

## 4.4: Skewness and Kurtosis

As usual, our starting point is a random experiment, modeled by a probability space  $(\Omega, \mathcal{F}, P)$ . So to review,  $\Omega$  is the set of outcomes,  $\mathcal{F}$  the collection of events, and  $\mathbb{P}$  the probability measure on the sample space  $(\Omega, \mathcal{F})$ . Suppose that  $X$  is a real-valued random variable for the experiment. Recall that the mean of  $X$  is a measure of the center of the distribution of  $X$ . Furthermore, the variance of  $X$  is the second moment of  $X$  about the mean, and measures the spread of the distribution of  $X$  about the mean. The third and fourth moments of  $X$  about the mean also measure interesting (but more subtle) features of the distribution. The third moment measures *skewness*, the lack of symmetry, while the fourth moment measures *kurtosis*, roughly a measure of the fatness in the tails. The actual numerical measures of these characteristics are standardized to eliminate the physical units, by dividing by an appropriate power of the standard deviation. As usual, we assume that all expected values given below exist, and we will let  $\mu = \mathbb{E}(X)$  and  $\sigma^2 = \text{var}(X)$ . We assume that  $\sigma > 0$ , so that the random variable is really random.

### Basic Theory

#### Skewness

The *skewness* of  $X$  is the third moment of the standard score of  $X$ :

$$\text{skew}(X) = \mathbb{E} \left[ \left( \frac{X - \mu}{\sigma} \right)^3 \right] \quad (4.4.1)$$

The distribution of  $X$  is said to be *positively skewed*, *negatively skewed* or *unskewed* depending on whether  $\text{skew}(X)$  is positive, negative, or 0.

In the unimodal case, if the distribution is positively skewed then the probability density function has a long tail to the right, and if the distribution is negatively skewed then the probability density function has a long tail to the left. A symmetric distribution is unskewed.

Suppose that the distribution of  $X$  is symmetric about  $a$ . Then

1.  $\mathbb{E}(X) = a$
2.  $\text{skew}(X) = 0$ .

**Proof**

By assumption, the distribution of  $a - X$  is the same as the distribution of  $X - a$ . We proved part (a) in the section on properties of expected value. Thus,  $\text{skew}(X) = \mathbb{E}[(X - a)^3] / \sigma^3$ . But by symmetry and linearity,  $\mathbb{E}[(X - a)^3] = \mathbb{E}[(a - X)^3] = -\mathbb{E}[(X - a)^3]$ , so it follows that  $\mathbb{E}[(X - a)^3] = 0$ .

The converse is not true—a non-symmetric distribution can have skewness 0. Examples are given in Exercises (30) and (31) below.

$\text{skew}(X)$  can be expressed in terms of the first three moments of  $X$ .

$$\text{skew}(X) = \frac{\mathbb{E}(X^3) - 3\mu\mathbb{E}(X^2) + 2\mu^3}{\sigma^3} = \frac{\mathbb{E}(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3} \quad (4.4.2)$$

**Proof**

Note that  $(X - \mu)^3 = X^3 - 3X^2\mu + 3X\mu^2 - \mu^3$ . From the linearity of expected value we have

$$\mathbb{E}[(X - \mu)^3] = \mathbb{E}(X^3) - 3\mu\mathbb{E}(X^2) + 3\mu^2\mathbb{E}(X) - \mu^3 = \mathbb{E}(X^3) - 3\mu\mathbb{E}(X^2) + 2\mu^3 \quad (4.4.3)$$

The second expression follows from substituting  $\mathbb{E}(X^2) = \sigma^2 + \mu^2$ .

Since skewness is defined in terms of an odd power of the standard score, it's invariant under a linear transformation with positive slope (a location-scale transformation of the distribution). On the other hand, if the slope is negative, skewness changes sign.

Suppose that  $a \in \mathbb{R}$  and  $b \in \mathbb{R} \setminus \{0\}$ . Then

1.  $\text{skew}(a + bX) = \text{skew}(X)$  if  $b > 0$
2.  $\text{skew}(a + bX) = -\text{skew}(X)$  if  $b < 0$

**Proof**

Let  $Z = (X - \mu)/\sigma$ , the standard score of  $X$ . Recall from the section on variance that the standard score of  $a + bX$  is  $Z$  if  $b > 0$  and is  $-Z$  if  $b < 0$ .

Recall that location-scale transformations often arise when physical units are changed, such as inches to centimeters, or degrees Fahrenheit to degrees Celsius.

