.8.

2.5.3 Marginal Cumulative Distribution Functions (MCDFs)

Marginal CDFs are simply CDFs that have been derived for a subset of the random variables in $\mathbf{X} = (X_1,\ldots,X_n)$ from initial knowledge of the joint PDF or joint CDF of $\mathbf{X}$. For example, ordering the elements of a continuous random variable $(X_1,\ldots,X_n)$ so that the first $m < n$ random variables are of primary interest, the MCDF of $(X_1,\ldots,X_m)$ can be defined as

$$\begin{aligned} F_{1\ldots m}(b_1,\ldots,b_m) &= P_{X_1\ldots X_m}(x_i \le b_i, i=1,\ldots,m)(\text{Def.}\,\textit{of CDF}) \\ &= P(x_i \le b_i, i=1,\ldots,m; x_i < \infty, i=m+1,\ldots,n)(\textit{equivalent events}) \\ &= F(b_1,\ldots,b_m,\infty,\ldots,\infty)(\textit{Def. in terms of joint CDF}) \\ &= \int_{-\infty}^{b_1}\cdots\int_{-\infty}^{b_m}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\ldots,x_n)\,dx_n\ldots dx_1\;(\textit{Def. in terms of joint PDF}) \\ &= \int_{-\infty}^{b_1}\cdots\int_{-\infty}^{b_m} f_{1\ldots m}(x_1,\ldots,x_m)\,dx_m\ldots dx_1\;(\textit{Def. in terms of marginal PDF}). \end{aligned}$$

In the case of an arbitrary subset $(X_{j_1},\ldots,X_{j_m})$, $m < n$ of the random variables $(X_1,\ldots,X_n)$, the MCDF in terms of the joint CDF or marginal PDF can be represented as

$$F_{j_1\ldots j_m}(b_{j_1},\ldots,b_{j_m}) = F(\mathbf{b}) = \int_{-\infty}^{b_{j_1}}\cdots\int_{-\infty}^{b_{j_m}} f_{j_1\ldots j_m}(x_{j_1},\ldots,x_{j_m})\,dx_{j_m}\ldots dx_{j_1}$$

where b_{j_i} is the j_ith entry in $\mathbf{b}$ and $b_i = \infty$ if $i \notin \{j_1,\ldots,j_m\}$.

Examples of marginal CDFs in the trivariate case are presented in Section 2.8. The discrete case is analogous, with summation replacing integration.

2.6 Conditional Density Functions

Suppose that we have knowledge of the probability space corresponding to an experiment involving outcomes of the n-dimensional random variable $\mathbf{X}_{(n)} = (X_1,\ldots,X_m, X_{m+1},\ldots,X_n)$, and we are interested in assigning probabilities to the event $(x_1,\ldots,x_m) \in C$ *given that* $(x_{m+1},\ldots,x_n) \in D$. In practical terms, this relates to an experiment in which n different characteristics were recorded for each outcome and we are specifically interested in analyzing a subset of these characteristics *given that* a fixed set of possibilities will occur with certainty for the remaining characteristics. Note that this is different from asking for the *probability* of observing the event $(x_1,\ldots,x_m) \in$ C *and* $(x_{m+1},\ldots,x_n) \in D$, for we are saying that $(x_{m+1},\ldots,x_n) \in D$ *will happen with certainty.* How do we assign the appropriate probability in this case? Questions of this type can be addressed through the use of **conditional PDFs**, which can be derived from knowledge of the joint density function $f(x_1,\ldots,x_n)$.

The key to the definition of a conditional PDF is to establish the equivalence between events for the m-dimensional random variable $(X_1,\ldots,X_m)$ and

$(n-m)$-dimensional random variable $(X_{m+1},\ldots,X_n)$ with events for the n-dimensional random variable $(X_1,\ldots,X_n)$. Then conditional probabilities in the probability space for $(X_1,\ldots,X_n)$ can be used to define a conditional PDF.

2.6.1 Bivariate Case

Let $f(x_1, x_2)$ be the joint density function and $R(\mathbf{X})$ be the range of the bivariate random variable (X_1, X_2). The event for the bivariate random variable that is equivalent to the event $x_1 \in C$ occurring for the scalar random variable X_1 is given by $A = \{(x_1, x_2): x_1 \in C, (x_1, x_2) \in R(X)\}$. That is, A is the set of all possible outcomes for the two-tuple (x_1, x_2) that result in the first coordinate x_1 residing in C. Similarly, the event for the bivariate random variable that is equivalent to the event D occurring for the random variable X_2 is given by $B = \{(x_1, x_2): x_2 \in D, (x_1, x_2) \in R(\mathbf{X})\}$. Then the probability that $x_1 \in C$ *given that* $x_2 \in D$ can be defined by the conditional probability

$$P_{X_1|X_2}(C|D) = P(x_1 \in C|x_2 \in D) = P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ for } P(B) \neq 0,$$

where $A \cap B = \{(x_1, x_2): x_1 \in C, x_2 \in D, (x_1, x_2) \in R(\mathbf{X})\}$.

In the case of a discrete random variable, the foregoing conditional probability is represented by

$$P_{X_1|X_2}(C|D) = P(A|B) = \frac{\sum\limits_{(x_1,x_2)\in A\cap B} f(x_1,x_2)}{\sum\limits_{(x_1,x_2)\in B} f(x_1,x_2)}$$

Given our convention that $f(x_1, x_2) = 0$ whenever $(x_1, x_2) \notin R(\mathbf{X})$, we can ignore the set-defining condition $(x_1, x_2) \in R(\mathbf{X})$ in both the sets $A \cap B$ and B, and represent the conditional probability as

$$P_{X_1|X_2}(C|D) = \frac{\sum_{x_1\in C}\sum_{x_2\in D} f(x_1,x_2)}{\sum_{x_1\in R(X_1)}\sum_{x_2\in D} f(x_1,x_2)} = \sum_{x_1\in C}\left[\frac{\sum_{x_2\in D} f(x_1,x_2)}{\sum_{x_2\in D} f_2(x_2)}\right]$$

where we have used the fact that $f_2(x_2) = \sum_{x_1\in R(X_1)} f(x_1,x_2)$. The expression in brackets is the conditional density function we seek, since it is the function that would be summed over the elements in C to assign probability to the event $x_1 \in C$, given $x_2 \in D$, for any event C. We will denote the conditional density of X_1, given $x_2 \in D$, by the notation $f(x_1|x_2 \in D)$. If D is a singleton set $\{d\}$, we will also represent the conditional density function as $f(x_1|x_2 = d)$.

In the case of a continuous bivariate random variable, the probability that $x_1 \in C$ given that $x_2 \in D$ would be given by (assuming $P_{X_2}(D) = P(B) \neq 0$)

$$P_{X_1|X_2}(C|D) = P(x_1 \in C|\, x_2 \in D) = P(A|B) = \frac{\int_{(x_1,x_2)\in A\cap B} f(x_1,x_2)\, dx_1\ dx_2}{\int_{(x_1,x_2)\in B} f(x_1,x_2)\, dx_1\ dx_2}$$

Using our convention that $f(x_1,x_2) = 0\ \forall\ (x_1, x_2) \notin R(\mathbf{X})$, we can also represent the conditional probability as

$$P_{X_1|X_2}(C|D) = \frac{\int_{x_1 \in C} \int_{x_2 \in D} f(x_1, x_2)\, dx_2\, dx_1}{\int_{-\infty}^{\infty} \int_{x_2 \in D} f(x_1, x_2)\, dx_2\, dx_1} = \int_{x_1 \in C} \left[\frac{\int_{x_2 \in D} f(x_1, x_2)\, dx_2}{\int_{x_2 \in D} f_2(x_2)\, dx_2} \right] dx_1,$$

where we have used the fact that $f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$. The expression in brackets is the conditional density function we seek, since it is the function that would be integrated over the elements in C to assign probability to the event $x_1 \in C$, given $x_2 \in D$, for any event C. As in the discrete case, we will use the notation $f(x_1|x_2 \in D)$ or $f(x_1|x_2 = d)$ to represent the conditional density function. In both the discrete and continuous cases, we will eliminate the random variable subscripts on $P_{X_1|X_2}(\cdot)$ when the random variable context of the probability set function is clear.

Once derived, a conditional PDF exhibits all of the standard properties of a PDF. The significance of the term *conditional* PDF is to indicate that the density of X_1 was derived from the joint density for (X_1, X_2) conditional on a specific event for X_2. Otherwise, there are no special general properties of a conditional PDF that distinguishes it from any other PDF.

We provide examples of the derivation and use of discrete and continuous conditional PDF's in the following examples.

Example 2.21 ***Conditional PDF in a Bivariate Discrete Case***

Recall the dice example, Example 2.15, where $f(x_1,x_2) = (1/36)\, I_{\{1,\ldots,6\}}(x_1)\, I_{\{1,\ldots,6\}}(x_2 - x_1)$. The conditional density function for X_1, given that $x_2 = 5$, is given by

$$f(x_1 \,|\, x_2 = 5) = \frac{f(x_1, 5)}{f_2(5)} = \frac{\frac{1}{36} I_{\{1,\ldots,6\}}(x_1)\, I_{\{1,\ldots,6\}}(5 - x_1)}{\frac{6-|5-7|}{36} I_{\{2,\ldots,12\}}(5)} = \frac{1}{4}\, I_{\{1,\ldots,4\}}(x_1).$$

The probability of rolling a 3 or less on the red die, given that the total of the two dice will be 5, is then

$$P(x_1 \leq 3 | x_2 = 5) = \sum_{x_1=1}^{3} f(x_1 | x_2 = 5) = \frac{3}{4}.$$

Note that the *unconditional* probability that $x_1 \leq 3$ is equal to 1/2.
The conditional density function for X_1, given that $x_2 \in D = \{7, 11\}$, is given by

$$f(x_1 \,|\, x_2 \in D) = \frac{\sum_{x_2 \in D} f(x_1, x_2)}{\sum_{x_2 \in D} f_2(x_2)} = \frac{\frac{1}{36} I_{\{1,\ldots,6\}}(x_1) \left[I_{\{1,\ldots,6\}}(7 - x_1) + I_{\{1,\ldots,6\}}(11 - x_1) \right]}{\frac{8}{36}}$$
$$= \tfrac{1}{8} I_{\{1,\ldots,4\}}(x_1) + \tfrac{1}{4} I_{\{5,6\}}(x_1)$$

The probability of rolling a 3 or less on the red die, given that the total of the two dice will be either a 7 or 11, is then

$$P(x_1 \leq 3 | x_2 \in D) = \sum_{x_1=1}^{3} f(x_1 | x_2 \in D) = \frac{3}{8}.$$

