5.2: Properties of Continuous Probability Density Functions

The graph of a continuous probability distribution is a curve. Probability is represented by area under the curve. We have already met this concept when we developed relative frequencies with histograms in Chapter 2. The relative area for a range of values was the probability of drawing at random an observation in that group. Again with the Poisson distribution, the graph in Example 4.3.3 used boxes to represent the probability of specific values of the random variable. In this case, we were being a bit casual because the random variables of a Poisson distribution are discrete, whole numbers, and a box has width. Notice that the horizontal axis, the random variable X, purposefully did not mark the points along the axis. The probability of a specific value of a continuous random variable will be zero because the area under a point is zero. Probability is area.

The curve is called the **probability density function** (abbreviated as **pdf**). We use the symbol $f(x))$ to represent the curve. $f(x))$ is the function that corresponds to the graph; we use the density function $f(x))$ to draw the graph of the probability distribution.

Area under the curve is given by a different function called the **cumulative distribution function** (abbreviated as **cdf**). The cumulative distribution function is used to evaluate probability as area. Mathematically, the cumulative probability density function is the integral of the pdf, and the probability between two values of a continuous random variable will be the integral of the pdf between these two values: the area under the curve between these values. Remember that the area under the pdf for all possible values of the random variable is one, certainty. Probability thus can be seen as the relative percent of certainty between the two values of interest.

- The outcomes are measured, not counted.
- The entire area under the curve and above the x-axis is equal to one.
- Probability is found for *intervals* of x values rather than for individual x values.
- $P(c<X<d)$ is the probability that the random variable X is in the interval between the values c and d. $P(c<X<d)$ is the area under the curve, above the x-axis, to the right of c and the left of d.
- $P(X=c)=0$ The probability that X takes on any single individual value is zero. The area below the curve, above the x-axis, and between $x=c$ and $x=c$ has no width, and therefore no area ($\text{area } =0$). Since the probability is equal to the area, the probability is also zero.
- $P(c<X<d)$ is the same as $P(c \leq x \leq d)$ because probability is equal to area.

We will find the area that represents probability by using geometry, formulas, technology, or probability tables. In general, integral calculus is needed to find the area under the curve for many probability density functions. When we use formulas to find the area in this textbook, the formulas were found by using the techniques of integral calculus.

There are many continuous probability distributions. When using a continuous probability distribution to model probability, the distribution used is selected to model and fit the particular situation in the best way.

In this chapter, we will study the uniform distribution, the exponential distribution, and the normal distribution. The following graphs illustrate these distributions.

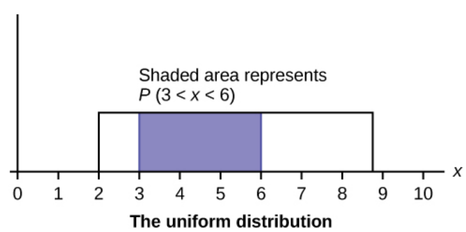

Figure 5.2.1 The graph shows a Uniform Distribution with the area between $x=3$ and $x=6$ shaded to represent the probability that the value of the random variable X is in the interval between three and six.

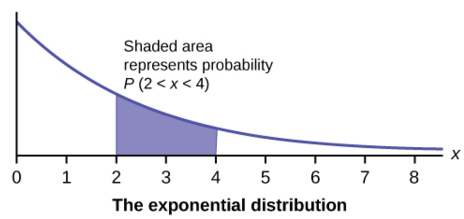

Figure 5.2.2 The graph shows an Exponential Distribution with the area between $x=2$ and $x=4$ shaded to represent the probability that the value of the random variable X is in the interval between two and four.

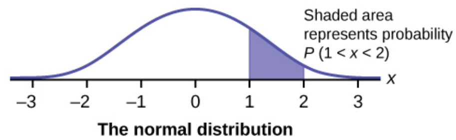

Figure 5.2.3 The graph shows the Standard Normal Distribution with the area between $x = 1$ and $x = 2$ shaded to represent the probability that the value of the random variable X is in the interval between one and two.

For continuous probability distributions, PROBABILITY = AREA.

Example 5.2.1

Consider the function $f(x) = \frac{1}{20}$ for $0 \leq x \leq 20$. The graph of $f(x) = \frac{1}{20}$ is a horizontal line. However, since $0 \leq x \leq 20$, $f(x)$ is restricted to the portion between $x = 0$ and $x = 20$, inclusive.

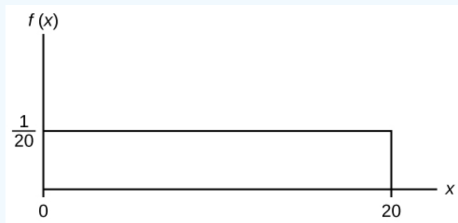

Figure 5.2.4

$f(x) = \frac{1}{20}$ for $0 \leq x \leq 20$.