### Kurtosis

The *kurtosis* of  $X$  is the fourth moment of the standard score:

$$\text{kurt}(X) = \mathbb{E} \left[ \left( \frac{X - \mu}{\sigma} \right)^4 \right] \quad (4.4.4)$$

*Kurtosis* comes from the Greek word for *bulging*. Kurtosis is always positive, since we have assumed that  $\sigma > 0$  (the random variable really is random), and therefore  $\mathbb{P}(X \neq \mu) > 0$ . In the unimodal case, the probability density function of a distribution with large kurtosis has fatter tails, compared with the probability density function of a distribution with smaller kurtosis.

$\text{kurt}(X)$  can be expressed in terms of the first four moments of  $X$ .

$$\text{kurt}(X) = \frac{\mathbb{E}(X^4) - 4\mu\mathbb{E}(X^3) + 6\mu^2\mathbb{E}(X^2) - 3\mu^4}{\sigma^4} = \frac{\mathbb{E}(X^4) - 4\mu\mathbb{E}(X^3) + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4} \quad (4.4.5)$$

Proof

Note that  $(X - \mu)^4 = X^4 - 4X^3\mu + 6X^2\mu^2 - 4X\mu^3 + \mu^4$ . From linearity of expected value, we have

$$\mathbb{E}[(X - \mu)^4] = \mathbb{E}(X^4) - 4\mu\mathbb{E}(X^3) + 6\mu^2\mathbb{E}(X^2) - 4\mu^3\mathbb{E}(X) + \mu^4 = \mathbb{E}(X^4) - 4\mu\mathbb{E}(X^3) + 6\mu^2\mathbb{E}(X^2) - 3\mu^4 \quad (4.4.6)$$

The second expression follows from the substitution  $\mathbb{E}(X^2) = \sigma^2 + \mu^2$ .

Since kurtosis is defined in terms of an even power of the standard score, it's invariant under linear transformations.

Suppose that  $a \in \mathbb{R}$  and  $b \in \mathbb{R} \setminus \{0\}$ . Then  $\text{kurt}(a + bX) = \text{kurt}(X)$ .

Proof

As before, let  $Z = (X - \mu)/\sigma$  denote the standard score of  $X$ . Then the standard score of  $a + bX$  is  $Z$  if  $b > 0$  and is  $-Z$  if  $b < 0$ .

We will show in [below](#) that the kurtosis of the standard normal distribution is 3. Using the standard normal distribution as a benchmark, the *excess kurtosis* of a random variable  $X$  is defined to be  $\text{kurt}(X) - 3$ . Some authors use the term *kurtosis* to mean what we have defined as *excess kurtosis*.

### Computational Exercises

As always, be sure to try the exercises yourself before expanding the solutions and answers in the text.

#### Indicator Variables

Recall that an *indicator random variable* is one that just takes the values 0 and 1. Indicator variables are the building blocks of many counting random variables. The corresponding distribution is known as the *Bernoulli distribution*, named for Jacob Bernoulli.

Suppose that  $X$  is an indicator variable with  $\mathbb{P}(X = 1) = p$  where  $p \in (0, 1)$ . Then

1.  $\mathbb{E}(X) = p$
2.  $\text{var}(X) = p(1 - p)$
3.  $\text{skew}(X) = \frac{1 - 2p}{\sqrt{p(1 - p)}}$
4.  $\text{kurt}(X) = \frac{1 - 3p + 3p^2}{p(1 - p)}$

Proof

Parts (a) and (b) have been derived before. All four parts follow easily from the fact that  $X^n = X$  and hence  $\mathbb{E}(X^n) = p$  for  $n \in \mathbb{N}_+$ .

Open the binomial coin experiment and set  $n = 1$  to get an indicator variable. Vary  $p$  and note the change in the shape of the probability density function.

### Dice

Recall that a *fair* die is one in which the faces are equally likely. In addition to fair dice, there are various types of crooked dice. Here are three:

- An *ace-six flat die* is a six-sided die in which faces 1 and 6 have probability  $\frac{1}{4}$  each while faces 2, 3, 4, and 5 have probability  $\frac{1}{8}$  each.
- A *two-five flat die* is a six-sided die in which faces 2 and 5 have probability  $\frac{1}{4}$  each while faces 1, 3, 4, and 6 have probability  $\frac{1}{8}$  each.
- A *three-four flat die* is a six-sided die in which faces 3 and 4 have probability  $\frac{1}{4}$  each while faces 1, 2, 5, and 6 have probability  $\frac{1}{8}$  each.