□

Example 2.22 ***Conditional PDF in a Bivariate Continuous Case***

Recall Example 2.20 regarding the proportion of daily capacity at which two processing plants operate. The conditional density function of plant 1's capacity, given that plant 2 operates at less than half of capacity, is given by

$$f(x_1 \mid x_2 \le .5) = \frac{\int_{-\infty}^{.5} f(x_1, x_2)\, dx_2}{\int_{-\infty}^{.5} f_2(x_2)\, dx_2} = \frac{\int_0^{.5} (x_1 + x_2)\, I_{[0,1]}(x_1)\, dx_2}{\int_0^{.5} (x_2 + 1/2)\, dx_2}$$
$$= \frac{.5\, x_1 + .125}{.375}\, I_{[0,1]}(x_1) = \left(\frac{4}{3}\, x_1 + \frac{1}{3}\right) I_{[0,1]}(x_1)$$

The probability that $x_1 \le .5$, given that $x_2 \le .5$, is given by

$$P(x_1 \le .5 \mid x_2 \le .5) = \int_0^{.5} \left(\frac{4}{3} x_1 + \frac{1}{3}\right) dx_1 = \frac{1}{3}.$$

Recall that the *unconditional* probability that $x_1 \le .5$ was .375. □

2.6.2 Conditioning on Elementary Events in Continuous Cases-Bivariate

A problem arises in the continuous case when defining a conditional PDF for X_1, conditional on an *elementary* event occurring for X_2. Namely, because all elementary events are assigned probability zero in the continuous case, with the integral over a singleton set being zero, our definition of the conditional density, as presented earlier, yields

$$f(x_1 \mid x_2 = b) = \frac{\int_b^b f(x_1, x_2)\, dx_2}{\int_b^b f_2(x_2)\, dx_2} = \frac{0}{0},$$

which is an indeterminate form. Thus $f(x_1|x_2 = b)$ is undefined so that $P(x_1 \in A \mid x_2 = b)$ is undefined as well. This is different than the discrete case, where

$$f(x_1 \mid x_2 = b) = \frac{f(x_1, b)}{f_2(b)}$$

is well-defined, provided $f_2(b) \ne 0$.

The problem is circumvented by redefining the conditional probability, $P(x_1 \in A|x_2 = b)$, in the continuous case in terms of a limit as

$$P(x_1 \in A \mid x_2 = b) = \lim_{\varepsilon \to 0^+} P(x_1 \in A \mid x_2 \in [b - \varepsilon, b + \varepsilon])$$
$$= \lim_{\varepsilon \to 0^+} \left[\frac{\int_{x_1 \in A} \int_{b-\varepsilon}^{b+\varepsilon} f(x_1, x_2)\, dx_2\ dx_1}{\int_{b-\varepsilon}^{b+\varepsilon} f_2(x_2) dx_2}\right]$$

where $\lim_{\varepsilon \to 0^+}$ means we are examining a limit for a sequence of ε-values that approach zero from positive values (i.e., $\varepsilon > 0$). The idea is to examine the limiting value of a sequence of probabilities that are conditioned on a corresponding sequence of events, $[b - \varepsilon, b + \varepsilon]$ for $\varepsilon \to 0^+$, that converge to the elementary event $\{b\}$. The following lemma will facilitate the identification of the limit.

Lemma 2.2
Mean Value Theorem for Integrals

If $g(x)$ is continuous $\forall x \in [c_1, c_2]$, then $\exists x_0 \in [c_1, c_2]$ such that $\int_{c_1}^{c_2} g(x)dx = g(x_0)(c_2 - c_1)$.[15]

To use the mean value theorem, and to ensure that the limit of the conditional probabilities exists, we assume that there exists a choice of $\varepsilon > 0$ such that $f_2(x_2)$ and $f(x_1, x_2)$ are continuous in x_2, $\forall x_2 \in [b - \varepsilon, b + \varepsilon]$, and that $f_2(b) > 0$. Then, by the mean value theorem,

$$P(x_1 \in A | x_2 = b) = \lim_{\varepsilon \to 0^+} \left[\frac{\int_{x_1 \in A} \int_{b-\varepsilon}^{b+\varepsilon} f(x_1, x_2)\, dx_2 dx_1}{\int_{b-\varepsilon}^{b+\varepsilon} f_2(x_2) dx_2} \right] = \lim_{\varepsilon \to 0^+} \left[\frac{2\varepsilon \int_{x_1 \in A} f(x_1, x_2^0)\, dx_1}{2\varepsilon f_2\,(x_2^*)} \right]$$

where both x_2^0 and $x_2^* \in [b - \varepsilon, b + \varepsilon]$, and x_2^0 will generally depend on the value of x_1.[16] The 2 ε's in the numerator and denominator cancel each other, and as $\varepsilon \to 0^+$, the interval $[b - \varepsilon, b + \varepsilon]$ reduces to $[b,b] = b$, so that in the limit, both x_2^0 and $x_2^* = b$. The limiting value of the conditional probability is then

$$P(x_1 \in A | x_2 = b) = \int_{x_1 \in A} \frac{f(x_1, b)}{f_2(b)} dx.$$

Since the choice of event A is arbitrary, it follows that the appropriate conditional probability density in this case is

$$f(x_1 \,|\, x_2 = b) = \frac{f(x_1, b)}{f_2\,(b)},$$

which is precisely of the same form as the discrete case. Thus, the definition of conditional density functions, when conditioning on elementary events, will be identical for continuous and discrete random variables, provided $f_2(b) \neq 0$.

Example 2.23
PDF Conditioned on an Elementary Event

Recall Example 2.22. The conditional PDF for plant 1's proportion of capacity X_1, given that plant 2's capacity proportion is $x_2 = .75$, can be defined as

$$f(x_1 \,|\, x_2 = .75) = \frac{f(x_1, .75)}{f_2\,(.75)} = \frac{(x_1 + .75) I_{[0,1]}(x_1)}{1.25} = \left(\frac{4}{5} x_1 + \frac{3}{5}\right) I_{[0,1]}(x_1)$$

The probability that $x_1 \leq .5$, given that $x_2 = .75$, is then given by $P(x_1 \leq .5 \mid x_2 = .75) = \int_0^{.5} \left(\frac{4}{5} x_1 + \frac{3}{5}\right) dx_1 = .4$ □

[15]R. Courant and F. John, *Introduction to Calculus and Analysis*, New York, John Wiley-Interscience, 1965, p. 143.

[16]In applying the mean value theorem to the numerator, we treat $f(x_1, x_2)$ as a function of the single variable x_2, fixing the value of x_1 for each application.

2.6.3 Conditioning on Elementary Events in Continuous Cases: *n*-Variate Case

The preceding concepts of discrete and continuous conditional PDFs in the bivariate case can be generalized to the n-variate case, as indicated in the following definition, which subsumes $n = 2$ as a special case:

Definition 2.23
Conditional Probability Density Functions

Let $f(x_1,\ldots,x_n)$ be the joint density function for the n-dimensional random variable $(X_1,\ldots,X_n)$. The conditional density function for the m-dimensional random variable$(X_1, \ldots, X_m)$, given that $(X_{m+1}, \ldots, X_n) \in D$ and $P_{X_{m+1},\ldots,X_n}(D)>0$, is as follows:

***Discrete Case*:**

$$f(x_1, \ldots, x_m | x_{m+1}, \ldots, x_n) \in D) = \frac{\sum_{(x_{m+1},\ldots,x_n)\in D} f(x_1, \ldots, x_n)}{\sum_{(x_{m+1},\ldots,x_n)\in D} f_{m+1,\ldots,n}(x_{m+1}, \ldots, x_n)}$$

***Continuous Case*:**

$$f(x_1, \ldots, x_m | x_{m+1}, \ldots, x_n) \in D) = \frac{\int_{(x_{m+1},\ldots,x_n)\in D} f(x_1, \ldots, x_n) dx_{m+1} \ldots dx_n}{\int_{(x_{m+1},\ldots,x_n)\in D} f_{m+1,\ldots,n}(x_{m+1}, \ldots, x_n) dx_{m+1} \ldots dx_n}$$

If D is equal to the elementary event $(d_{m+1}, \ldots, d_n)$ then the definition of the conditional density in both the discrete and continuous cases can be represented as

$$f(x_1, \ldots, x_m | x_i = d_i, i = m+1, \ldots, n) = \frac{f(x_1, \ldots, x_m, d_{m+1}, \ldots, d_n)}{f_{m+1,\ldots,n}\,(d_{m+1}, \ldots, d_n)}$$

when the marginal density in the denominator is positive valued.[17]

For example, if $n = 3$, then the conditional density function of (X_1, X_2), given that $x_3 \in D$, would be defined as

$$f(x_1, x_2 | x_3 \in D) = \frac{\sum_{x_3\in D} f(x_1, x_2, x_3)}{\sum_{x_3\in D} f_3(x_3)}$$

in the discrete case, with integration replacing summation in the continuous case. If $D = d_3$, then for both the discrete and continuous cases,

$$f(x_1, x_2 | x_3 = d_3) = \frac{f(x_1, x_2, d_3)}{f_3\,(d_3)}.$$

[17]In the continuous case, it is also presumed that f and $f_{m+1,\ldots,n}$ are continuous in $(x_{m+1}, \ldots, x_n)$ within some neighborhood of points around the point where the conditional density is evaluated in order to justify the conditional density definition via a limiting argument analogous to the bivariate case. Motivation for the conditional density expression when conditioning on an elementary event in the continuous case can then be provided by extending the mean-value theorem argument used in the bivariate case. See R.G. Bartle, *Real Analysis*, p. 429 for a statement of the general mean value theorem for integrals.

An example of conditional PDFs in the trivariate case will be presented in Section 2.8.

In summary, if we begin with the joint density function appropriate for assigning probabilities to events involving the n-dimensional random variable $(X_1,\ldots,X_n)$, we can derive a conditional probability density function that is the PDF appropriate for assigning probabilities to events for an m-dimensional subset of the random variables in $(X_1,\ldots,X_n)$, *given* (or *conditional*) on an event for the remaining $n-m$ random variables. The construction of the conditional density involves both the joint density of $(X_1,\ldots, X_n)$ and the marginal density of the $(n-m)$ dimensional random variable on which we are conditioning. In the special case where we are conditioning on an elementary event, the conditional density function simply becomes the ratio of the joint density function divided by the marginal density function, replacing the arguments of these functions with their conditioned values for those arguments corresponding to random variables on which we are conditioning (which represents *all* of the arguments of the marginal density, and a subset of the arguments of the joint density).

2.6.4 Conditional CDFs

We can define the concept of a **conditional CDF** by simply using a conditional density function in the definition of the CDF. For example, for the bivariate random variable (X_1, X_2), we can define

$$F(b_1 \mid x_2 \in D) = P(x_1 \leq b_1 \mid x_2 \in D) = \int_{-\infty}^{b_1} f(x_1 \mid x_2 \in D)\, dx_1$$

as one such conditional CDF, representing the CDF of X_1, conditional on $x_2 \in D$. Once defined, the conditional CDF possesses no special properties that distinguish it in concept from any other CDF. The reader is asked to contemplate the various conditional CDFs that can be defined for the n-dimensional random variable $(X_1,\ldots,X_n)$.

