4

Continuous Random Variables and Probability Distributions

4.1 Probability Density Functions

Probability Density Functions

A discrete random variable (rv) is one whose possible values either constitute a finite set or else can be listed in an infinite sequence (a list in which there is a first element, a second element, etc.).

A random variable whose set of possible values is an entire interval of numbers is not discrete.

Probability Density Functions

Recall from Chapter 3 that a random variable X is continuous if

(1) possible values comprise either a single interval on the number line (for some $A < B$, any number x between A and B is a possible value) or a union of disjoint intervals, and

(2) $P(X = c) = 0$ for any number c that is a possible value of X.

Example 1

If in the study of the ecology of a lake, we make depth measurements at randomly chosen locations, then X = the depth at such a location is a continuous rv.

Here A is the minimum depth in the region being sampled, and B is the maximum depth.

Probability Density Functions

One might argue that although in principle variables such as height, weight, and temperature are continuous, in practice the limitations of our measuring instruments restrict us to a discrete (though sometimes very finely subdivided) world.

However, continuous models often approximate real-world situations very well, and continuous mathematics (the calculus) is frequently easier to work with than mathematics of discrete variables and distributions.

Probability Distributions for Continuous Variables

Probability Distributions for Continuous Variables

Suppose the variable X of interest is the depth of a lake at a randomly chosen point on the surface.

Let M = the maximum depth (in meters), so that any number in the interval [0, M] is a possible value of X.

If we "discretize" X by measuring depth to the nearest meter, then possible values are nonnegative integers less than or equal to M.

The resulting discrete distribution of depth can be pictured using a probability histogram.

Probability Distributions for Continuous Variables

If we draw the histogram so that the area of the rectangle above any possible integer k is the proportion of the lake whose depth is (to the nearest meter) k, then the total area of all rectangles is 1. A possible histogram appears in Figure 4.1(a).
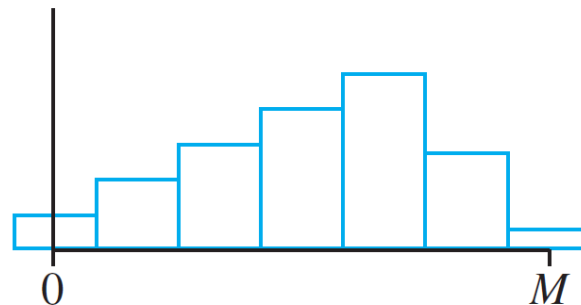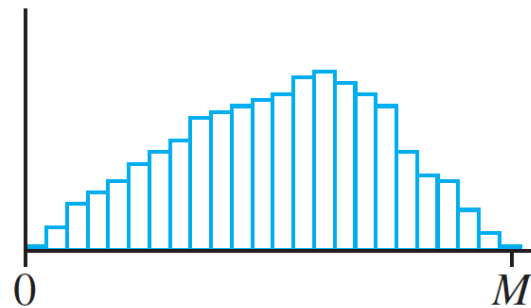

Probability histogram of depth measured to the nearest meter

Figure 4.1(a)

Probability Distributions for Continuous Variables

If depth is measured much more accurately and the same measurement axis as in Figure 4.1(a) is used, each rectangle in the resulting probability histogram is much narrower, though the total area of all rectangles is still 1. A possible histogram is pictured in Figure 4.1(b).
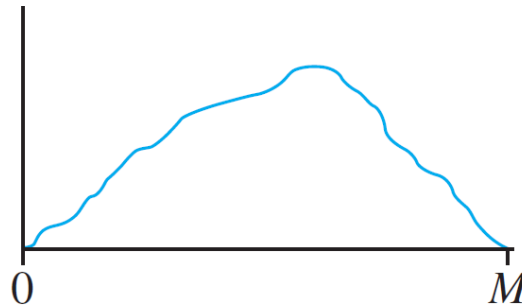

Probability histogram of depth measured to the nearest centimeter

Figure 4.1(b)

Probability Distributions for Continuous Variables

It has a much smoother appearance than the histogram in Figure 4.1(a). If we continue in this way to measure depth more and more finely, the resulting sequence of histograms approaches a smooth curve, such as is pictured in Figure 4.1(c).

A limit of a sequence of discrete histograms

Figure 4.1(c)

Probability Distributions for Continuous Variables

Because for each histogram the total area of all rectangles equals 1, the total area under the smooth curve is also 1.