A flat die, as the name suggests, is a die that is not a cube, but rather is shorter in one of the three directions. The particular probabilities that we use ( $\frac{1}{4}$  and  $\frac{1}{8}$ ) are fictitious, but the essential property of a flat die is that the opposite faces on the shorter axis have slightly larger probabilities than the other four faces. Flat dice are sometimes used by gamblers to cheat.

A standard, fair die is thrown and the score  $X$  is recorded. Compute each of the following:

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1.  $\frac{7}{2}$
2.  $\frac{35}{12}$
3. 0
4.  $\frac{303}{175}$

An ace-six flat die is thrown and the score  $X$  is recorded. Compute each of the following:

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1.  $\frac{7}{2}$
2.  $\frac{15}{4}$
3. 0
4.  $\frac{37}{25}$

A two-five flat die is thrown and the score  $X$  is recorded. Compute each of the following:

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1.  $\frac{7}{2}$
2.  $\frac{11}{4}$
3. 0
4.  $\frac{197}{121}$

A three-four flat die is thrown and the score  $X$  is recorded. Compute each of the following:

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$

3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1.  $\frac{7}{2}$
2.  $\frac{9}{4}$
3. 0
4.  $\frac{59}{27}$

All four die distributions above have the same mean  $\frac{7}{2}$  and are symmetric (and hence have skewness 0), but differ in variance and kurtosis.

Open the dice experiment and set  $n = 1$  to get a single die. Select each of the following, and note the shape of the probability density function in comparison with the computational results above. In each case, run the experiment 1000 times and compare the empirical density function to the probability density function.

1. fair
2. ace-six flat
3. two-five flat
4. three-four flat

### Uniform Distributions

Recall that the continuous uniform distribution on a bounded interval corresponds to selecting a point *at random* from the interval. Continuous uniform distributions arise in geometric probability and a variety of other applied problems.

Suppose that  $X$  has uniform distribution on the interval  $[a, b]$ , where  $a, b \in \mathbb{R}$  and  $a < b$ . Then

1.  $\mathbb{E}(X) = \frac{1}{2}(a + b)$
2.  $\text{var}(X) = \frac{1}{12}(b - a)^2$
3.  $\text{skew}(X) = 0$
4.  $\text{kurt}(X) = \frac{9}{5}$

Proof

Parts (a) and (b) we have seen before. For parts (c) and (d), recall that  $X = a + (b - a)U$  where  $U$  has the uniform distribution on  $[0, 1]$  (the *standard uniform distribution*). Hence it follows from the formulas for [skewness](#) and [kurtosis](#) under linear transformations that  $\text{skew}(X) = \text{skew}(U)$  and  $\text{kurt}(X) = \text{kurt}(U)$ . Since  $\mathbb{E}(U^n) = 1/(n + 1)$  for  $n \in \mathbb{N}_+$ , it's easy to compute the skewness and kurtosis of  $U$  from the computational formulas [skewness](#) and [kurtosis](#). Of course, the fact that  $\text{skew}(X) = 0$  also follows trivially from the symmetry of the distribution of  $X$  about the mean.

Open the special distribution simulator, and select the continuous uniform distribution. Vary the parameters and note the shape of the probability density function in comparison with the moment results in the last exercise. For selected values of the parameter, run the simulation 1000 times and compare the empirical density function to the probability density function.

### The Exponential Distribution

Recall that the *exponential distribution* is a continuous distribution on  $[0, \infty)$  with probability density function  $f$  given by

$$f(t) = re^{-rt}, \quad t \in [0, \infty) \quad (4.4.7)$$

where  $r \in (0, \infty)$  is the *rate parameter*. This distribution is widely used to model failure times and other “arrival times”. The exponential distribution is studied in detail in the chapter on the Poisson Process.

Suppose that  $X$  has the exponential distribution with rate parameter  $r > 0$ . Then

1.  $\mathbb{E}(X) = \frac{1}{r}$
2.  $\text{var}(X) = \frac{1}{r^2}$
3.  $\text{skew}(X) = 2$
4.  $\text{kurt}(X) = 9$

Proof

These results follow from the computational formulas for [skewness](#) and [kurtosis](#) and the general moment formula  $\mathbb{E}(X^n) = n!/r^n$  for  $n \in \mathbb{N}$ .