2.7 Independence of Random Variables

From our previous discussion of independence of events, we know that A and B are independent *iff* $P(A \cap B) = P(A)P(B)$. This concept can be applied directly to determine whether two particular events for the n-dimensional random variable $(X_1,\ldots,X_n)$ are independent. The general definition of independence of events (Definition 1.19) can also be applied to examine the independence of k specific events for the random variable $(X_1,\ldots,X_n)$.

The concept of independence of events will now be extended further to the idea of **independence of random variables**, which is related to the question of whether the n events (recall the abbreviated set definition notation of Definition 2.7) $\{x_i \in A_i\} \equiv \{(x_1,\ldots,x_n): x_i \in A_i, (x_1,\ldots,x_n) \in R(\mathbf{X})\}$, $i = 1,\ldots,n$, are independent for *all* possible choices of the events $A_1,\ldots,A_n$. If so, the *n random variables* are said to be independent. In effect, the concept is one of **global independence of**

events for random variables – we define an event A_i for each of the n random variables in $(X_1,\ldots,X_n)$ and, *no matter how we define the events* (which is the meaning of the term "global" here), the events $\{x_i \in A_i\}$, $i = 1,\ldots,n$, are independent. Among other things, we will see that this implies that the probability assigned to *any* event A_i for *any* random variable X_i in $(X_1,\ldots,X_n)$ is unaffected by conditioning on *any* event B for the remaining random variables (assuming $P(B) > 0$ for the existence of the conditional probability).

2.7.1 Bivariate Case

We seek to establish a condition that will ensure that the events $\{x_1 \in A_1\}$ and $\{x_2 \in A_2\}$ are independent for *all* possible choices of the events A_1 and A_2. This can be accomplished by applying independence conditions to events in the probability space, $\{R(\mathbf{X}),\Upsilon,P\}$ for the bivariate random variable $\mathbf{X} = (X_1, X_2)$. The events $x_1 \in A_1$ and $x_2 \in A_2$ are equivalent, respectively, to the following events for the bivariate random variable:

$$B_1 = \{(x_1,x_2) : x_1 \in A_1, (x_1,x_2) \in R(\mathbf{X})\} \text{and } B_2 = \{(x_1,x_2) : x_2 \in A_2, (x_1,x_2) \in R(\mathbf{X})\}.$$

The two events B_1 and B_2 are independent *iff* $P(B_1 \cap B_2) = P(B_1)P(B_2)$, which can also be represented using our abbreviated notation as $P(x_1 \in A_1, x_2 \in A_2) = P(x_1 \in A_1)P(x_2 \in A_2)$. Requiring the independence condition to hold for *all* choices of the events A_1 and A_2 leads to the definition of the independence condition for random variables.

Definition 2.24 ***Independence of Random Variables: Bivariate***

The random variables X_1 and X_2 are said to be independent *iff* $P(x_1 \in A_1, x_2 \in A_2) = P(x_1 \in A_1)\, P(x_2 \in A_2)$ for all events A_1, A_2.

There is an equivalent characterization of independence of random variables in terms of PDFs that can be useful in practice and that also further facilitates the investigation of the implications of random variable independence.

Theorem 2.6 ***Bivariate Density Factorization for Independence of Random Variables***

The random variables X_1 and X_2 with joint PDF $f(x_1, x_2)$ and marginal PDFs $f_i(x_i)$, $i = 1, 2$, are independent iff the joint density factors into the product of the marginal densities as $f(x_1, x_2) = f_1(x_1)\, f_2(x_2)\ \forall\ (x_1, x_2)$.[18]

[18]Technically, the factorization need not hold at points of discontinuity for the joint density function of a continuous random variable. However, if the random variables are independent, there will always exist a density function for which the factorization can be formed. This has to do with the fundamental non-uniqueness of PDFs in the continuous case, which can be redefined arbitrarily at a countable number of isolated points without affecting the assignment of any probabilities of events through integration. There are few practical benefits of this non-uniqueness, and we suppress this technical anomaly here.

Proof **Discrete Case** Let A_1 and A_2 be any two events for X_1 and X_2, respectively. Then if the joint density function $f(x_1, x_2)$ factors,

$$P(x_1 \in A_1, x_2 \in A_2) = \sum_{x_1 \in A_1} \sum_{x_2 \in A_2} f(x_1, x_2) = \sum_{x_1 \in A_1} f_1(x_1) \sum_{x_2 \in A_2} f_2(x_2) = P(x_1 \in A_1) P(x_2 \in A_2)$$

so that X_1 and X_2 are independent. Thus, factorization is sufficient for independence. Now assume (X_1, X_2) are independent random variables. Let $A_1 = \{a_1\}$ and $A_2 = \{a_2\}$ for any choice of elementary events, $a_i \in R(X_i)$, corresponding to the random variable $X_i, i = 1,2$, respectively. Then, by independence,

$$P(x_1 = a_1, x_2 = a_2) = f(a_1, a_2) = P(x_1 = a_1) P(x_2 = a_2) = f_1(a_1) f_2(a_2)$$

If $a_i \notin R(X_i)$, then $f_i(a_i) = 0$ and $f(a_1, a_2) = 0$ for $i = 1,2$, and thus factorization will automatically hold. Thus, factorization is necessary for independence.
Continuous case Let A_1 and A_2 be any two events for X_1 and X_2, respectively. Then if the joint density function $f(x_1, x_2)$ factors,[19]

$$\begin{aligned} P(x_1 \in A_1, x_2 \in A_2) &= \int_{x_2 \in A_2} \int_{x_1 \in A_1} f(x_1, x_2) dx_1 dx_2 \\ &= \int_{x_1 \in A_1} f_1(x_1) dx_1 \int_{x_2 \in A_2} f_2(x_2) dx_2 \\ &= P(x_1 \in A_1) P(x_2 \in A_2), \end{aligned}$$

so that X_1 and X_2 are independent. Thus, factorization is sufficient for independence. Now assume (X_1, X_2) are independent random variables. Let $A_i = \{x_i : x_i \leq a_i\}$ for arbitrary choice of $a_i, i = 1,2$. Then by independence,

$$\begin{aligned} P(x_1 \leq a_1, x_2 \leq a_2) &= \int_{-\infty}^{a_2} \int_{-\infty}^{a_1} f(x_1, x_2)\, dx_1 \; dx_2 \\ &= P(x_1 \leq a_1) P(x_2 \leq a_2) = \int_{-\infty}^{a_1} f_1(x_1)\, dx_1 \int_{-\infty}^{a_2} f_2(x_2)\, dx_2 . \end{aligned}$$

Differentiating the integrals with respect to a_1 and a_2 yields $f(a_1, a_2) = f_1(a_1) f_2(a_2)$ wherever the joint density function is continuous. Thus, the factorization condition stated in the theorem is *necessary* for independence. ■

In other words, two random variables are independent *iff* their joint PDF can be expressed equivalently as the product of their respective marginal PDFs (the condition not being required to hold at points of discontinuity in the continuous case). An important implication of independence of X_1 and X_2 is that the

[19] Any points of discontinuity can be ignored in the definitions of the probability integrals without affecting the probability assignments.

conditional and marginal PDFs of the respective random variables are identical.[20] For example, assuming independence,

$$f(x_1 \mid x_2 \in B) = \frac{\int_{x_2 \in B} f(x_1, x_2)\, dx_2}{\int_{x_2 \in B} f_2(x_2)\, dx_2} = \frac{f_1(x_1) \int_{x_2 \in B} f_2(x_2)\, dx_2}{\int_{x_2 \in B} f_2(x_2)\, dx_2} = f_1(x_1)$$

(in the discrete case, replace integration by summation). The fact that conditional and marginal PDFs are identical implies that the probability of $x_1 \in A$, for any event A, is unaffected by the occurrence or nonoccurrence of event B for X_2. For example, in the continuous case,

$$P(x_1 \in A | x_2 \in B) = \int_{x_1 \in A} f(x_1 | x_2 \in B) dx_1 = \int_{x_1 \in A} f_1(x_1) dx_1 = P(x_1 \in A)$$

(replace integration by summation in the discrete case). The result holds for any events involving X_2 for which the conditional density function is defined. The roles of X_1 and X_2 can be reversed in the preceding discussion.

Example 2.24 ***Independence of Bivariate Continuous RVs***

Recall Example 2.16 concerning coating flaws in the manufacture of television screens. The horizontal and vertical coordinates of the coating flaw was the outcome of a bivariate random variable with joint density function

$$f(x_1, x_2) = \tfrac{1}{12} I_{[-2,2]}(x_1) I_{[-1.5,1.5]}(x_2).$$

Are the random variables independent?
Answer: The marginal densities of X_1 and X_2 are given by

$$f_1(x_1) = I_{[-2,2]}(x_1) \int_{-1.5}^{1.5} \frac{1}{12} dx_2 = .25 I_{[-2,2]}(x_1)$$

$$f_2(x_2) = I_{[-1.5,1.5]}(x_2) \int_{-2}^{2} \frac{1}{12} dx_1 = \frac{1}{3} I_{[-1.5,1.5]}(x_2)\,.$$

It follows that $f(x_1, x_2) = f_1(x_1) f_2(x_2)\ \forall (x_1, x_2)$, and the random variables are independent. Therefore, knowledge that an event for X_2 has occurred has no effect on the probability assigned to events for X_1, and vice versa. □

Example 2.25 ***Independence of Bivariate Discrete RVs***

Recall the dice example, Example 2.15. Are X_1 and X_2 independent random variables?

Answer: Examine the validity of the independence condition:

$$f(x_1, x_2) \stackrel{?}{=} f_1(x_1) f_2(x_2) \quad \forall (x_1, x_2),$$

[20] We will henceforth suppress constant reference to the fact that factorization might not hold for some points of discontinuity in the continuous case – it will be tacitly understood that results we derive based on the factorization of $f(x_1, x_2)$ may be violated at some isolated points. For example, for the case at hand, marginal and conditional densities may not be equal at some isolated points. Assignments of probability will be unaffected by this technical anomaly.

or, specifically,

$$\frac{1}{36} I_{\{1,2,\ldots,6\}}(x_1) I_{\{1,2,\ldots,6\}}(x_2 - x_1) \stackrel{?}{=} \frac{1}{6} I_{\{1,2,\ldots,6\}}(x_1) \left(\frac{6 - |x_2 - 7|}{36}\right) I_{\{2,\ldots,12\}}(x_2) \quad \forall (x_1, x_2)$$

The random variables X_1 and X_2 are *not* independent, since, for example, letting $x_1 = 2$ and $x_2 = 4$ results in $1/36 \neq 1/72$. Therefore, knowledge that an event for X_2 has occurred *can* affect the probability assigned to events for X_1, and vice versa. □

2.7.2 *n*-Variate

The independence concept can be extended beyond the bivariate case to the case of **independence of random variables** $X_1,\ldots,X_n$. The formal definition of independence in the n-variate case is as follows:

Definition 2.25 *Independence of Random Variables (n-Variate)*

> The random variables $X_1, X_2,\ldots,X_n$ are said to be independent *iff* $P(\mathrm{x_i} \in A_i, i = 1,\ldots,n) = \prod_{i=1}^{n} P(x_i \in A_i)$ for all choices of the events $A_1,\ldots,A_n$.