The probability that the depth at a randomly chosen point is between a and b is just the area under the smooth curve between a and b. It is exactly a smooth curve of the type pictured in Figure 4.1(c) that specifies a continuous probability distribution.
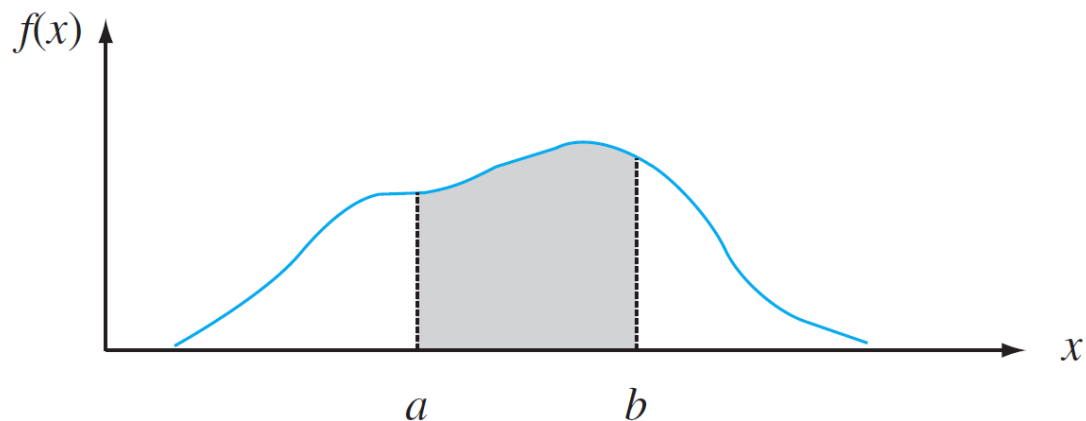
Probability Distributions for Continuous Variables

Definition

Let X be a continuous rv. Then a **probability distribution** or **probability density function** (pdf) of X is a function $f(x)$ such that for any two numbers a and b with $a \le b$,

$$P(a \le X \le b) = \int_a^b f(x)dx$$

Probability Distributions for Continuous Variables

That is, the probability that *X* takes on a value in the interval [*a*, *b*] is the area above this interval and under the graph of the density function, as illustrated in Figure 4.2.

$P(a \leq X \leq b)$ = the area under the density curve between *a* and *b*

Figure 4.2

The graph of *f*(*x*) is often referred to as the *density curve*.

Probability Distributions for Continuous Variables

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

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

2. $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx$ = area under the entire graph of $f(x)$

$$= 1$$

Example 4

The direction of an imperfection with respect to a reference line on a circular object such as a tire, brake rotor, or flywheel is, in general, subject to uncertainty.
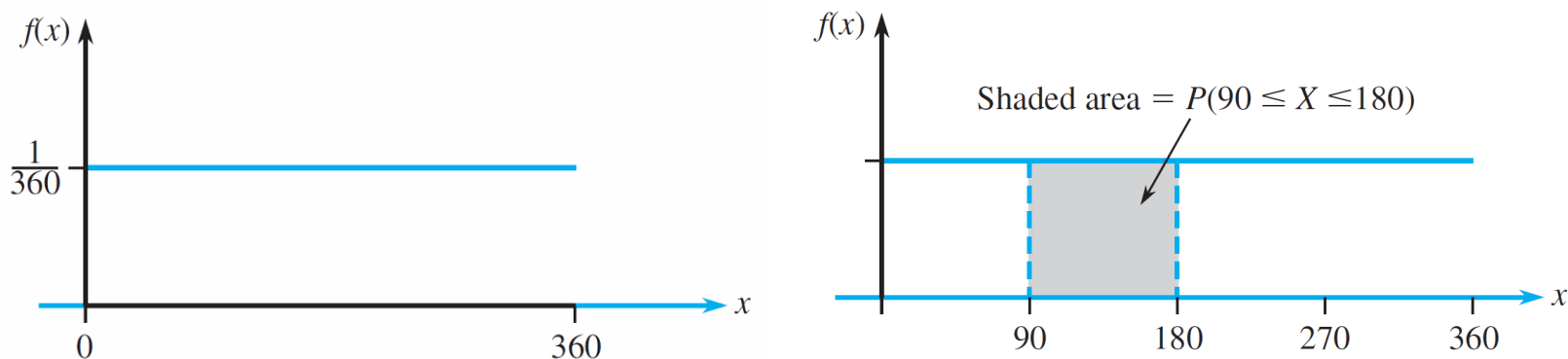
Consider the reference line connecting the valve stem on a tire to the center point, and let *X* be the angle measured clockwise to the location of an imperfection. One possible pdf for *X* is

$$f(x) = \begin{cases} \dfrac{1}{360} & 0 \leq x < 360 \\ 0 & \text{otherwise} \end{cases}$$