Note that the skewness and kurtosis do not depend on the rate parameter  $r$ . That's because  $1/r$  is a scale parameter for the exponential distribution

Open the gamma experiment and set  $n = 1$  to get the exponential distribution. Vary the rate parameter and note the shape of the probability density function in comparison to the moment results in the last exercise. For selected values of the parameter, run the experiment 1000 times and compare the empirical density function to the true probability density function.

### Pareto Distribution

Recall that the *Pareto distribution* is a continuous distribution on  $[1, \infty)$  with probability density function  $f$  given by

$$f(x) = \frac{a}{x^{a+1}}, \quad x \in [1, \infty) \quad (4.4.8)$$

where  $a \in (0, \infty)$  is a parameter. The Pareto distribution is named for Vilfredo Pareto. It is a heavy-tailed distribution that is widely used to model financial variables such as income. The Pareto distribution is studied in detail in the chapter on Special Distributions.

Suppose that  $X$  has the Pareto distribution with shape parameter  $a > 0$ . Then

1.  $\mathbb{E}(X) = \frac{a}{a-1}$  if  $a > 1$
2.  $\text{var}(X) = \frac{a}{(a-1)^2(a-2)}$  if  $a > 2$
3.  $\text{skew}(X) = \frac{2(1+a)}{a-3} \sqrt{1 - \frac{2}{a}}$  if  $a > 3$
4.  $\text{kurt}(X) = \frac{3(a-2)(3a^2+a+2)}{a(a-3)(a-4)}$  if  $a > 4$

Proof

These results follow from the standard computational formulas for [skewness](#) and [kurtosis](#) and the general moment formula  $\mathbb{E}(X^n) = \frac{a}{a-n}$  if  $n \in \mathbb{N}$  and  $n < a$ .

Open the special distribution simulator and select the Pareto distribution. Vary the shape parameter and note the shape of the probability density function in comparison to the moment results in the last exercise. For selected values of the parameter, run the experiment 1000 times and compare the empirical density function to the true probability density function.

### The Normal Distribution

Recall that the *standard normal distribution* is a continuous distribution on  $\mathbb{R}$  with probability density function  $\phi$  given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R} \quad (4.4.9)$$

Normal distributions are widely used to model physical measurements subject to small, random errors and are studied in detail in the chapter on Special Distributions.

Suppose that  $Z$  has the standard normal distribution. Then

1.  $\mathbb{E}(Z) = 0$
2.  $\text{var}(Z) = 1$
3.  $\text{skew}(Z) = 0$
4.  $\text{kurt}(Z) = 3$

Proof

Parts (a) and (b) were derived in the previous sections on expected value and variance. Part (c) follows from symmetry. For part (d), recall that  $\mathbb{E}(Z^4) = 3\mathbb{E}(Z^2) = 3$ .

More generally, for  $\mu \in \mathbb{R}$  and  $\sigma \in (0, \infty)$ , recall that the normal distribution with mean  $\mu$  and standard deviation  $\sigma$  is a continuous distribution on  $\mathbb{R}$  with probability density function  $f$  given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R} \quad (4.4.10)$$

However, we also know that  $\mu$  and  $\sigma$  are location and scale parameters, respectively. That is, if  $Z$  has the standard normal distribution then  $X = \mu + \sigma Z$  has the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ .

If  $X$  has the normal distribution with mean  $\mu \in \mathbb{R}$  and standard deviation  $\sigma \in (0, \infty)$ , then

1.  $\text{skew}(X) = 0$
2.  $\text{kurt}(X) = 3$

Proof

The results follow immediately from the formulas for [skewness](#) and [kurtosis](#) under linear transformations and the previous result.

Open the special distribution simulator and select the normal distribution. Vary the parameters and note the shape of the probability density function in comparison to the moment results in the last exercise. For selected values of the parameters, run the experiment 1000 times and compare the empirical density function to the true probability density function.

### The Beta Distribution

The distributions in this subsection belong to the family of *beta distributions*, which are continuous distributions on  $[0, 1]$  widely used to model random proportions and probabilities. The beta distribution is studied in detail in the chapter on Special Distributions.