The motivation for the definition is similar to the argument used in the bivariate case. For $B_i = \{(x_1, \ldots, x_n) : x_i \in A_i, (x_1, \ldots, x_n) \in R(\mathbf{X})\}, i = 1, \ldots, n$ to be independent events, we require (recall Definition 1.19)

$$P(\bigcap_{j \in J} B_j) = \prod_{j \in J} P(B_j) \; \forall J \subset \{1, 2, \ldots, n\} \text{ with } N(J) \geq 2.$$

If we require this condition to hold for *all* possible choices of the events $(B_1,\ldots,B_n)$, then the totality of the conditions can be represented as

$$P(x_i \in A_i, i = 1, \ldots, n) = P\left(\bigcap_{i=1}^{n} B_i\right) = \prod_{i=1}^{n} P(B_i) = \prod_{i=1}^{n} P(x_i \in A_i)$$

for all choices of the events $A_1,\ldots,A_n$ (or, equivalently, for corresponding choices of $B_1,\ldots,B_n$). Any of the other conditions required for independence of events, i.e.,

$$P\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} P(B_j) \text{ with } J \subset \{1, 2, \ldots, n\} \text{ and } N(J) < n,$$

are implied by the preceding condition upon letting $A_j = R(X_j)$ (or equivalently, $B_j = R(\mathbf{X})$) for $j \in \bar{J}$.

The generalization of the joint density factorization theorem is given as Theorem 2.7. The proof is a direct extension of the arguments used in proving Theorem 2.6, and is left to the reader.

Theorem 2.7
Density Factorization for Independence of Random Variables (n-Variate Case)

The random variables $X_1, X_2,\ldots,X_n$ with joint PDF $f(x_1,\ldots,x_n)$ and marginal PDFs $f_i(x_i)$, $i = 1,\ldots,n$, are independent *iff* the joint density can be factored into the product of the marginal densities as $f(x_1, \ldots, x_n) = \prod_{i=1}^n f_i(x_i)\ \forall (x_1, \ldots, x_n)$.[21]

An example of the application of Theorem 2.7 is given in Section 2.8.

2.7.3 Marginal Densities Do Not Determine an *n*-Variate Density Without Independence

If $(X_1,\ldots,X_n)$ are independent random variables, then knowing the marginal densities $f_i(x_i)$, $i = 1,\ldots,n$ is equivalent to knowing the joint density function for $(X_1,\ldots,X_n)$, since then $f(x_1,\ldots,x_n) = \prod_{i=1}^n f_i(x_i)$. However, if the random variables in the collection $(X_1,\ldots,X_n)$ are *not* independent, then knowing each of the marginal densities of the X_i's is generally *not* sufficient to determine the joint density function for $(X_1,\ldots,X_n)$. In fact, it can be shown that an uncountably infinite family of different joint density functions can give rise to the same collection of marginal density functions.[22] We provide the following counter example in the bivariate case to the proposition that knowledge of the marginal PDFs is sufficient for determining the n-variate PDF.

Example 2.26
Marginal Densities Do Not Imply n-Variate Densities

Examine the function

$$f_\alpha(x_1,x_2) = [1+\alpha(2x_1 - 1)(2x_2 - 1)]I_{[0,1]}(x_1)I_{[0,1]}(x_2).$$

The reader should verify that $f_\alpha(x_1, x_2)$ is a PDF $\forall\ \alpha \in [-1,1]$. For any choice of $\alpha \in [-1,1]$, the marginal density function for X_1 is given by $f_1(x_1) = \int_{-\infty}^{\infty} f_\alpha(x_1, x_2)\, dx_2 = I_{[0,1]}(x_1)$. Similarly, the marginal density of X_2, for any choice of $\alpha \in [-1,1]$, is given by $f_2(x_2) = \int_{-\infty}^{\infty} f_\alpha(x_1, x_2) dx_1 = I_{[0,1]}(x_2)$.

Since the *same* marginal density functions are associated with each of an uncountably infinite collection of bivariate density functions, it is clear that knowledge of $f_1(x_1)$ and $f_2(x_2)$ is insufficient to determine which is the appropriate joint density function for (X_1, X_2). If we knew the marginal densities of X_1 and X_2, as stated, and if X_1 and X_2 are *independent* random variables, then we would know that $f(x_1, x_2) = I_{[0,1]}(x_1)\, I_{[0,1]}(x_2)$. □

[21] The same technical proviso regarding points of discontinuity in the case of continuous random variables hold as in the bivariate case. See Footnote 18.

[22] E.J. Gumbel (1958) *Distributions a' plusieurs variables dont les marges sont données*, *C.R. Acad. Sci., Paris*, 246, pp. 2717–2720.

2.7.4 Independence Between Random Vectors and Between Functions of Random Vectors

The independence concepts can be extended so that they apply to independence among two or more random *vectors*. Essentially, all that is required is to interpret the X_i's as *multivariate* random variables in the appropriate definitions and theorems presented heretofore, and the statements are valid. Motivation for the validity of the extensions can be provided using arguments that are analogous to those used previously. For example, to extend the previous bivariate result to two random *vectors*, let $\mathbf{X}_1 = (X_{11},\ldots,X_{1m})$ be an m-dimensional random variable and $\mathbf{X}_2 = (X_{21},\ldots,X_{2n})$ be an n-dimensional random variable. Then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent *iff*

$$P(\mathbf{x}_1 \in A_1, \mathbf{x}_2 \in A_2) = P((\mathbf{x}_{11},\ldots,x_{1m}) \in A_1, (x_{21},\ldots,x_{2n}) \in A_2)$$
$$= P((x_{11},\ldots,x_{1m}) \in A_1)P((x_{21},\ldots,x_{2n}) \in A_2) = P(\mathbf{x}_1 \in A_1)P(\mathbf{x}_2 \in A_2)$$

for *all* event pairs A_1, A_2. Furthermore, in terms of joint density factorization, $\mathbf{X}_1$ and $\boldsymbol{X}_2$ are independent *iff*

$$f(\mathbf{x}_1, \mathbf{x}_2) = f(x_{11}, \ldots, x_{1m}, x_{21}, \ldots, x_{2n})$$
$$= f_1(x_{11}, \ldots, x_{1m}) f_2(x_{21}, \ldots, x_{2n})$$
$$= f_1(\mathbf{x}_1) f_2(\mathbf{x}_2)\ \forall(\mathbf{x}_1, \mathbf{x}_2)$$

The reader can contemplate the myriad of other independence conditions that can be constructed for discrete and continuous random *vectors*.

Implications of the extended independence definitions and theorems are qualitatively similar to the implications identified previously for the case where the X_i's were interpreted as scalars. For example, if $\mathbf{X}_1 = (X_{11},\ldots,X_{1m})$ and $\mathbf{X}_2 = (X_{21},\ldots,X_{2n})$ are independent random variables, then

$$P((x_{11},\ldots,x_{1m}) \in A_1 | (x_{21},\ldots,x_{2n}) \in A_2)) = P((x_{11},\ldots,x_{1m}) \in A_1),$$

i.e., conditional and unconditional probability of events for the random variable $\mathbf{X}_1$ are identical (and similarly for $\mathbf{X}_2$) for all choices of A_1 and A_2 for which the conditional probability is defined.

It is also useful to note some results concerning the independence of random variables that are defined as functions of other independent random variables. We begin with the simplest case of two independent random variables X_1 and X_2.

Theorem 2.8 *If X_1 and X_2 are independent random variables, and if the random variables Y_1 and Y_2 are defined by $y_1 = Y_1(x_1)$ and $y_2 = Y_2(x_2)$, then Y_1 and Y_2 are independent random variables.*

Proof The event involving outcomes of X_i that is equivalent to the event $y_i \in A_i$ is given by $B_i = \{x_i\colon Y_i(x_i) \in A_i, x_i \in R(X_i)\}$ *for* $i = 1, 2$. Then

$$P(y_1 \in A_1, y_2 \in A_2) = P(x_1 \in B_1, x_2 \in B_2)$$
$$= P(x_1 \in B_1)P(x_2 \in B_2) \quad \text{(by independence of } x_1, x_2)$$
$$= P(y_1 \in A_1)P(y_2 \in A_2),$$

and since this holds for every event pair A_1, A_2, the random variables Y_1 and Y_2 are independent. ■

Example 2.27 ***Independence of Functions of Continuous RVs***

A large service station sells unleaded and premium-grade gasoline. The quantities sold of each type of fuel on a given day is the outcome of a bivariate random variable with density function[23]

$$f(x_1,x_2) = \frac{1}{20}e^{-(.1x_1+.5x_2)}I_{(0,\infty)}(x_1)I_{(0,\infty)}(x_2),$$

where the x_i's are measured in thousands of gallons. The marginal densities are given by (reader, please verify)

$$f_1(x_1) = \frac{1}{10}e^{-.1x_1}\, I_{(0,\infty)}(x_1) and\, f_2(x_2) = \frac{1}{2}e^{-.5x_2}\, I_{(0,\infty)}(x_2)$$

and so the random variables are independent. The prices of unleaded and premium gasoline are \$3.25 and \$3.60 per gallon, respectively. The wholesale cost of gasoline plus federal state and local taxes amounts to \$2.80 and \$3.00 per gallon, respectively. Other daily variable costs in selling the two products amount to $C_i(x_i) = 20\, x_i^2$, $i = 1, 2$. Are daily profits above variable costs for the two products independent random variables?