Example 4

cont'd

The pdf is graphed in Figure 4.3.

The pdf and probability from Example 4

Figure 4.3

Example 4

Clearly $f(x) \geq 0$. The area under the density curve is just the area of a rectangle:

$$(\text{height})(\text{base}) = \left(\frac{1}{360}\right)(360) = 1.$$

The probability that the angle is between 90° and 180° is

$$P(90 \leq X \leq 180) = \int_{90}^{180} \frac{1}{360}\ dx$$

$$= \frac{x}{360} \bigg|_{x=90}^{x=180}$$

$$= \frac{1}{4} = .25$$

18

Example 4
cont'd

The probability that the angle of occurrence is within 90° of the reference line is

$$P(0 \leq X \leq 90) + P(270 \leq X < 360) = .25 + .25 = .50$$

Probability Distributions for Continuous Variables

Because whenever $0 \leq a \leq b \leq 360$ in Example 4.4 and $P(a \leq X \leq b)$ depends only on the width $b - a$ of the interval, X is said to have a uniform distribution.

Definition

A continuous rv X is said to have a **uniform distribution** on the interval $[A, B]$ if the pdf of X is

$$f(x; A, B) = \begin{cases} \dfrac{1}{B - A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

Probability Distributions for Continuous Variables

The graph of any uniform pdf looks like the graph in Figure 4.3 except that the interval of positive density is [A, B] rather than [0, 360].

In the discrete case, a probability mass function (pmf) tells us how little "blobs" of probability mass of various magnitudes are distributed along the measurement axis.

In the continuous case, probability density is "smeared" in a continuous fashion along the interval of possible values. When density is smeared uniformly over the interval, a uniform pdf, as in Figure 4.3, results.

Probability Distributions for Continuous Variables

When X is a discrete random variable, each possible value is assigned positive probability.

This is not true of a continuous random variable (that is, the second condition of the definition is satisfied) because the area under a density curve that lies above any single value is zero:

$$P(X = c) = \int_{c}^{c} f(x)\,dx = \lim_{\varepsilon \to 0} \int_{c-\varepsilon}^{c+\varepsilon} f(x)\,dx = 0$$

Probability Distributions for Continuous Variables

The fact that $P(X = c) = 0$ when X is continuous has an important practical consequence: The probability that X lies in some interval between a and b does not depend on whether the lower limit a or the upper limit b is included in the probability calculation:

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) \quad (4.1)$$

If X is discrete and both a and b are possible values (e.g., X is binomial with $n = 20$ and $a = 5$, $b = 10$), then all four of the probabilities in (4.1) are different.

Probability Distributions for Continuous Variables

The zero probability condition has a physical analog. Consider a solid circular rod with cross-sectional area = 1 in^2.

Place the rod alongside a measurement axis and suppose that the density of the rod at any point x is given by the value $f(x)$ of a density function. Then if the rod is sliced at points a and b and this segment is removed, the amount of mass removed is $\int_a^b f(x)\,dx$; if the rod is sliced just at the point c, no mass is removed.

Mass is assigned to interval segments of the rod but not to individual points.

Example 5

"Time headway" in traffic flow is the elapsed time between the time that one car finishes passing a fixed point and the instant that the next car begins to pass that point.