The graph of $f(x) = \frac{1}{20}$ is a horizontal line segment when $0 \leq x \leq 20$.

The area between $f(x) = \frac{1}{20}$ where $0 \leq x \leq 20$ and the x-axis is the area of a rectangle with base $= 20$ and height $= \frac{1}{20}$.

$$\text{AREA} = 20\left(\frac{1}{20}\right) = 1$$

Suppose we want to find the area between \(bf{f(x)) = \frac{1}{20}}\) and the *x*-axis where $\mathbf{0 < x < 2}$.

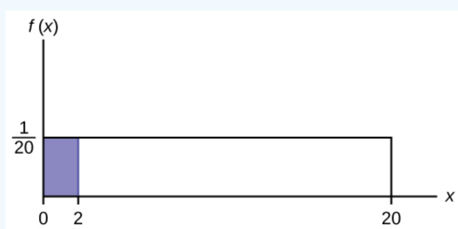

Figure 5.2.5

$$\text{AREA} = (2 - 0)\left(\frac{1}{20}\right) = 0.1$$

$$(2 - 0) = 2 = \text{base of rectangle}$$

REMINDER

area of a rectangle = (base)(height).

The area corresponds to a probability. The probability that x is between zero and two is 0.1, which can be written mathematically as $P(0 < x < 2) = P(x < 2) = 0.1$.

Suppose we want to find the area between $\mathbf{f(x) = \frac{1}{20}}$ and the *x*-axis where $\mathbf{4 < x < 15}$.

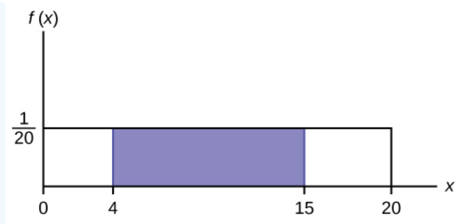

Figure 5.2.6

$\text{AREA} = (15 - 4)\left(\frac{1}{20}\right) = 0.55$

$(15 - 4) = 11 = \text{the base of a rectangle}$

The area corresponds to the probability $P(4 < x < 15) = 0.55$.

Suppose we want to find $P(x = 15)$. On an x-y graph, $x = 15$ is a vertical line. A vertical line has no width (or zero width). Therefore, $P(x = 15) = \text{(base)(height)} = (0)\left(\frac{1}{20}\right) = 0$

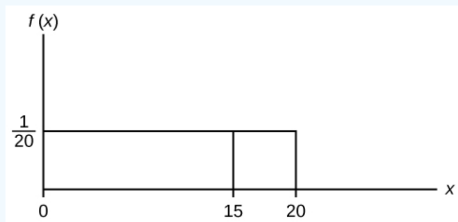

Figure 5.2.7

$P(X \leq x)$, which can also be written as $P(X < x)$ for continuous distributions, is called the cumulative distribution function or CDF. Notice the "less than or equal to" symbol. We can also use the CDF to calculate $P(X > x)$. The CDF gives "area to the left" and $P(X > x)$ gives "area to the right." We calculate $P(X > x)$ for continuous distributions as follows: $P(X > x) = 1 - P(X < x)$.

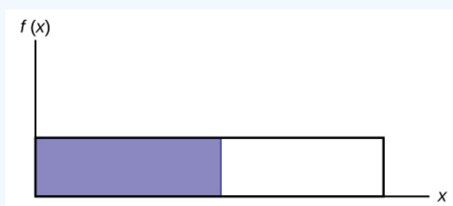

Figure 5.2.8

Label the graph with $f(x)$ and x. Scale the x and y axes with the maximum x and y values. $f(x) = \frac{1}{20}, 0 \leq x \leq 20$.

To calculate the probability that x is between two values, look at the following graph. Shade the region between $x = 2.3$ and $x = 12.7$. Then calculate the shaded area of a rectangle.

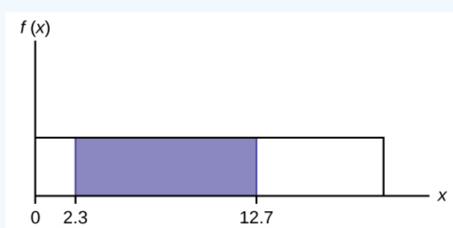

Figure 5.2.9

$P(2.3 < x < 12.7) = (\text{ base })(\text{ height }) = (12.7 - 2.3)\left(\frac{1}{20}\right) = 0.52$

Exercise 5.2.1

Consider the function $f(x) = \frac{1}{8}$ for $0 \leq x \leq 8$. Draw the graph of $f(x))$ and find $P(2.5 < x < 7.5)$.