Answer: Yes. Note that the profit levels in the two cases are $\Pi_1 = 450x_1 - 20x_1^2$ and $\Pi_2 = 600x_2 - 20x_2^2$, respectively. Since Π_1 is only a function of x_1, Π_2 is only a function of x_2, and X_1 and X_2 are independent, then Π_1 and Π_2 are independent by Theorem 2.8. □

A more general theorem explicitly involving random *vectors* is stated as follows:

Theorem 2.9 *Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be a collection of n independent random vectors, and let the random vectors $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ be defined by $\mathbf{y}_i = \mathbf{Y}_i(\mathbf{x}_i)$, $i = 1,\ldots,n$. Then the random vectors $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ are independent.*

Proof The event involving outcomes of the random vector $\mathbf{X}_i$ that is equivalent to the event A_i for the random vector $\mathbf{Y}_i$ is given by $B_i = \{\mathbf{x}_i\colon \mathbf{Y}_i(\mathbf{x}_i) \in A_i,\ \mathbf{x}_i \in R(\mathbf{X}_i)\}$, $i = 1,\ldots,n$. Then

$$\begin{aligned} P(\mathbf{y}_i \in A_i, i=1,\ldots,n) &= P(\mathbf{x}_i \in B_i, i=1,\ldots,n) \\ &= \prod_{i=1}^{n} P(\mathbf{x}_i \in B_i) \text{ (by independence of random vectors)} \\ &= \prod_{i=1}^{n} P(\mathbf{y}_i \in A_i) \end{aligned}$$

and since this holds for every collection of events $A_1,\ldots,A_n$, the random vectors $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ are independent by a vector interpretation of the random variables in Definition 2.19. ■

[23] This must be an approximation – why?

Example 2.28 ***Independence of Functions of Discrete RVs***

Examine the experiment of independently tossing two fair coins and rolling three fair dice. Let X_1 and X_2 represent whether heads ($x_i = 1$) or tails ($x_i = 0$) appears on the first and second coins, respectively, and let X_3, X_4, and X_5 represent the number of dots facing up on each of the three dice, respectively. Since the random variables are independent, the probability density of $X_1,\ldots,X_5$ can be written as

$$f(x_1,\ldots,x_5) = \prod_{i=1}^{2} \frac{1}{2} I_{\{0,1\}}(x_i) \prod_{i=3}^{5} \frac{1}{6} I_{\{1,\ldots,6\}}(x_i)$$

Define two new random *vectors* $\mathbf{Y}_1$ and $\mathbf{Y}_2$ using the vector functions

$$\mathbf{y}_1 = \begin{bmatrix} y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ x_1\ x_2 \end{bmatrix} = \mathbf{Y}_1(x_1, x_2),$$

$$\mathbf{y}_2 = \begin{bmatrix} y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} x_3 + x_4 + x_5 \\ x_3\ x_4 / x_5 \end{bmatrix} = \mathbf{Y}_2(x_3, x_4, x_5)$$

Then since the vector $\mathbf{y}_1$ is a function of (x_1, x_2), $\mathbf{y}_2$ is a function of (x_3, x_4, x_5), and since the random vectors (X_1, X_2) and (X_3,X_4,X_5) are independent (why?), Theorem 2.9 indicates that the random vectors $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent. This is clearly consistent with intuition, since outcomes of the vector $\mathbf{Y}_1$ obviously have nothing to do with outcomes of the vector $\mathbf{Y}_2$. The reader should note that within vectors, the random variables are *not* independent, i.e., Y_{11} and Y_{12} are not independent, and neither are Y_{21} and Y_{22}. □

2.8 Extended Example of Multivariate Concepts in the Continuous Case

We now further illustrate some of the concepts of this chapter with an example involving a trivariate continuous random variable. Let (X_1, X_2, X_3) have the PDF $f(x_1, x_2, x_3) = (3/16)\, x_1\, x_2^2\, e^{-x_3} I_{[0,2]}(x_1)\, I_{[0,2]}(x_2)\, I_{[0,\infty)}(x_3)$.

a. What is the marginal density of X_1? of X_2? of X_3?
Answer:

$$f_1(x_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)dx_2dx_3$$
$$= \tfrac{3}{16}x_1 I_{[0,2]}(x_1)\int_{-\infty}^{\infty} x_2^2 I_{[0,2]}(x_2)dx_2 \int_{-\infty}^{\infty} e^{-x_3} I_{[0,\infty]}(x_3)dx_3$$
$$= \tfrac{3}{16}x_1 I_{[0,2]}(x_1)\left(\tfrac{8}{3}\right)(1) = \tfrac{1}{2}x_1 I_{[0,2]}(x_1).$$

Similarly,

$$f_2(x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)dx_1dx_3 = \frac{3}{8}x_2^2 I_{[0,2]}(x_2)$$

$$f_3(x_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2,x_3)dx_1dx_2 = e^{-x_3} I_{[0,\infty]}(x_3).$$

b. What is the probability that $x_1 \geq 1$?

Answer: $P(x_1 \geq 1) = \int_1^\infty f_1(x_1)dx_1 = \int_1^2 \frac{1}{2}x_1 dx_1 = \left.\frac{x_1^2}{4}\right|_1^2 = .75.$

c. Are the three random variables independent?
Answer: Yes. It is clear that $f(x_1, x_2, x_3) = f_1(x_1)\, f_2(x_2)\, f_3(x_3)\ \forall\ (x_1,x_2,x_3)$.

d. What is the marginal cumulative distribution function for X_1? for X_3?
Answer: By definition,

$$F_1(b) = \int_{-\infty}^{b} f_1(x_1)\, dx_1 = \int_{-\infty}^{b} \frac{1}{2}x_1 I_{[0,2]}(x_1)\, dx_1$$
$$= \frac{1}{2}\left.\frac{x_1^2}{2}\right|_0^b I_{[0,2]}(b) + I_{(2,\infty)}(b) = \frac{b^2}{4} I_{[0,2]}(b) + I_{(2,\infty)}(b),$$
$$F_3(b) = \int_{-\infty}^{b} f_3(x_3)\, dx_3 = \int_{-\infty}^{b} e^{-x_3} I_{[0,\infty)}(x_3)\, dx_3$$
$$= -e^{-x_3}\Big|_0^b I_{[0,\infty)}(b) = (1 - e^{-b})\, I_{[0,\infty)}(b).$$

e. What is the probability that $x_1 \leq 1$? that $x_3 > 1$?
Answer: $P(x_1 \leq 1) = F_1(1) = .25$. $P(x_3 > 1) = 1 - F_3(1) = e^{-1} = .3679$.

f. What is the joint cumulative distribution function for X_1, X_2, X_3?
Answer: By definition:

$$F(b_1, b_2, b_3) = \int_{-\infty}^{b_1}\int_{-\infty}^{b_2}\int_{-\infty}^{b_3} f(x_1, x_2, x_3)\, dx_3\ dx_2\ dx_1$$
$$= \int_{-\infty}^{b_1} \frac{1}{2}x_1\ I_{[0,2]}(x_1)\, dx_1 \int_{-\infty}^{b_2} \frac{3}{8}\int_{-\infty}^{b_2} x_2^2\, I_{[0,2]}(x_2)\, dx_2 \int_{-\infty}^{b_3} e^{-x_3}\, I_{[0,\infty)}(x_3)\, dx_3$$
$$= \left[\frac{b_1^2}{4} I_{[0,2]}(b_1) + I_{(2,\infty)}(b_1)\right]\left[\frac{3b_2^3}{24} I_{[0,2]}(b_2) + I_{(2,\infty)}(b_2)\right]\left[\left(1 - e^{-b_3}\right) I_{[0,\infty)}(b_3)\right]$$

g. What is the probability that $x_1 \leq 1,\ x_2 \leq 1,\ x_3 \leq 10$?
Answer: $F(1,1,10) = (1/4)(3/24)(1 - e^{-10}) = .031$.

h. What is the conditional PDF of X_1, given that $x_2 = 1$ and $x_3 = 0$?
Answer: By definition, $f(x_1|x_2 = 1, x_3 = 0) = \frac{f(x_1,1,0)}{f_{23}(1,0)}$. Also,

$$f_{23}(x_2,x_3) = \int_{-\infty}^{\infty} f(x_1,x_2,x_3)dx_1 = \frac{3}{8}x_2^2 I_{[0,2]}(x_2)e^{-x_3}I_{[0,\infty)}(x_3). \text{ Thus,}$$

$$f(x_1|x_2 = 1, x_3 = 0) = \frac{(\frac{3}{16})\, x_1\ I_{[0,2]}(x_1)}{\frac{3}{8}} = \frac{1}{2}x_1 I_{[0,2]}(x_1)$$

i. What is the probability that $x_1 \in [0, 1/2]$, given that $x_2 = 1$ and $x_3 = 0$?

Answer:

$$P(x_1 \in [0, \tfrac{1}{2}] \,|\, x_2 = 1, x_3 = 0) = \int_0^{1/2} f(x_1 \,|\, x_2 = 1, x_3 = 0)\, dx_1$$
$$= \int_0^{1/2} \frac{1}{2} x_1 \; I_{[0,2]}\,(x_1)\, dx_1 = \frac{x_1^2}{4}\bigg|_0^{1/2} = \frac{1}{16}$$

j. Let the two random variables Y_1 and Y_2 be defined by $y_1 = Y_1(x_1, x_2) = x_1^2 x_2$ and $y_2 = Y_2(x_3) = x_3/2$. Are the random variables Y_1 and Y_2 independent? **Answer**: Yes, they are independent. The bivariate random variable (X_1, X_2) is independent of the random variable X_3 since $f(x_1, x_2, x_3) = f_{12}(x_1, x_2)\, f_3(x_3)$, i.e., the joint density function factors into the product of the marginal density of (X_1, X_2) and the marginal density of X_3. Then, since y_1 is a function of only (x_1, x_2) and y_2 is a function of only x_3, Y_1 and Y_2 are independent random variables, by Theorem 2.9.

Keywords, Phrases, and Symbols

[), interval, closed lower bound and open upper bound
(], interval, open lower bound and closed upper bound
[], interval, closed bounds
(), interval, open bounds
Abbreviated set notation
CDF
Classes of discrete and continuous density functions
Composite random variable
Conditional cumulative distribution function
Conditional density function
Continuous density component
Continuous joint PDF
Continuous PDF
Continuous random variable
Cumulative distribution function
Density factorization for independence
Discrete density component
Discrete joint PDF
Discrete PDF
Discrete random variable
Duality between CDFs and PDFs
$\exists$, there exists
Equivalent events
Event A is relatively certain
Event A is relatively impossible
Event A occurs with probability one
Event A occurs with probability zero
$F(b)$
$f(x_1, \ldots, x_m | (x_{m+1}, \ldots, x_n) \in B)$
$f(x_1, \ldots, x_n)$
$f_{1\ldots m}(x_1, \ldots, x_m)$
Increasing function
Independence of random variables
Induced probability space, $\{R(X), \Upsilon_X, P_X\}$
$\bar{J}$, complement of J
Marginal cumulative distribution function
Marginal PDF
MCDF
Mixed discrete-continuous random variables
MPDF
Multivariate cumulative distribution function
Multivariate random variable
$\Leftrightarrow$, mutual implication or *iff*
Outcome of the random variable, x
$P(x \leq b)$
PDF
$R(X)$
Random variable, X
Real-valued vector function
Truncation function
$X(w)$
$X: S \rightarrow R$

Problems

1. Which of the following are valid PDFs? Justify your answer.

a. $f(x) = (.2)^x\,(.6)^{1-x}\,I_{\{0,1\}}\,(x)$

b. $f(x) = (.3)\,(.7)^x\,I_{\{0,1,2,\ldots\}}\,\{x\}$

c. $f(x) = .6\,e^{-x/4}\,\mathrm{I}_{(0,\infty)}\,(x)$

d. $f(x) = x^{-1}\,I_{[1,e]}\,(x)$

2. Graph each of the *probability density* functions in Problem 1.

3. Sparkle Cola, Inc., manufactures a cola drink. The cola is sold in 12 oz. bottles. The probability distribution associated with the random variable whose outcome represents the actual quantity of soda place in a bottle of Sparkle Cola by the soda bottling line is specified to be

$$f(x) = 50\left[e^{-100(12-x)}I_{(-\infty,12]}(x) + e^{-100(x-12)}I_{(12,\infty)}(x)\right].$$

In order to be considered full, a bottle must contain within .25 oz. of 12 oz. of soda.

a. Define a random variable whose outcome indicates whether or not a bottle is considered full.

b. What is the range of this random variable?

c. Define a PDF for the random variable. Use it to assign probability to the event that a bottle is "considered full."

d. The PDF $f(x)$ is only an approximation. Why?