Let X = the time headway for two randomly chosen consecutive cars on a freeway during a period of heavy flow. The following pdf of X is essentially the one suggested in "The Statistical Properties of Freeway Traffic" (*Transp. Res.,* vol. 11: 221–228):

$$f(x) = \begin{cases} .15e^{-.15(x-.5)} & x \geq .5 \\ 0 & \text{otherwise} \end{cases}$$

Example 5

cont'd

The graph of $f(x)$ is given in Figure 4.4; there is no density associated with headway times less than .5, and headway density decreases rapidly (exponentially fast) as x increases from .5.
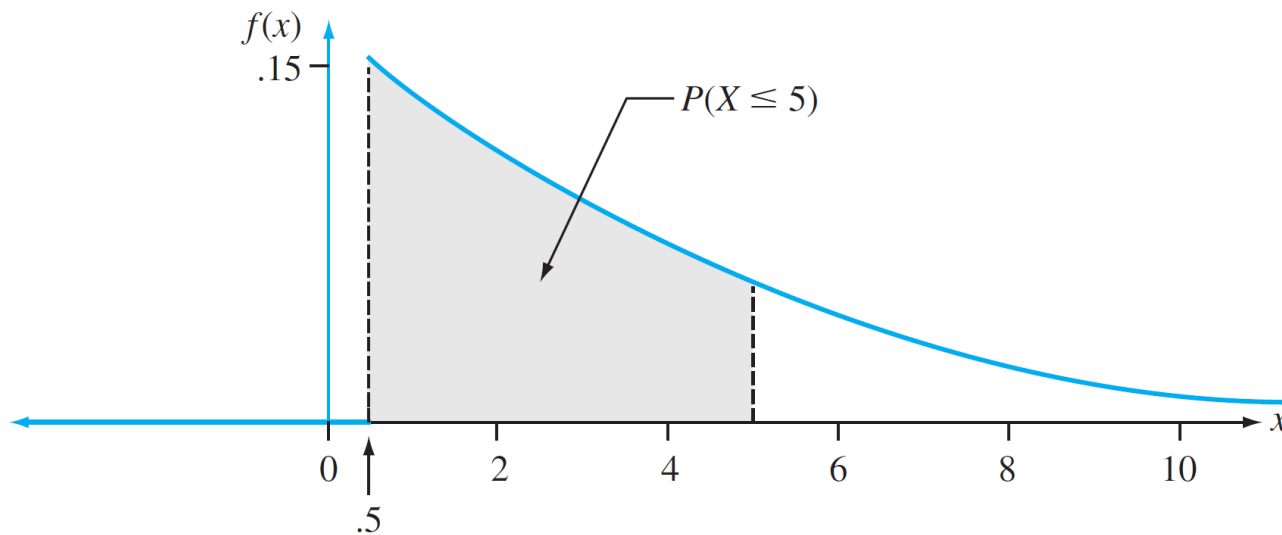
The density curve for time headway in Example 5

Figure 4.4

Example 5

cont'd

Clearly, $f(x) \geq 0$; to show that $\int_{-\infty}^{\infty} f(x)\,dx = 1$, we use the calculus result

$$\int_a^\infty e^{-kx}\,dx = (1/k)e^{-k \cdot a}.$$

Then

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{.5}^{\infty} .15e^{-.15(x-.5)}\,dx$$

$$= .15e^{.075} \int_{.5}^{\infty} e^{-.15x}\,dx$$

$$= .15e^{.075} \cdot \frac{1}{.15} e^{-(.15)(.5)}$$

$$= 1$$

Example 5

cont'd

The probability that headway time is at most 5 sec is

$$P(X \leq 5) = \int_{-\infty}^{5} f(x)\, dx$$

$$= \int_{.5}^{5} .15 e^{-.15}(x - .5)\, dx$$

$$= .15 e^{.075} \int_{.5}^{5} e^{-15x}\, dx$$

$$= .15 e^{.075} \cdot \left(-\frac{1}{.15} e^{-.15x} \Big|_{x=.5}^{x=5} \right)$$

Example 5

cont'd

$$= e^{.075}(-e^{-.75} + e^{-.075})$$

$$= 1.078(-.472 + .928)$$

$$= .491$$

$$= P(\text{less than 5 sec})$$

$$= P(X < 5)$$

Probability Distributions for Continuous Variables

Unlike discrete distributions such as the binomial, hypergeometric, and negative binomial, the distribution of any given continuous rv cannot usually be derived using simple probabilistic arguments.

Just as in the discrete case, it is often helpful to think of the population of interest as consisting of X values rather than individuals or objects.

The pdf is then a model for the distribution of values in this numerical population, and from this model various population characteristics (such as the mean) can be calculated.