Suppose that  $X$  has probability density function  $f$  given by  $f(x) = 6x(1 - x)$  for  $x \in [0, 1]$ . Find each of the following:

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1.  $\frac{1}{2}$
2.  $\frac{1}{20}$
3. 0
4.  $\frac{15}{7}$

Suppose that  $X$  has probability density function  $f$  given by  $f(x) = 12x^2(1 - x)$  for  $x \in [0, 1]$ . Find each of the following:

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1.  $\frac{3}{5}$
2.  $\frac{1}{25}$
3.  $-\frac{2}{7}$
4.  $\frac{33}{14}$

Suppose that  $X$  has probability density function  $f$  given by  $f(x) = 12x(1 - x)^2$  for  $x \in [0, 1]$ . Find each of the following:

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1.  $\frac{2}{5}$
2.  $\frac{1}{25}$
3.  $\frac{2}{7}$
4.  $\frac{33}{14}$

Open the special distribution simulator and select the beta distribution. Select the parameter values below to get the distributions in the last three exercises. In each case, note the shape of the probability density function in relation to the calculated moment results. Run the simulation 1000 times and compare the empirical density function to the probability density function.

1.  $a = 2, b = 2$
2.  $a = 3, b = 2$
3.  $a = 2, b = 3$

Suppose that  $X$  has probability density function  $f$  given by  $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$  for  $x \in (0, 1)$ . Find

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1.  $\frac{1}{2}$
2.  $\frac{1}{8}$
3. 0
4. 96

The particular beta distribution in the last exercise is also known as the (standard) *arcsine distribution*. It governs the last time that the Brownian motion process hits 0 during the time interval  $[0, 1]$ . The arcsine distribution is studied in more generality in the chapter on Special Distributions.

Open the Brownian motion experiment and select the last zero. Note the shape of the probability density function in relation to the moment results in the last exercise. Run the simulation 1000 times and compare the empirical density function to the probability density function.

### Counterexamples

The following exercise gives a simple example of a discrete distribution that is not symmetric but has skewness 0.

Suppose that  $X$  is a discrete random variable with probability density function  $f$  given by  $f(-3) = \frac{1}{10}, f(-1) = \frac{1}{2}, f(2) = \frac{2}{5}$ . Find each of the following and then show that the distribution of  $X$  is not symmetric.

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

Answer

1. 0
2. 3
3. 0
4.  $\frac{5}{3}$

The PDF  $f$  is clearly not symmetric about 0, and the mean is the only possible point of symmetry.

The following exercise gives a more complicated continuous distribution that is not symmetric but has skewness 0. It is one of a collection of distributions constructed by [Erik Meijer](#).

Suppose that  $U$ ,  $V$ , and  $I$  are independent random variables, and that  $U$  is normally distributed with mean  $\mu = -2$  and variance  $\sigma^2 = 1$ ,  $V$  is normally distributed with mean  $\nu = 1$  and variance  $\tau^2 = 2$ , and  $I$  is an indicator variable with  $\mathbb{P}(I = 1) = p = \frac{1}{3}$ . Let  $X = IU + (1 - I)V$ . Find each of the following and then show that the distribution of  $X$  is not symmetric.

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\text{skew}(X)$
4.  $\text{kurt}(X)$

### Solution

The distribution of  $X$  is a *mixture* of normal distributions. The PDF is  $f = pg + (1 - p)h$  where  $g$  is the normal PDF of  $U$  and  $h$  is the normal PDF of  $V$ . However, it's best to work with the random variables. For  $n \in \mathbb{N}_+$ , note that  $I^n = I$  and  $(1 - I)^n = 1 - I$  and note also that the random variable  $I(1 - I)$  just takes the value 0. It follows that

$$X^n = IU^n + (1 - I)V^n, \quad n \in \mathbb{N}_+ \quad (4.4.11)$$

So now, using standard results for the normal distribution,

1.  $\mathbb{E}(X) = p\mu + (1 - p)\nu = 0$  .
2.  $\text{var}(X) = \mathbb{E}(X^2) = p(\sigma^2 + \mu^2) + (1 - p)(\tau^2 + \nu^2) = \frac{11}{3}$
3.  $\mathbb{E}(X^3) = p(3\mu\sigma^2 + \mu^3) + (1 - p)(3\nu\tau^2 + \nu^3) = 0$  so  $\text{skew}(X) = 0$
4.  $\mathbb{E}(X^4) = p(3\sigma^4 + 6\sigma^2\mu^2 + \mu^4) + (1 - p)(3\tau^4 + 6\tau^2\nu^2 + \nu^4) = 31$  so  $\text{kurt}(X) = \frac{279}{121} \approx 2.306$

The graph of the PDF  $f$  of  $X$  is given below. Note that  $f$  is not symmetric about 0. (Again, the mean is the only possible point of symmetry.)

The PDF of  $X$   


This page titled [4.4: Skewness and Kurtosis](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.