4. A health maintenance organization (HMO) is currently treating 10 patients with a deadly bacterial infection. The best-known antibiotic treatment is being used in these cases, and this treatment is effective 95 percent of the time. If the treatment is not effective, the patient expires.

a. Define a random variable whose outcome represents the number of patients being treated by the HMO that survive the deadly bacterial infection. What is the range of this random variable? What is the event space for outcome of this random variable?

b. Define the appropriate PDF for the random variable you defined in (a). Define the probability set function appropriate for assigning probabilities to events regarding the outcome of the random variable.

c. Using the probability space you defined in (a) and (b), what is the probability that all 10 of the patients survive the infection?

d. What is the probability that no more than two patients expire?

e. If 50 percent of the patients were to expire, the government would require that the HMO suspend operations, and an investigation into the medical practices of the HMO would be conducted. Provide an argument in defense of the government's actions in this case.

5. Star Enterprises is a small firm that produces a product that is simple to manufacture, involving only one variable input. The relationship between input and output levels is given by $q = x^{.5}$, where q is the quantity of product produced and x is the quantity of variable input used. For any given output and input prices, Star Enterprises operates at a level of production that maximizes its profit over variable cost. The possible prices in dollars facing the firm on a given day is represented by a random variable V with $R(V) = \{10,20,30\}$ and PDF

$$f(v) = .2I_{\{10\}}(v) + .5I_{\{20\}}(v) + .3I_{\{30\}}(v).$$

Input prices vary independently of output prices, and input price on a given day is the outcome of W with $R(W) = \{1,2,3\}$ and PDF

$$g(w) = .4I_{\{1\}}(w) + .3I_{\{2\}}(w) + .3I_{\{3\}}(w).$$

a. Define a random variable whose outcome represents Star's profit above variable cost on a given day. What is the range of the random variable? What is the event space?

b. Define the appropriate PDF for profit over variable cost. Define a probability set function appropriate for assigning probability to events relating to profit above variable cost.

c. What is the probability that the firm makes at least \$100 profit above variable cost?

d. What is the probability that the firm makes a positive profit on a given day? Is making a positive profit a certain event? Why or why not?

e. *Given* that the firm makes at least \$100 profit above variable cost, what is the probability that it makes at least \$200 profit above variable cost?

6. The ACME Freight Co. has containerized a large quantity of 4-gigabyte memory chips that are to be shipped to a personal computer manufacturer in California. The shipment contains 1,000 boxes of memory chips, with each box containing a dozen chips. The chip manufacturer calls and says that due to an error in manufacturing, each box contains exactly one defective chip. The defect can be detected through an easily administered nondestructive continuity test using an ohmmeter. The chip maker requests that ACME break open the container, find the defective chip in each box, discard them, and then reassemble the container for shipment. The testing of each chip requires 1 min to accomplish.

a. Define a random variable representing the amount of testing time required to find the defective chip in a box of chips. What is the range of the random variable? What is the event space?

b. Define a PDF for the random variable you have defined in (a). Define a probability set function appropriate for assigning probabilities to events relating to testing time required to find the defective chip in a box of chips.

c. What is the probability that it will take longer than 5 min to find the defective chip in a box of chips?

d. If ACME uses 28-hour-shift workers for one shift each to perform the testing, what is the probability that testing of all of the boxes in the container will be completed?

7. Intelligent Electronics, Inc., manufactures monochrome liquid crystal display (LCD) notebook computer screens. The number of hours an LCD screen functions until failure is represented by the outcome of a random variable X having range $R(X) = [0,\infty)$ and PDF

$$f(x) = .01 \exp\left(-\frac{x}{100}\right) I_{[0,\infty)}(x).$$

The value of x is measured in thousands of hours. The company has a 1-year warranty on its LCD screen, during which time the LCD screen will be replaced free of charge if it fails to function.

a. Assuming that the LCD screen is used for 8,760 hours per year, what is the probability that the firm will have to perform warranty service on an LCD screen?

b. What is the probability that the screen functions for at least 50,000 hours? *Given* that the screen has already functioned for 50,000 hours, what is the probability that it will function for at least *another* 50,000 hours?

8. People Power, Inc., is a firm that specializes in providing temporary help to various businesses. Job applicants are administered an aptitude test that evaluates mathematics, writing, and manual dexterity skills. After the firm analyzed thousands of job applicants who took the test, it was found that the scores on the three tests could be viewed as outcomes of random variables with the following joint density function (the tests are graded on a 0–1 scale, with 0 the lowest score and 1 the highest):

$$f(x_1, x_2, x_3) = .80(2x_1 + 3x_2)\, x_3 \prod_{i=1}^{3} I_{[0,1]}(x_i).$$

a. A job opening has occurred for an office manager. People Power, Inc., requires scores of $> .75$ on both the mathematics and writing tests for a job applicant to be offered the position. Define the marginal density function for the mathematics and writing scores. Use it to define a probability space in which probability questions concerning events for the mathematics and writing scores can be answered. What is the probability that a job applicant who has just entered the office to take the test will qualify for the office manager position?

b. A job opening has occurred for a warehouse worker. People Power, Inc., requires a score of $> .80$ on the manual dexterity test for a job applicant to be offered the position. Define the marginal density function for the dexterity score. Use it to define a probability space in which probability questions concerning events for the dexterity score can be answered. What is the probability that a job applicant who has just entered the office to take the test will qualify for the warehouse worker position?

c. Find the conditional density of the writing test score, given that the job applicant achieves a score of $> .75$ on the mathematics test. Given that the job applicant scores $> .75$ on the mathematics test, what is the probability that she scores $> .75$ on the writing test? Are the two test scores independent random variables?

d. Is the manual dexterity score independent of the writing and mathematics scores? Why or why not?

9. The weekly average price (in dollars/foot) and total quantity sold (measured in thousands of feet) of copper wire manufactured by the Colton Cable Co. can be viewed as the outcome of the bivariate random variable (P,Q) having the joint density function:

$$f(p,q) = 5\,pe^{-pq}\,I_{[.1,.3]}(p)\,I_{(0,\infty)}(q).$$

a. What is the probability that total dollar sales in a week will be less than \$2,000?

b. Find the marginal density of price. What is the probability that price will exceed \$.25/ft?

c. Find the conditional density of quantity, given price = .20. What is the probability that > 5,000 ft of cable will be sold in a given week?

d. Find the conditional density of quantity, given price = .10. What is the probability that > 5,000 ft of cable will be sold in a given week? Compare this result to your answer in (c). Does this make economic sense? Explain.

10. A personal computer manufacturer produces both desktop computers and notebook computers. The monthly proportions of customer orders received for desktop and notebook computers that are shipped within 1 week's time can be viewed as the outcome of a bivariate random variable (X,Y) with joint probability density

$$f(x,y) = (2-x-y)\,I_{[0,1]}(x)\,I_{[0,1]}(y).$$

a. In a given month, what is the probability that more than 75 percent of notebook computers and 75 percent of desktop computers are shipped within 1 week of ordering?

b. Assuming that an equal number of desktop and notebook computers are ordered in a given month, what is the probability that more than 75 percent of all orders received will be shipped within 1 week?

c. Are the random variables independent?

d. Define the conditional probability that less than 50 percent of the notebook orders are shipped within 1 week, given that x proportion of the desktop orders are shipped within 1 week (the probability will be a function of the proportion x). How does this probability change as x increases?

11. A small nursery has seven employees, three of whom are salespersons, and four of whom are gardeners who tend to the growing and caring of the nursery stock. With such a small staff, employee absenteeism can be critical. The number of salespersons and gardeners absent on any given day is the outcome of a bivariate random variable (X,Y). The nonzero values of the joint density function are given in tabular form as:

		Y				
		0	1	2	3	4
	0	.75	.025	.01	.01	.03
X	1	.06	.03	.01	.01	.003
	2	.025	.01	.005	.005	.002
	3	.005	.004	.003	.002	.001

a. What is the probability that more than two employees will be absent on any given day?

b. Find the marginal density function of the number of gardeners that are absent. What is the probability that more than two gardeners will be absent on any given day?

c. Are the number of gardener absences and the number of salesperson absences independent random variables?

d. Find the conditional density function for the number of salespersons that are absent, given that there are no gardeners absent. What is the probability that there are no salespersons absent, given that there are no gardeners absent? Is the conditional probability higher or lower given that there is at least one gardener absent?

12. The joint density of the bivariate random variable (X,Y) is given by

$$f(x,y) = xy\,I_{[0,1]}(x)\,I_{[0,2]}(y).$$

a. Find the joint cumulative distribution function of (X,Y). Use it to find the probability that $x \le .5$ and $y \le 1$.

b. Find the marginal cumulative distribution function of X. What is the probability that $x \le .5$?

c. Find the marginal density of X from the marginal cumulative distribution of X.

13. The joint cumulative distribution function for (X,Y) is given by

$$F(x,y) = \left(1 - e^{-x/10} - e^{-y/2} + e^{-(x+5y)/10}\right) I_{(0,\infty)}(x)\,I_{(0,\infty)}(y).$$

a. Find the joint density function of (X,Y).

b. Find the marginal density function of X.

c. Find the marginal cumulative distribution function of X.

14. The cumulative distribution of the random variable X is given by

$$F(x) = (1 - p^{x+1})\, I_{\{0,1,2,\ldots\}}(x), \text{for some choice of } p \in (0,1).$$

a. Find the density function of the random variable X.

b. What is the probability that $x \leq 8$ if $p = .75$?

c. What is the probability that $x \leq 1$ given that $x \leq 8$?

15. The federal mint uses a stamping machine to make coins. Each stamping produces 10 coins. The number of the stamping at which the machine breaks down and begins to produce defective coins can be viewed as the outcome of a random variable, X, having a PDF with general functional form $f(x) = \alpha\,(1 - \beta)^{x-1}\, I_{\{1,\,2,\,3,\,\ldots\}}(x)$, where $\beta \in (0,1)$.

a. Are there any constraints on the choice of α for $f(x)$ to be a PDF? If so, precisely what are they?

b. Is the random variable X a discrete or a continuous random variable? Why?

c. It is known that the probability the machine will break down on the first stamping is equal to .05. What is the specific functional form of the PDF $f(x)$? What is the probability that the machine will break down on the tenth stamping?

d. Continue to assume the results in (a–c). Derive a functional representation for the cumulative distribution function corresponding to the random variable X. Use it to assign the appropriate probability to the event that the machine does not break down for at least 10 stampings.

e. What is the probability that the machine does not break down for at least 20 stampings, *given* that the machine does not break down for at least 10 stampings?

16. The daily quantity demanded of unleaded gasoline in a regional market can be represented as $Q = 100 - 10p + E$, where $p \in [0,8]$, and E is a random variable having a probability density given by $f(e) = 0.025 I_{[-20,20]}(e)$.

Quantity demanded, Q, is measured in thousands of gallons, and price, p, is measured in dollars.

a. What is the probability of the quantity demanded being greater than 70,000 gal if price is equal to \$4? if price is equal to \$3?

b. If the average variable cost of supplying Q amount of unleaded gasoline is given by $C(Q) = Q^{.5}/2$, define a random variable that can be used to represent the daily profit above variable cost from the sale of unleaded gasoline.

c. If price is set equal to \$4, what is the probability that there will be a positive profit above variable cost on a given day? What if price is set to \$3? to \$5?

d. If price is set to \$6, what is the probability that quantity demanded will equal 40,000 gal?

17. For each of the cumulative distribution functions listed below, find the associated PDFs. For each CDF, calculate $P(x \leq 6)$.

a. $F(b) = (1 - e^{-b/6})\, I_{(0,\infty)}(b)$

b. $F(b) = (5/3)\,(.6 - .6^{\text{trunc}(b)+1}) I_{[0,\infty)}(b)$

18. An economics class has a total of 20 students with the following age distribution:

# of students	age
10	19
4	20
4	21
1	24
1	29

Two students are to be selected randomly, without replacement, from the class to give a team report on the state of the economy. Define a random variable whose outcome represents the average age of the two students selected. Also, define a discrete PDF for the random variable. Finally, what is the probability space for this experiment?

19. Let X be a random variable representing the *minimum* of the two numbers of dots that are facing up after a pair of fair dice is rolled. Define the appropriate probability density for X. What is the probability space for the experiment of rolling the fair dice and observing the minimum of the two numbers of dots?

20. A package of a half-dozen light bulbs contains two defective bulbs. Two bulbs are randomly selected from the package and are to be used in the same light fixture. Let the random variable X represent the number of light bulbs selected that function properly (i.e., that are not defective). Define the appropriate PDF for X. What is the probability space for the experiment?

21. A committee of three students will be randomly selected from a senior-level political science class to present an assessment of the impacts of an antitax initiative to some visiting state legislators. The class consists of five economists, eight political science majors, four business majors, and three art majors. Referring to the experiment of drawing three students randomly from the class, let the bivariate random variable (X,Y) be defined by $x =$ number of economists on the committee, and $y =$ number of business majors on the committee.

a. What is the range of the bivariate random variable (X,Y)? What is the PDF, $f(x,y)$, for this bivariate random variable? What is the probability space?

b. What is the probability that the committee will contain at least one economist and at least one business major?

c. What is the probability that the committee will consist of only political science and art majors?

d. *On the basis of the probability space you defined in (a) above*, is it possible for you to assign probability to the event that the committee will consists entirely of art majors? Why or why not? If you answer yes, calculate this probability using $f(x,y)$ from (a).

e. Calculate the marginal density function for the random variable X. What is the probability that the committee contains three economists?

f. Define the conditional density function for the number of business majors on the committee, *given* that the committee contains two economists. What is the probability that the committee contains less than one business major, *given* that the committee contains two economists?

g. Define the conditional density function for the number of business majors on the committee, *given* that the committee contains at least two economists. What is the probability that the committee contains less than one business major, *given* that the committee contains at least two economists?

h. Are the random variables X and Y independent? Justify your answer.

22. The Imperial Electric Co. makes high-quality portable compact disc players for sale in international and domestic markets. The company operates two plants in the United States, where one plant is located in the Pacific Northwest and one is located in the South. At either plant, once a disc player is assembled, it is subjected to a stringent quality-control inspection, at which time the disc player is either approved for shipment or else sent back for adjustment before it is shipped. On any given day, the proportion of the units produced at each plant that require adjustment before shipping, and the total production of disc players at the company's two plants, are outcomes of a trivariate random variable, with the following joint PDF:

$$f(x,y,z) = \tfrac{2}{3}(x+y)\,e^{-x}\, I_{(0,\infty)}(x)\, I_{(0,1)}(y)\, I_{(0,1)}(z),$$

where
$x =$ total production of disc players at the two plants, measured in thousands of units,
$y =$ proportion of the units produced at the Pacific Northwest plant that are shipped without adjustment, and
$z =$ proportion of the units produced in the southern plant that are shipped without adjustment.

a. In this application, the use of a *continuous* trivariate random variable to represent proportions and total production values must be viewed as only an *approximation* to the underlying real-world situation. Why? In the remaining parts, assume the approximation is acceptably accurate, and use the approximation to answer questions where appropriate.

b. What is the probability that less than 50 percent of the disc players produced in each plant will be shipped without adjustment and that production will be less than 1,000 units on a given day?

c. Derive the marginal PDF for the total production of disc players at the two plants. What is the probability that less than 1,000 units will be produced on a given day?

d. Derive the marginal PDF for the bivariate random variable (Y,Z). What is the probability that more than 75 percent of the disc players will be shipped without adjustment from each plant?

e. Derive the conditional density function for X, *given* that 50 percent of the disc players are shipped from the Pacific Northwest plant without adjustment. What is the probability that 1,500 disc players will be produced by the Imperial Electric Co. on a day for which 50 percent of the disc players are shipped from the Pacific Northwest plant without adjustment?

f. Answer (e) for the case where 90 percent of the disc players are shipped from the Pacific Northwest plant without adjustment.

g. Are the random variables (X, Y, Z) independent random variables?

h. Are the random variables (Y, Z) independent random variables?

23. ACE Rentals, a car-rental company, rents three types of cars: compacts, mid-size sedans, and large luxury cars. Let (x_1, x_2, x_3) represent the number of compacts, mid-size sedans, and luxury cars, respectively, that ACE rents per day. Let the sample space for the possible outcomes of (X_1, X_2, X_3) be given by

$S = \{ (x_1,x_2,x_3) : x_1,x_2, \text{ and } x_3 \in (0,1,2,3)\}$

(ACE has an inventory of nine cars, evenly distributed among the three types of cars).
The discrete PDF associated with (X_1, X_2, X_3) is given by

$$f(x_1, x_2, x_3) = \left[\frac{.004(3 + 2x_1 + x_2)}{(1 + x_3)}\right] \prod_{i=1}^{3} I_{\{0,1,2,3\}}(x_i).$$

The compact car rents for \$20/day, the mid-size sedan rents for \$30/day, and the luxury car rents for \$60/day.

a. Derive the marginal density function for X_3. What is the probability that all three luxury cars are rented on a given day?

b. Derive the marginal density function for (X_1, X_2). What is the probability of more than one compact and more than one mid-size sedan being rented on a given day?

c. Derive the conditional density function for X_1, given $x_2 \leq 2$. What is the probability of renting no more than one compact care, given that two or more mid-size sedans are rented?

d. Are X_1, X_2, and X_3 jointly independent random variables? Why or why not? Is (X_1, X_2) independent of X_3?

e. Derive the conditional density function for (X_1, X_2), given that $x_3 = 0$. What is the probability of renting more than one compact and more than one mid-size sedan given that no luxury cars are rented?

f. If it costs \$150/day to operate ACE Rentals, define a random variable that represents the daily profit made by the company. Define an appropriate density function for this random variable. What is the probability that ACE Rentals makes a positive daily profit on a given day?

24. If (X_1, X_2) and (X_3, X_4) are independent bivariate random variables, are X_2 and X_3 independent random variables? Why or why not?

25. The joint density function of the discrete trivariate random variable (X_1, X_2, X_3) is given by

$$f(x_1, x_2, x_3) = .20\, I_{\{0,1\}}(x_1)\, I_{\{0,1\}}(x_2)\, I_{\{|x_1 - x_2|\}}(x_3) + .05\, I_{\{0,1\}}(x_1)\, I_{\{0,1\}}(x_2)\, I_{\{1-|x_1 - x_2|\}}(x_3).$$

a. Are (X_1, X_2), (X_1, X_3), and (X_2, X_3) *each* pairwise independent random variables?

b. Are X_1, X_2, X_3 jointly independent random variables?

26. SUPERCOMP, a retail computer store, sells personal computers and printers. The number of computers and printers sold on any given day varies, with the probabilities of the various possible sales outcomes being given by the following table:

		Number of computers sold					
		0	1	2	3	4	
	0	.03	.03	.02	.02	.01	Probabilities of elementary events
Number	1	.02	.05	.06	.02	.01	
of	2	.01	.02	.10	.05	.05	
printers	3	.01	.01	.05	.10	.10	
	4	.01	.01	.01	.05	.15	

a. If SUPERCOMP has a profit margin (product sales price – product unit cost) of \$100 per computer sold and \$50 per printer sold, define a random variable representing aggregate profit margin from the sale of computers and printers on a given day. What is the range of this random variable?

b. Define a discrete density function appropriate for use in calculating probabilities of all events concerning aggregate profit margin outcomes on a given day.

c. What is the probability that the aggregate profit margin is $\geq$ \$300 on a given day?

d. The daily variable cost of running the store is \$200/day. What is the probability that SUPERCOMP's aggregate profit margin on computer and printer sales will equal or exceed variable costs on a given day?

e. Assuming that events involving the number of computers and printers sold are independent from day to day, what is the probability that for any given 6-day business week, aggregate profit margins equal or exceed variable cost all 6 days?

27. Given the function definitions below, determine which can be used as PDFs (PDFs) and which cannot. Justify your answers.

a. $f(x) = \begin{cases} \left(\frac{1}{4}\right)^x & \text{for} \quad x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$

b. $f(x) = \left(\frac{1}{4}\right)^x I_{(0,\infty)}(x)$

c. $f(x, y) = \begin{cases} (2x + y)/100, & \text{for} \quad x \text{ and } y = 0, 1, 2, 3, 4, \text{ and } y \le x \\ 0 & \text{otherwise} \end{cases}$

d. $f(x, y) = 6xy^2 I_{[0,1]}(x) I_{[0,1]}(y)$

28. Given the function definitions below, determine which can be used as cumulative distribution functions (CDFs) and which cannot. Justify your answers.

a. $F(c) = \dfrac{e^c}{1 + e^c}$ for $c \in (-\infty, \infty)$

b. $F(c) = \begin{cases} 1 - x^{-2}, & \text{for} \quad c \in (1, \infty) \\ 0 & \text{otherwise} \end{cases}$

c. $F(c) = \begin{cases} 1 - (.5)^{floor(c)} & \text{for} \quad c \ge 1 \\ 0 & \text{otherwise.} \end{cases}$

where *floor* $(c) \equiv$ round down the value c.

d.

$$F(c_1, c_2) = \begin{cases} 1 \ \textit{if} \ c_1 \text{ and } c_2 \in (1, \infty) \\ c_1^3 I_{[0,1]}(c_1) \textit{if} \ c_2 \in (1, \infty) \\ c_2^2 I_{[0,1]}(c_2) \textit{if} \ c_1 \in (1, \infty) \\ c_1^3 c_2^2 I_{[0,1]}(c_1) I_{[0,1]}(c_2) \text{ for } c_1 \text{ and } c_2 \in (-\infty, 1) \end{cases}$$

e. $F(c_1, c_2) = (1 - e^{-c_1})(1 - e^{-c_2}) I_{[0,\infty)}(c_1) I_{[0,\infty)}(c_2)$

29. For those functions in (28) that are actually cumulative distribution functions (CDFs), use the duality principle to derive the PDFs (PDFs) that are associated with the CDFs.

30. The daily quantity demanded of milk in a regional market, measured in 1,000's of gallons, can be represented during the summer months as the outcome of the following random variable:

$Q = 200 - 50p + V$,

where V is a random variable having a probability density defined by

$f(v) = 0.02 I_{[-25,25]}(v)$ and p is the price of milk, in dollars per gallon.

a. What is the probability that the quantity demanded will be greater than 100,000 gal if price is equal to \$2? if price is equal to \$2.25?

b. If the variable cost of supplying Q amount of milk is given by the cost function $C(Q) = 20\,Q^{.5}$, define a random variable that represents the daily profit above variable cost from the sale of milk.

c. If price is equal to \$2, what is the probability that there will be a positive profit above variable cost on a given day? What if price is set to \$2.25?

d. Is there any conceptual problem with using the demand function listed above to model quantity demanded if $p = 4$? If so, what is it?

31. A small locally-owned hardware store in a western college town accepts both cash and checks for purchasing merchandise from the store. From experience, the store accountant has determined that 2 percent of the checks that are written for payment are "bad" (i.e., they are refused by the bank) and cannot be cashed. The accountant defines the following probability model $(R(X), f(x))$ for the outcome of a random variable X denoting the number of bad checks that occur in n checks received on a given day at the store:

$$f(x) = \begin{cases} \frac{n!}{(n-x)!x!} (.02)^x (.98)^{n-x} \text{ for} & x \in R(X) = \{0, 1, 2, \ldots, n\} \\ 0 & \text{elsewhere} \end{cases}$$

If the store receives 10 checks for payment on a given day, what is the probability that:

a. Half are bad?

b. No more than half are bad?

c. None are bad?

d. None are bad, *given that* no more than half are bad?

32. Let an outcome of the random variable T represent the time, in minutes, that elapses between when an order is placed at a ticket counter by a customer and when the ticket purchase is completed. The following probability model $(R(T), f(t))$ governs the behavior of the random variable T:

$$f(t) = \begin{cases} 3e^{-3t} \text{ for } t & \in R\{T\} = [0, \infty) \\ 0 & \text{elsewhere} \end{cases}$$

a. What is the probability that the customer waits less than 3 min to have her ticket order completed?

b. Derive the cumulative distribution function for T. Use it to define the probability that it takes longer than 10 min to have the ticket order completed.

c. Given that the customer's wait will be less than 3 min, what is the probability that it will be less than 1 min?

d. *Given* that the customer has already waited more than 3 min, what is the probability that the customer will wait *at least* another 3 min to have the ticket order completed?

33. Outcomes of the random variable Z represent the number of customers that are waiting in a queue to be serviced at Fast Lube, a quick stop automobile lubrication business, when the business opens at 9 A.M. on any given Saturday. The probability model $(R(Z), f(z))$ for the random variable Z is given by:

$$f(z) = \begin{cases} .5^{z+1} \text{ for } z & \in R(Z) = \{0, 1, 2, 3, \ldots\} \\ 0 & \text{elsewhere} \end{cases}$$

a. Derive the cumulative distribution function for Z.

b. What is the probability that there will be less than 10 people waiting?

c. What is the probability that there will be more than 3 people waiting?

d. Given that no more than two people will be waiting, what is the probability that there will be no customers when business opens at 9 A.M.?

34. The daily wholesale price and quantity sold of ethanol in a Midwestern regional market during the summer months is represented by the outcome of a bivariate random variable (P,Q) having the following probability model $(R(P,Q), f(p,q))$:

$$f(p,q)\begin{cases} .5pe^{-pq} \text{ for } & (p,q) \in R(P,Q) = [2,4] \times [0,\infty) \\ 0 & \text{elsewhere} \end{cases}$$

where price is measured in dollars and quantity is measured in 100,000 gal units (e.g., $q = 2$ means 200,000 gal were sold).

a. Derive the marginal probability density of price. Use it to determine the probability that price will exceed \$3.

b. Derive the marginal cumulative distribution function for price. Use it to verify your answer to part (a) above.

c. Derive the marginal probability density of quantity. Use it to determine the probability that quantity sold will be less than \$500,000 gal.

d. Let the random variable $D = PQ$ denote the daily total dollar sales of ethanol during the summer months. What is the probability that daily total dollar sales will exceed \$300,000?

e. Are P and Q independent random variables?

35. The BigVision Electronic Store sells a large 73 inch diagonal big screen TV. The TV comes with a standard 1 year warranty on parts and labor so that if anything malfunctions on the TV in the first year of ownership, the company repairs or replaces the TV for free. The store also sells an "extended warranty" which a customer can purchase that extends warranty coverage on the TV for another 2 years, for a total of 3 years of coverage. The daily numbers of TVs and extended warranties sold can be viewed as the outcome of a bivariate random variable (T, W) with probability model $(R(T,W), f(t,w))$ given by

$$f(t,w)\begin{cases} (2t + w)/100, \text{ for } & t \text{ and } w = 0, 1, 2, 3, 4, \text{ and } w \leq t \\ 0 & \text{otherwise} \end{cases}$$

a. What is the probability that all of the TVs sold on a given day will be sold with extended warranties?

b. Derive the marginal density function for the number of TVs sold. Use it to define the probability that ≤ 2 TVs are sold on a given day?

c. Derive the marginal density function for the number of warranties sold. What is the probability that ≥ 3 warranties are sold on a given day?

d. Are T and W independent random variables?

36. The following function is proposed as a cumulative distribution function for the bivariate random variable (X,Y):

$$F(x,y) = \left(1 + e^{-(x/10 + y/20)} - e^{-x/10} - e^{-y/20}\right) I_{(0,\infty)}(x) I_{(0,\infty)}(y)$$

a. Verify that the function has the appropriate properties to serve as a cumulative distribution function.

b. Derive the marginal cumulative distribution function of Y.

c. Derive the marginal PDF of Y.

d. Derive the joint PDF of (X,Y).

e. What is the probability that $X \leq 10$ and $Y \leq 20$?

f. Are X and Y independent random variables?

37. For each of the joint PDFs listed below, determined which random variables are independent and which are not.

a. $f(x,y) = e^{-(x+y)} I_{[0,\infty)}(x) I_{[0,\infty)}(y)$

b. $f(x,y) = \dfrac{x(1+y)}{300} I_{\{1,2,3,4,5\}}(x) I_{\{1,2,3,4,5\}}(y)$

c. $f(x,y,z) = 8xyz I_{[0,1]}(x) I_{[0,1]}(y) I_{[0,1]}(z)$

d. $f(x_1,x_2,x_3) = \dfrac{.5^{x_1} .2^{x_2} .75^{x_3}}{10} \prod_{i=1}^{3} I_{\{0,1,2,\ldots\}}(x_i)$

38. The daily wholesale price and quantity sold of ethanol in a Midwestern regional market during the summer months is represented by the outcome of a bivariate random variable (P,Q) having the following probability model $\{R(P,Q), f(p,q)\}$:

$$f(p,q) = \begin{cases} .5pe^{-pq} & \text{for} \quad (p,q) \in R(P,Q) = [2,4] \times [0,\infty) \\ 0 & \text{elsewhere} \end{cases}$$

where price is measured in dollars and quantity is measured in 100,000 gal units (e.g., $q = 2$ means 200,000 gal were sold).

a. Derive the conditional-on-p PDF for quantity sold.

b. What is the probability that quantity sold exceeds 50,000 gal if price = \$2. What is the probability that quantity sold exceeds 50,000 gal if price = \$4. Does this make economic sense?

c. What is the probability that quantity sold exceeds 50,000 gal if price is *greater than or equal to* \$3.00?

39. Let the random variable X represent the product of the number of dots facing up on each die after a pair of fair dice is rolled. Let Y represent the sum of the number of dots facing up on the pair of dice.

a. Define a probability model $(R(X), f(x))$ for the random variable X.

b. What is the probability that $X \geq 16$?

c. Define a probability model $(R(X,Y), f(x,y))$ for the random vector (X,Y).

d. What is the probability that $X \geq 16$ and $Y \geq 8$?

e. Are X and Y independent random variables?

f. Define the conditional PDF of X given that $Y = 7$.

g. What is the probability that $X \geq 10$ given that $Y = 7$?

40. The production of a certain volatile commodity is the outcome of a stochastic production function given by $Y = L^{.5} K^{.25} e^{v}$, where v is a random variable having the cumulative distribution function $F(v) = \frac{1}{1+e^{-2(v-1)}}$, L denotes units of labor and K denotes units of capital.

a. If labor is applied at 9 units and capital is applied at 16 units, what is the probability that output will exceed 12 units?

b. Given the input levels applied in (a), what is the probability that output will be between 12 and 16 units?

c. What level of capital and labor should be applied so that the probability of producing a positive profit is *maximized* when output price is \$10, labor price is \$5, and capital price is \$10?

d. What is the value of the maximum probability of obtaining positive profit?

Springer
http://www.springer.com/978-1-4614-5021-4

Mathematical Statistics for Economics and Business
Mittelhammer, R.C.
2013, XXIX, 755 p., Hardcover
ISBN: 978-1-4614-5021-4